2018

Assessing Ratio-Based Fatigue Indexes Using a Single Channel EEG

Lucas B. Coffey
University of North Florida

ASSESSING RATIO-BASED FATIGUE INDEXES USING A SINGLE CHANNEL EEG

By

Lucas B. Coffey

A thesis prospectus submitted to the
School of Computing
in partial fulfillment of the requirements for the degree of

Master of Science in Computing and Information Sciences

UNIVERSITY OF NORTH FLORIDA
SCHOOL OF COMPUTING

April, 2018

The thesis "Assessing Ratio-based Fatigue Indexes using a Single Channel EEG" submitted by Lucas Coffey in partial fulfillment of the requirements for the degree of Master of Science in Computing and Information Sciences has been

Approved by the thesis committee: Date

Dr. Karthikeyan Umapathy
Thesis Advisor and Committee Chairperson

Dr. Ching-Hua Chuan

Dr. Debbie Wang

Accepted for the School of Computing:

Dr. Sherif Elfayoumy
Director of the School

Accepted for the College of Computing, Engineering, and Construction:

Dr. Mark Tumeo
Dean of the College

Accepted for the University:

Dr. John Kantner
Dean of the Graduate School

ACKNOWLEDGEMENT

I would like to express my special appreciation to my advisor, Dr. Karthikeyan Umapathy for being a tremendous mentor for me. I would like to thank him for all his advice and inputs regarding my thesis. His advice on this thesis and the suggestions to improve materials presented were invaluable. I would also like to thank Dr. Debbie Wang and Dr. Ching-Hua Chuan, for serving as my committee members, despite their busy schedules. Their suggestions and advice were tremendously helpful.

A special thanks to Mr. Jim Littleton for reviewing my thesis report, and ensuring its compliance with the University's requirements. I would like to acknowledge Dr. Roger Eggen for his guidance through the Graduate School and School of Computing processes. In addition, I want to give special thanks to the School of Computing and the Department of Psychology at the University of North Florida for allowing me to use all the required resources to complete this thesis.

Finally, a special thanks to my wife Ashley for her unwavering support throughout this endeavor, without which this insurmountable task would not have been accomplished.

CONTENTS

FIGURES

TABLES

EQUATIONS

ABSTRACT

Driver fatigue is a state of reduced mental alertness which impairs the performance of a range of cognitive and psychomotor tasks, including driving. According to the National Highway Traffic Safety Administration, driver fatigue was responsible for 72,000 accidents that lead to more than 800 deaths in 2015. A reliable method of driver fatigue detection is needed to prevent such accidents. There has been a great deal of research into studying driver fatigue via electroencephalography (EEG) to analyze brain wave data. These research works have produced three competing EEG data-based ratios that have the potential to detect driver fatigue.

Research has shown these three ratios trend downward as fatigue increases. However, no empirical research has been conducted to determine whether drivers begin to feel fatigue at a certain Percent Change from an alert state to a fatigue state in one or more of these ratios. If a Percent Change could be identified for which drivers begin to feel fatigue, then it could be used as a method of fatigue detection in real-time system. This research focuses on answering this question by collecting brain wave data via an EEG device over a 60-minute driving session for 10 University of North Florida (UNF) students. A frequency distribution and cluster analysis was done to identify a common Percent Change for the participants who experienced fatigue. The results of the analysis were compared to a subset of users who did not experience fatigue to validate the findings. The project was approved by the UNF IRB on Nov. 1, 2016 (reference number 475514-4).

Chapter 1

INTRODUCTION

Driver fatigue is a state of reduced mental alertness which impairs the performance of a range of cognitive and psychomotor tasks, including driving (Lal & Craig, 2001). Driver fatigue tends to manifest itself after periods of long driving, causing a significant problem for transportation employees that are forced to drive for extended periods of time. According to the National Highway Traffic Safety Administration, driver fatigue was responsible for 72,000 accidents that lead to more than 800 deaths in 2015 (National Highway Traffic Safety Adminstration, 2017). A reliable detection method for sensing driver fatigue in real time could potentially save lives.

A great deal of research into devising such a detection method has been conducted. Some of these techniques include eye movement tracking, heart rate analysis, video based facial recognition, and electroencephalography (EEG) (Lal & Craig, 2001). The former methods have not proven to be reliable methods for driver fatigue detection thus far. For example, eye movement and facial recognition rely on camera technology which can be subject to error when natural occurrences, such as sun glare, occur. Much of the research surrounding heart rate technology relies on embedding sensors into a steering wheel (Minnesota Department of Transportation, 2015). This results in a need for the driver to maintain constant contact with the wheel in order not to affect the detection algorithm.

These limitations present a significant problem for implementing a real-time driver fatigue detection system using such methods.

EEG devices present a much more practical and reliable solution because they are not vulnerable to external influence, and they can be easily embedded in a wearable device. EEGs are devices that measure electrical potential from the scalp that is generated when neurons fire within the brain. Research has proven that certain frequencies of this electrical potential are indicative of certain mental states (Lal & Craig, 2001). EEG devices can also be paired with modern computing devices to create brain-computer interfaces. Such an interface allows developers to interact with brain wave data programmatically, enabling the potential for creating a real-time fatigue detection system.

However, before such a system could be designed, a method for detecting fatigue must be established. During the research for this thesis three different EEG ratios were identified that show noticeable trends as fatigue increases. However, these research works do not identify the point at which fatigue beings to occur. Before these ratios can be used for real time fatigue detection purposes, it is crucial that research is conducted to identify a reliable method of driver fatigue detection. In this thesis, we examined Percent Change from an alert state to a fatigue state as a method for detecting driver fatigue by capturing and analyzing data from 10 participants who simulated driving. By proving that a relationship exists between Percent Change and driver fatigue that is common across the participant base we may identify a potential method for detecting driver fatigue. The identification of such a method would help in determining when an alarm could be triggered by a real-time system using EEG data before the driver falls asleep.

1.1 Problem Statement

Fatigue is a problem that plagues many in the civilian and professional world. Driver fatigue has been said to cause 20-30% of all roadway accidents, resulting in billions of dollars in road related injury costs (Lal & Craig, 2001). This problem is exacerbated by long periods of monotonous driving, particularly at night (Lal & Craig, 2001). In safety critical occupations, such as professional truck driving, this problem can potentially lead to major, life threating accidents that could potentially be prevented. This begs the question of how we might prevent such accidents.

Fatigue has been the subject of EEG studies for years. This has led to a widely held understanding that certain brain frequencies are indicative of different mental states such as alertness, relaxation, sleep, and deep sleep (Lal & Craig, 2001). Armed with this knowledge, developers can couple EEGs with modern computing devices to create a Brain-Computer Interface (BCI) that could potentially detect driver fatigue in real-time.

Fatigue can be detected by tracing variations in the Alpha, Beta, and Theta brain wave spectrums provided by BCI devices. Previous research has analyzed the relative power spectrum of these waves to detect mental fatigue, but has often resulted in situations where minor fatigue went undetected (Kar , Bhagat, & Routray, 2010). To account for this, many researchers have turned to a new method of evaluation where ratios of slow to fast moving waves were used to set a fatigue index (Kar , Bhagat, & Routray, 2010) (Jap, Lal, Fischer, & Bekiaris, 2009) (Ming-ai , Cheng , & Jin-Fu, 2010).

While ratio-based approaches have shown promise, there has been no research to determine at which point fatigue begins to occur for a driver. The purpose of this thesis is to identify the Percent Change from an alert state to a fatigue state for each of these ratios while driving, and to determine if this Percent Change is common to the majority of the participants in our study. If such a trend were to be identified, it could be used as a detection mechanism for a real-time fatigue detection system in future work.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

2.1 What is Fatigue?

While there has been no globally accepted definition of fatigue, it has been described as "a transitory period between awake and sleep and if uninterrupted, can lead to sleep" (Lal & Craig, 2001). Fatigue has been known to reduce attention, compromise judgment, and impact decision making skills (Occupation Safety and Health Administration, 2015). Lamond and Dawson (2002) likened a driver that has gone without sleep for a 28-hour period to that of a person driving under the influence of the alcohol.

Fatigue can be divided into physical and mental states. Physical fatigue is associated with physiological symptoms such as muscle fatigue, while mental fatigue is considered to by psychological in nature. Mental fatigue is typically gradual and can leave the person in a state of relaxation, which will ensure reduced attention and alertness. Boredom has also been linked to mental fatigue because it causes similar symptoms of reduced attention and weariness. This happens when the mind perceives external stimuli to be low or identical. This reduces sensory impulses that are fed to the central nervous system, inducing a bored state (Lal & Craig, 2001).

Driver fatigue can be thought of as an extension of mental fatigue where drivers feel similar feelings of weariness after prolonged periods of driving. Factors that contribute to driver fatigue are monotonous periods of driving, length of journey, time of day, and irregular work schedules (Lal & Craig, 2000). While driver fatigue can manifest itself in all drivers, it is a particular problem in the transportation industry.

A need exists to measure the physiological features of driver fatigue so there can be an intervention before an accident occurs. EEG studies have been used to study such features in the past. EEG devices allow the mental state of a person to be measured by detecting micro-volts of electricity that are generated by neural activity. This activity can be broken down into a discrete set of frequencies, all of which are indicative of a certain mental state that can help in the development of a driver fatigue detection mechanism. However, to interact with these data programmatically, so that a real-time fatigue detection system can be created, a brain-computer interface is needed.

2.2 What are Brain Computer Interfaces?

A Brain-Computer Interface (BCI) has been defined as "a method of communication based on neural activity generated by the brain and is independent of its normal pathways of peripheral nerves and muscles" (Vallabhaneni, Wang, & He, 2005). To expand upon this, consider the process of human movement. The initial process begins with the intent to move. This intent prompts signals from the brain to the nervous system, which in turn activates the appropriate muscles (Graimann, Allison, & Pfurt, 2010). A BCI can circumvent this process by measuring the electrical signals in the brain, the signals created by the intent to move, and relay them to some computer application for processing (Graimann, Allison, & Pfurt, 2010). This same process can be used to perform a multitude of activities such as playing a game, manipulating a mouse cursor, or controlling household objects. Brain signals can be measured in one of two ways, invasively or non-invasively.

2.2.1 Invasive

Invasive methods typically require surgery so that electrodes can be placed on the surface of the brain. These electrodes record a signal from the brain which is then documented in an intra-cortical recording where it can be used by an application (Graimann, Allison, & Pfurt, 2010). This offers a better signal quality in comparison to the non-invasive approach, but also carries a much more significant risk to the subject due to the need for brain surgery (Lebedev & Nicolelis, 2006).

2.2.2 Non-Invasive

In contrast, a non-invasive approach does not require surgery. Instead sensors are placed on top of the scalp to measure the fluctuations in brain activity by means of EEGs. EEGs are the more commonly known BCI technique. They often resemble helmets or caps that are capable of connecting to a computing device. EEGs have proven to be a valuable research method for years, likely attributed to their inexpensive nature and high usability.

EEG history can be dated back to 1870 when Richard Canton first discovered electrical activity in animal brains followed by the invention of the actual first EEG device by Hans Burner in 1924 (Neurosky, 2015). Burner's invention lead to the classification of such brain activity into distinct rhythms or frequencies. These frequencies have been proved to be indicative of certain mental states such as alertness, relaxedness, and sleep (Rubin, 2009).

It is important to note that while EEGs can provide a means for understanding brain activity, they do not have the capability to monitor thoughts or feelings. EEGs provide a passive means of monitoring electric activity that is generated by neurons firing within the brain. Like all electricity, this type of activity can be measured and classified. In the case of EEGs, the data are usually classified into five different types of frequencies (see Sections 2.3.2-2.3.6).

2.3 Brain Waves

At the core of human thought are neurons. Neurons exist to send and receive information to and from other neurons via synapses (Shier, Butler, & Lewis, 2009). Think of neurons as a power source and a synapse as a wire that can transfer that power. When neurons communicate with one another they generate tiny busts of electrical activity which can be measured by an EEG device (Rubin, 2009). This activity can be classified according to frequency of electrical activity, ranging from 0-42 HZ (Rubin, 2009). Currently, there are five commonly recognized frequencies that the brain operates at: Alpha, Beta, Delta, Theta, and Gamma.

2.3.1 Fast Fourier Transform (FFT)

It is important to understand how EEG data are converted into the frequency spectrum from its original voltage. To re-iterate how brainwave data are generated, when neurons communicate with one another they generate electrical potential within the brain. The EEG electrodes capture this potential which is generally known as "raw data". In signal processing terms, the raw data exists in the time domain. To more easily classify and understand these data they must be transformed into the frequency domain. This results in the classification of the five frequencies listed below. This transformation is done via Fast Fourier Transform (FFT).

Fast Fourier Transform (FFT) is a signal processing technique that converts a signal from its representation in the time domain (a measure of amplitude over time) to the frequency domain (a measure of amplitude over frequency for a given time-period) (Cochran, Cooley, Favin, & He, 1967). "FFT decomposes the recorded EEG into a voltage by frequency spectral graph commonly called the "power spectrum, with power being the square of the EEG magnitude, and magnitude being the integral average of the amplitude of the EEG signal, measured from + peak to – peak, across the time sampled" (Rocha, Thomaz, Rocha, & Vieito, 2017). This allows researchers to analyze the power of different frequencies of a signal. For EEG data purposes, the raw data are commonly converted into the five frequencies described below.

2.3.2 Alpha (8-13 Hz)

The Alpha spectrum ranges from 8-13 HZ (Hu, Guo, Liu, & Wang, 2017). Alpha waves are associated with deep relaxation and are usually present when the eyes are closed (Rubin, 2009). They typically attenuate when the eyes are opened and during periods of drowsiness and sleep (Baars & Gage, 2007). They can be observed in the posterior and occipital sections of the brain with peak-peak amplitudes of around 50 micro-volts (Teplan, 2002).

2.3.3 Beta (13-30 Hz)

The Beta spectrum ranges from 13-30 HZ (Hu, Guo, Liu, & Wang, 2017). Beta wave activity is present when we are focused and in a state of alertness (Lal & Craig, 2001). If you are solving a complex problem then your brain is likely operating in the Beta spectrum.

2.3.4 Delta (1-4 Hz)

Delta waves are the slowest waves in the spectrum ranging from 1-4 HZ (Hu, Guo, Liu, & Wang, 2017). They occur during deep sleep and in rare physical conditions such as coma and vegetative states (Baars & Gage, 2007). Delta waves increase as we become more and more out of touch with the world (Baars & Gage, 2007).

2.3.5 Theta (4-8 Hz)

The Theta spectrum ranges from 4-8 HZ (Hu, Guo, Liu, & Wang, 2017). Theta waves have an unknown origin within the brain (Baars & Gage, 2007). They typically occur in states of deep relaxation such as meditation and during Rapid Eye Movement (REM) sleep (BrainWorks, 2011). They have also been known to appear in some short-term memory tasks such as memory retrieval (Baars & Gage, 2007).

2.3.6 Gamma (> 30 HZ)

The Gamma spectrum ranges from all waves greater than 30 HZ (Hu, Guo, Liu, & Wang, 2017). Gamma waves are the newest spectrum to be discovered and represent the highest frequencies in the spectrum. Gamma waves were once thought to be "spare brain noise", but have been linked to moments of high information processing (BrainWorks, 2011). It is thought that individuals with exceptional memory skills spend an exceptional amount of time in this spectrum (Transparent Coroporation, 2013).

2.4 Measuring EEG Power Spectrum Data

There are two means of measuring EEG frequency power spectrum data, absolute and relative power. Absolute power is measured as (μV^2 /Hz) for each frequency band

(Yuvaraj, et al., 2014). Relative power is measured as the ratio of the absolute power of some frequency band (using the mean frequency) over the total absolute power for all given frequency bands in question (Yuvaraj, et al., 2014). These distinctions are important when discussing EEG data.

2.5 Related Work

2.5.1 Eye Movement

Numerous research methods and studies regarding fatigue detection were conducted in the past. Electrooculography (EOG) was used as a research method for assessing the role of eye movement as it relates to fatigue. An EOG is a device similar to an EEG that measures the voltage associated with eye movement by placing a set of electrodes around the eye (Siddiqui & Shaikh, 2013). This has the possibility for effectiveness in fatigue detection because research has linked eye movements to fatigue. The transition from fast eye movement and normal blinking to no eye movement and a fast blinking rate are an occurrence in the fatigue process (Lal & Craig, 2002). Others have studied the duration of blinks and the number of blinks to determine their relationship to fatigue (Hsieh & Tai, 2013). However, eye movement based detection methods may not be ideal for fatigue detection because of the inconvenient placement of electrodes around the eye that can be aggravating and distracting to drivers. Other efforts have been made to use camera technology to study eye movement, but this technique is not reliable because it can be affected by sun glare and general lighting conditions within the vehicle (Hsieh & Tai, 2013).

2.5.2 Heart Rate

Heart rate has also been discussed as an indicator of fatigue. A decrease in heart rate usually occurs as driver fatigue begins to set in; however, research has also shown that there is a decrease in heart rate during prolonged driving (Lal & Craig, 2002). This would require further research into this field to determine if heart beats per minute would be a good indicator of fatigue. Other studies have shown a dramatic change in heart rate variability (time in between heart beats) accompany fatigue (Lal & Craig, 2002). However, more research must be done in this field to determine if it is a reliable indicator of fatigue.

2.5.3 Video Detection/Facial Recognition

Experiments into analyzing certain visual indicators of fatigue have been done. Facial tone, blinking, eyelid closure, yawning, and nodding have been studied as fatigue indicators (Sigari, Pourshahabi, Soryani, & Fathy, 2014). However, such systems can be impacted by features such as skin tone, eye color, and reflection (Veeraraghavan & Papanikolopoulos, 2001). Additionally, because these systems rely on camera technology they can be affected by vehicle lighting conditions. Because of the issues stated above, video detection/facial recognition methods may not be suitable for fatigue detection.

2.5.4 EEG Based Methods

Previous research has attempted to evaluate fatigue based on the relative power index of individual frequency bands. A decrease in relative Alpha power was observed during fatigue conditions (Naqvia , Badruddin, Malik, Hazabbah, & Abdullah, 2014). Another driver fatigue study supported this by showing an increase in Alpha and Theta waves and a decrease in Beta waves over a 90-minute driving session where 13 male students (aging from 22 to 27) were studied (Zhaoa , Zhaoa , Liua, & Zheng, 2012). These researchers utilized a driving simulator that consisted of a car frame with a steering wheel, gas/brake pedals, manual gear shift, a horn, and a turn signal. Students were screened based on a questionnaire that ensured they did not work night shifts, use prescription medications, have serious medical conditions that would affect the study such as concomitant disease, alcoholism, drug abuse, etc. Prior to the study, students were asked to refrain from consuming alcohol, caffeine, tea, food, or smoking approximately 12 hours before the study. Each session was conducted in a dimly lit, sound attenuated and temperature controlled laboratory between the hours of 9 a.m. and 11:30 a.m. or 3 p.m. and 5 p.m. Volunteers were provided necessary time to familiarize themselves with the simulator. Once a driver was able to drive for a 15-minute period without errors, they were permitted to begin the actual study.

At the beginning of the driving session, participants were asked to report if they felt fatigued or not. Afterward an ‘oddball’ task was given to the participant to ascertain their current alert level. As an oddball task, participants were instructed to respond to a stimulus that occurred randomly. This task consisted of a red and green circular image that appeared on the screen at random times. When either image was displayed, the participant was instructed to respond to the images using a mouse. Oddball tasks were given at the beginning and end of the session to ascertain the differences in the participant's reaction time. The driving session lasted 90 minutes. At the end of the session, the volunteers were asked to report their feelings of fatigue again so that it could be used in the data analysis.

The EEG power spectrum data were averaged into 5 minute epochs at the beginning and end of each participant's session, and were subjected to EEG Power Spectral Density analysis for the frontal, central, parietal, occipital, and temporal regions of the brain. Additionally, the relative power of the Alpha, Beta, Theta, and Delta spectrums were compared for the same epochs to look for statistically significant changes in the rhythms. They found that Alpha and Theta waves increased and Beta decreased when the epochs were compared.

While the study by Zhaoa reaffirms that Alpha and Theta waves increase and Beta decreases as driver fatigue occurs, there has been evidence showing that only subtle changes may be observed during mild fatigue by utilizing absolute and relative power (Eoh, Chung , & Kim, 2005) (Kar , Bhagat, & Routray, 2010). This presents a difficult problem when trying to implement a detection algorithm. To counter this, researches have turned to analyzing different ratios of frequency spectrum data to enhance the contrast between different levels of fatigue (Kar , Bhagat, & Routray, 2010). This may give additional insight into the finer granularity of changes in brain waves that might be helpful in detecting fatigue.

2.5.5 EEG Ratio Based Indexes

There has been considerable research into analyzing the ratio between slow moving (Alpha and Theta) to fast moving waves (Beta) as an indicator of fatigue (Jap, Lal, Fischer, & Bekiaris, 2009). Jap , Lal, Fischer, & Bekiaris (2009) performed their study on 52 non-professional drivers (36 males and 16 females) between the ages of 20-70. A lifestyle appraisal questionnaire was used to screen participants to ensure they had no medical constraints such as severe concomitant disease, alcoholism, drug abuse, and psychological or intellectual problems. Prior to the study, participants were instructed to refrain from consuming caffeine, tea, or food as well as smoking approximately 4 hours before the study, and to avoid consuming alcohol 24 hours before the study. The study was conducted in a temperature controlled environment at approximately noon for each participant. The video game “Grand Prix 2” in conjunction with a car frame, a built in steering wheel, brakes, gas pedal, and gears were used as the simulation method.

Participants were asked to complete two driving sessions. The first session, the “alert driving” session, lasted approximately 10-15 minutes. This session was intended to provide a baseline for the ‘drowsy’ driving session. During these sessions, participants were provided a track with multiple cars and multiple stimuli on the road. Following this, the participants were asked to drive at a continuous speed of 60-80 km/h for one hour with minimal stimuli.

A 30 channel EEG was used to collect brain wave data. A video camera was also used to record the participant's face to identify physical signs of fatigue. After the study, the EEG raw data were sectioned into one second epochs and then were subjected to fast Fourier transformation (FFT) to derive the raw EEG data into four frequency components; Delta, Theta, Alpha, and Beta. Next, the frequency spectrum data were segmented into 3 consecutive time intervals for the alert driving session and Base Ratios 1, 2, 3, and 4 (See Equations 1, 2, 3, and 4 below) were calculated for each interval and then averaged together to determine an alert baseline.

Similarly, the monotonous driving session for each participant was divided into 10 sections. For each section, Base Ratios 1, 2, and 3 were calculated every 10 seconds and then averaged to obtain one value for each section. The data from each of the 10 sections were then compared to the baseline for each subject.

Analysis of variance (ANOVA) was performed to identify significant differences between the 10 time points during the monotonous driving session and the alert baseline for Base Ratios 1, 2, 3, and 4 listed below. This was done for five brain sites (central, frontal, occipital, parietal, and temporal). Results were reported as mean ± standard deviation (SD). Significant level was reported as $p < 0.05$.

$$Base\ Ratio1 = \frac{Alpha\ \ Waves + \ Theta\ Waves}{Beta\ Waves}$$

Equation 1. Base Ratio 1

$$Base\ Ratio\ 2 = \frac{Alpha\ \ Waves}{Beta\ Waves}$$

Equation 2. Base Ratio 2

$$Base\ Ratio\ 3 = \frac{Alpha\ \ Waves\ \ + Theta\ Waves}{Alpha\ Waves\ + \ Beta\ Waves}$$

Equation 3. Base Ratio 3

$$Base\ Ratio\ 4 = \frac{Theta\ Waves}{Beta\ Waves}$$

Equation 4. Base Ratio 4

The authors determined that Base Ratio 1 had the greatest range of change between sessions and thus may be a good indicator of fatigue. This work was extended by (Punsawad, Aempedchr, Wongsawat, & Panichkun) where the authors assigned a weight to each frequency such as shown in Base Ratios 5, 6, and 7:

$$Base\ Ratio\ 5 = \frac{0.5Alpha\ Waves + 0.5Theta\ Waves}{0.5Beta\ Waves}$$

Equation 5. Base Ratio 5

$$Base\ Ratio\ 6 = \frac{0.6\ Alpha\ Waves\ + 0.4 Theta\ Waves}{0.5\ Beta\ Waves}$$

Equation 6. Base Ratio 6

$$Base\ Ratio\ 7 = \frac{0.4 Alpha\ Waves + 0.6 Theta\ Waves}{0.5 Beta\ Waves}$$

Equation 7. Base Ratio 7

This system was tested on seven participants. No screening criteria were applied for this study. The simulation method used was a PlayStation 2 game called "Short Track Racing" in conjunction with a steering wheel, and gas and brake pedals. The researchers used a threshold method to trigger alarms during this study. The threshold was set by closing their eyes for 10 seconds prior to the session and calculating the ratios in question for this study. 50 percent of this value was used to set the threshold for each participant. If the participant fell below this level, they triggered an alarm.

The actual driving session lasted one hour. To reduce false positives, the researchers chose to not to start the detection system for the first half hour. The remaining 30 minutes were then subdivided into three 10 minute sections. Validation of the system was done by allowing the participants to self-report fatigue by pressing a button on the screen. After the study, the number of alarms for the given three sections were compared to the number of participant fatigue reports to determine if fatigue had been sensed for that given time period.

Base Ratio 7 proved to have the highest level of accuracy, but it also did not capture as many alarms as participant fatigue indications. While on the bright side, this may mean there are less false positives, it could potentially be a problem in a safety critical system. It is noteworthy to mention that there is no known comparative study between the weighted ratios and the work performed by (Jap, Lal, Fischer, & Bekiaris, 2009).

Other work was done by (Ming-ai , Cheng , & Jin-Fu, 2010) where Base Ratio 8 (see Equation 8) was studied over an approximately four hour session.

$$Base\ Ratio\ 8 = \frac{Delta\ Waves + Theta\ Waves}{Alpha\ Waves + Beta\ Waves}$$

Equation 8. Base Ratio 8

This work proved to show a continual decline in the fatigue index as the 'drowsy' session continued. The data in this study were recorded using a SYMTOP NT-9200 Dynamic 16 channel EEG. The exact method for driving simulation is not listed in the paper, but appears to be a computer game with a steering wheel. No pre-study instructions were listed and the paper does not mention the exact number of participants.

In that study, they calculated a baseline index by recording EEG data in one minute intervals during the sober driving session. Afterward EEG data were collected for a four hour 'drowsy' driving session (again samples were collected in one minute epochs). During post-processing, the raw EEG data were subjected to Independent Component Analysis (ICA) to separate the different EEG channels. Following this FFT was performed to parse out the frequency spectrums. Next, the authors calculated the Power Spectrum Density of each frequency spectrum. Finally, the authors plugged these numbers into their fatigue index calculation. In comparison to the baseline, the "drowsy session" indexes were significantly higher giving the authors reason to believe that this index method is a reliable method for fatigue detection.

2.6 The need for a one channel EEG

Many of the EEG research mentioned above has utilized multi-channel EEG devices. Multi-channel devices are often used in clinical research settings because of their superior data quality. (Ratti, Waninger, Berka, Ruffini, & Verma, 2017) demonstrated that data quality was reduced when using single channel EEGs vs multi-channel EEGs. That study attributed this due to the sensor placement of single-channel devices on the frontal lobe, which are prone to noise generated from eye blinks. However, that same study noted that power metrics for the Neurosky Mindwave, a single channel device, were very similar to the multi-channel devices used for comparison in the study (Neurosky, 2015). That research was supported by (Hal, Rhodes, & Dunne, 2014) who detected drowsiness 81% of the time using the Neurosky Mindwave device.

It is also important to bring up the question of practicality and usability when considering EEG devices. Even a perfect detection algorithm will fail if the user perceives the detection method to be intrusive or uncomfortable. While multi-channel EEGs may provide additional granularity and data quality over single-channel devices, they present implementation problems in the real world. Imagine the difficulty of trying to drive cross-country wearing a device with 16 sensors. Additionally, if the algorithms are designed to consider all 16 channels, the loss of contact with one could give inaccurate results. This would mean that the driver would need to maintain constant contact with all 16 sensors. Such a system does not seem feasible. However, a sensor for a simple one channel EEG could easily be placed upon the forehead of a driver. This could even potentially be embedded in a hardhat which many industry professionals are required to wear on the job already. This is easier to implement than a multi-channel EEG that would require the user to keep in contact with multiple electrodes. In this sense, simpler is better and a one channel EEG would be more effective.

Chapter 3

RESEARCH METHODS

3.1 Research Statement

BCI devices have been used to study variability in the absolute and relative power spectrums of Alpha, Beta, and Theta waves to determine the best way to detect mental fatigue. Research has shown that ratio-based approaches are a good choice for detecting changes in fatigue due to their ability to detect minor levels of change in fatigue. While ratio-based approaches have shown promise, a critical gap exist in the research. To our knowledge, there has been no empirical research conducted to identify at what point an alarm should be issued using any of the fatigue detection ratios listed in the Section 2.4.5. If these approaches are to ever be used in a real-time fatigue detection system, a fatigue detection method needs to be established to determine when an alarm should be issued.

The purpose of this thesis is to identify if Percent Change from an alert state to a fatigue state is common across the participant base for any of the ratios stated in Section 2.4.5. For this purpose, data were be collected for a period of one hour, via an EEG, to capture the participant's brainwaves while they simulate driving. Every five minutes the participant were asked to state their level of fatigue based on the Stanford Sleepiness Scale (Stanford Sleepiness Scale, 2017) as shown in Table 1. The Stanford Sleepiness

Scale has been used by various researchers and clinicians with adult population to gather self-reported data on sleepiness.

After the data collection, an average of the first 15 minutes were taken to establish a baseline. Following this a Percent Change analysis was performed between the baseline and the point at which the participant indicates they are fatigued for all ratios. This percentage was compared across the participant population to determine if there is a trend between the baseline and the point of which participants identify themselves as fatigued. If a common Percent Change is discovered across the participant base, then it would indicate that Percent Change from a baseline alert state to a fatigue state could be a potential detection mechanism for which an alarm could be issued in a real-time fatigue detection system.

3.2 Neurosky Mindwave Headset

The EEG device used to collect data in the study is the Neurosky Mindwave Mobile (Neurosky, 2015). The device uses one dry electrode that is placed on the center of the forehead. The device utilizes a technology called Thinkgear. Thinkgear is a microchip that pre-processes the EEG signal and transmits the data over Bluetooth (Neurosky, 08). The EEG raw data that is passed from the Neurosky Mindwave device is simply a measurement of the electrical potential that is occurring within the brain at a given time. The device samples this activity at a rate of 512 HZ (Neurosky, 2014).

Raw data is measured in the time domain. To utilize this data in a way that is comparable to the other studies mentioned, the data must be transformed into the frequency domain. This is typically done using FFT. From a computational standpoint, performing FFT in a real-time system could be considered expensive and would require a very efficient algorithm to accommodate such overhead. Luckily, the Mindwave device is already optimized to do this via the Thinkgear technology that comes with the device. This allows the frequency domain data to be accessed via an API.

Frequency domain data is returned as a measure of power for each frequency band. It is important to note that this measure of power is done via a proprietary algorithm. Therefore, the details of how it is calculated are not provided. Additionally, these values have custom developed units instead of conventional units such as microvolts. Therefore, they are only meaningful for comparison with other values of the same data type. Despite not generating conventional units, EEG data produced by Neurosky Mindwave is relevant and appropriate for performing Percent Change analysis between subjects. The frequencies represented through this API are; Delta (0.5 - 2.75Hz), Theta (3.5 - 6.75Hz), Low-Alpha (7.5 - 9.25Hz), High-Alpha (10 - 11.75Hz), Low-Beta (13 - 16.75Hz), High-Beta (18 - 29.75Hz), Low-Gamma (31 - 39.75Hz), and Mid-Gamma (41 - 49.75Hz).

3.3 Driving Simulation

The purpose of this thesis is to identify if Percent Change from an alert state to a fatigue state is common across the participant base for any of the ratios stated in Section 2.4.5.

To do this, a realistic driving simulation must be setup to simulate driving conditions. We use STISIM driving simulator software build 2.06.00 (Systems Technology, Inc., 2016) to perform the simulation. This driving simulator was used to simulate night time driving with two-lanes. Periodically on-road obstacles such as another vehicle on the road and infrequent curvy roads were displayed. The simulator may show some off-road attractions such as a tree, shopping complex, and parked cars as a part of the driving simulation. Subjects were able to drive using steering wheel and gas & brake pedals.

3.4 Participant Population

The participant population consists of UNF students who are over the age of 18. All interested student volunteers who are 18 years or older were accepted regardless of their gender, race, or other demographic aspects. We contacted UNF professors via email and requested them to encourage students from their classes to participate voluntarily in the study. See Appendix A for the sample email script. Professors were asked to give a link to any students interested in which they can sign up to participate in this study. The link also contained instructions on where the study was held. To facilitate the sign-up process, we used the website Slotted.co to allow participants to select a date and time convenient to their schedule. See Appendix B for details of instructions that were provided to participants upon arrival to the lab where the experiment took place.

3.5 Factors Influencing Experiment

Fatigue has the potential to be affected by multiple factors: caffeine, prior night's sleep, time of day, hours worked among others. To control some of these influences a list of instructions was sent to the participant prior to the study and a controlled laboratory was setup for the experiment. Participants were asked to refrain from using caffeine 12 hours prior to the study and alcohol 24 hours prior to the study. The study took place in a UNF laboratory at approximately 6 p.m. During the data collection, participants were asked to wear a Neurosky EEG headset. While wearing the device, participants drove for approximately one hour via a driving simulator.

3.6 Data Collection

While the participant is driving, they were asked to wear the Neurosky Mindwave EEG device. To facilitate our data collection, we developed a program using Java to read EEG data produced by Neurosky Mindwave device. This Java program that ran in the background synced with the Mindwave device and continuously read packets via the Mindwave API over Bluetooth. Both EEG Power for each frequency were captured as well as the raw data. Both data types were written to a log file for persistence purposes so that the data can be further analyzed post-simulation.

Degree of Sleepiness	Scale Rating
Feeling active, vital, alert, or wide awake	1
Functioning at high levels, but not at peak; able to concentrate	2
Awake, but relaxed; responsive but not fully alert	3
Somewhat foggy, let down	4
Foggy; losing interest in remaining awake; slowed down	5
Sleepy, woozy, fighting sleep; prefer to lie down	6
No longer fighting sleep, sleep onset soon; having dream-like thoughts	7
Asleep	X

Table 1. Stanford Sleepiness Scale

Additionally, to assess the participant's level of fatigue, the driver was asked to state their current level of fatigue every five minutes based on the Stanford Sleepiness Scale. This data were manually recorded via an Excel spreadsheet by the study administrator. This self-reported data were used to determine the driver's Percent Change from the alert baseline to their fatigue state during post-simulation analysis.

3.7 Data Analysis

The three ratios that were identified as being the most successful in Section 2.4.5 were Base Ratios 1,7, and 8. It's important to re-iterate that the Neurosky API provides a coarser classification of certain wave bands (noted in Section 3.2) than the studies identified in the research. For example, the Alpha spectrum is sub-classified into Low Alpha and High Alpha. This provides an opportunity to study these ratios at a very granular level. To do such, the following ratios were identified from their original ratios. Base Ratio 1 was sub-divided into the following ratios:

$$Ratio\ 1 = \frac{Low\ Alpha\ \ Waves\ \ + Theta\ Waves}{Low\ Beta\ Waves}$$

Equation 9. Ratio 1

$$Ratio\ 2 = \frac{High\ Alpha\ \ Waves\ \ + Theta\ Waves}{High\ Beta\ Waves}$$

Equation 10. Ratio 2

$$Ratio\ 3 = \frac{High\ Alpha\ \ Waves\ \ + Theta\ Waves}{Low\ Beta\ Waves}$$

Equation 11. Ratio 3

$$Ratio\ 4 = \frac{Low\ Alpha\ \ waves\ \ + Theta\ Waves}{High\ Beta\ waves}$$

Equation 12. Ratio 4

$$Ratio\ 5 = \frac{(Low\ Alpha\ \ Waves + High\ Alpha\ Waves) + Theta\ waves}{(High\ Beta\ Waves + Low\ Beta\ Waves)}$$

Equation 13. Ratio 5

Base Ratio 7 (see Equation 7) was sub-divided into the following ratios:

$$Ratio\ 6 = \frac{0.6 Theta\ Waves + 0.4 LowAlpha\ Waves}{0.5 Low\ Beta\ Waves}$$

Equation 14. Ratio 6

$$Ratio\ 7 = \frac{0.6\ Theta\ Waves + 0.4 High\ Alpha\ Waves}{0.5 High\ Beta\ Waves}$$

Equation 15. Ratio 7

$$Ratio\ 8 = \frac{0.6 Theta\ Waves + 0.4 Low\ Alpha\ Waves}{0.5 High\ Beta\ Waves}$$

Equation 16. Ratio 8

$$Ratio\ 9 = \frac{0.6 Theta\ Waves + 0.4 High\ Alpha\ Waves}{0.5 Low\ Beta\ Waves}$$

Equation 17. Ratio 9

$$Ratio\ 10 = \frac{0.6 Theta\ Waves + (0.4 Low\ Alpha\ Waves +\ 0.4 High\ Alpha\ Waves)}{(0.5 Low\ Beta\ Waves +\ 0.5 High\ Beta\ Waves)}$$

Equation 18. Ratio 10

Base Ratio 8 (see Equation 8) was sub-divided into the following ratios:

$$Ratio\ 11 = \frac{Delta\ Waves + Theta\ Waves}{Low\ Alpha\ Waves + Low\ Beta\ Waves}$$

Equation 19. Ratio 11

$$Ratio\ 12 = \frac{Delta\ Waves + Theta\ Waves}{High\ Alpha\ Waves + High\ Beta\ Waves}$$

Equation 20. Ratio 12

$$Ratio\ 13 = \frac{Delta\ Waves + Theta\ Waves}{High\ Alpha\ Waves + Low\ Beta\ Waves}$$

Equation 21. Ratio 13

$$Ratio\ 14 = \frac{Delta\ Waves + Theta\ Waves}{Low\ Alpha\ Waves + High\ Beta\ Waves}$$

Equation 22. Ratio 14

$$Ratio\ 15 = \frac{Delta\ Waves + Theta\ Waves}{(Low\ Alpha\ Waves + High\ Alpha\ Waves) + (Low\ Beta\ Waves + High\ Beta\ Waves)}$$

Equation 23. Ratio 15

3.8 Institutional Review Board (IRB) Approval

As this thesis study utilizes human subjects as a part of the experiment, Institutional Review Board (IRB) approval was necessary. We submitted necessary research proposal and experimental materials to obtain UNF IRB approval. The IRB project was approved on Nov. 1, 2016 and reference number is 475514-4. See Appendix C for IRB approval letter and Appendix D for informed consent form that was provided to the participants.

Chapter 4

SYSTEM DESIGN

4.1 System Overview

To perform our study, a method of data retrieval was needed from the Neurosky Mindwave device. To do this a Java program was developed to interact with the Mindwave device by reading packets over Bluetooth. This program was executed for each session while the participant drove for a period of one hour. A persistence solution was also needed to store the data as packets are read. Because the Java program is constantly reading packets from the device, a database storage solution was not appropriate due to latency issues. Instead the data were written to a log file to reduce the overhead.

This process worked well for retrieving and storing the data. However, the data needed to be placed in a database so that queries could easily be run to perform analysis. To do that a second Java program was written to parse the data from the log file, and store it in an Oracle database.

Once the data were placed in the initial staging table, the actual data analysis (i.e., Percent Change analysis) could begin. First the data were averaged into five-minute time intervals (accomplished by simply averaging every sample within each five minute period). This approach was similar to the method used by (Jap, Lal, Fischer, & Bekiaris, 2009). Next, a baseline was calculated by averaging the first 15 minutes (the first 3 five-minute time intervals) of the data for each ratio of interest for the study. This was done to establish a baseline during the period when the driver should have been most awake. Again, this is approach was similar to the method used by (Jap, Lal, Fischer, & Bekiaris, 2009). After that, we identified the point at which the drivers claimed that they began to feel fatigue (this is when the driver stated that they were at level 5 on the Stanford Sleepiness Scale). At this point we analyzed what each ratio was at that moment in time. The Percent Change analysis was then performed for each ratio, where the Percent Change was measured from the baseline to the moment in which drivers began to feel fatigue. Following this, the results were compared across the participant base to look for a trend that might be useful in establishing a threshold for which an alarm could be rendered in a real-time fatigue detection system. This is explained in greater detail in Chapter 6.

4.2 EEG Data Reader

4.2.1 ThinkGear Communications Driver (TGCD)

The Java program communicated with the Neurosky Mindwave via the ThinkGear Communications driver (TGCD). This driver provides an API that allows for communications between the ThinkGear chip and the application (NeuroSky, 2014). The TGCD API is accessed via a Dynamic-Link Library (DLL). Because DLLs are typically used for C/C++/C#, a Java Native Interface (JNI) was used. JNIs are programming devices that allow code that is written in another language to be executed in Java (Oracle, 2016).

4.2.2 Pre-Study Validation

A calibration procedure did not exist for the Neurosky device. To ensure that accurate readings were being received, a series of tasks were given to the participants with expected outcomes from the EEG data being returned from the device. The first task was for the participants to close their eyes for 10 seconds. The expected result was to watch the power for Low-Alpha and High-Alpha waves to increase. For the second task the students were asked to solve a series of mathematical equations for a period of one minute. Because the participant was expected to transition to a state of concentration while doing this, the expected result was for Low-Beta and High-Beta power to increase while Low-Alpha and High-Alpha decreased. This validation was conducted for all 10 of the study participants prior to the one hour driving session to determine if the device was functioning properly.

4.2.3 Java Data Reader

The Java program that reads data from the Neurosky device operated on Port COM21. Communication was established in the API by calling the ThinkGear.Connect method (see Figure 1). After establishing a connection, the application requested to read packets from the device over Bluetooth at a rate of 57600 BAUD (this is the default setting).

```
private static int getConnection(String comPortName) {
        int connectionId = ThinkGear.GetNewConnectionId();

        if(connectionId < 0){
            System.out.println("ERROR.");
            System.exit(-1);
        }

        int errCode = -1;
        do{
             errCode = ThinkGear.Connect(connectionId, comPortName,
                       ThinkGear.BAUD_57600, ThinkGear.STREAM_PACKETS);
             if( errCode < 0 ) {
                 System.out.println("Connecting...");
             }

        }while(errCode < 0);

        System.out.println("Connection established.  Connection id: " + connectionId);
        return connectionId;
    }
```

Figure 1. Get Connection Method

As the application began to read data, the packets were inspected for a specific type of value. This was done by calling the ThinkGear.GetValueStatus method for each specific data type ex: Low Alpha, High Alpha, Beta etc. If the program determined that the value was present, it then extracted the value by calling ThinkGear.GetValueStatus (see Figure 2) for the data type.

```
if(ThinkGear.GetValueStatus(connectionId, ThinkGear.DATA_RAW) != 0){
    double power = ThinkGear.GetValue(connectionId, ThinkGear.DATA_RAW);
    logInsert(userID,"Raw",power,currentTimestamp);
}
```

Figure 2. Get Value Status Method

Once data were extracted the program wrote to a log file (see Figure 3 for an example). This process ensured that no packets were missed due to the latency associated with writing to a database. The program continued to read packets until the session was complete. After the data were collected for the study, a second program ran to parse the data from the log file and place them in the database so that the data analysis could be performed.

```
User: 27 Type: Raw Value: 229.0 Timestamp: 2016-12-11 17:32:19.921
User: 27 Type: Low_Alpha Value: 4840.0 Timestamp: 2016-12-11 17:32:19.941
User: 27 Type: High_Alpha Value: 5286.0 Timestamp: 2016-12-11 17:32:19.941
User: 27 Type: Low_Beta Value: 6271.0 Timestamp: 2016-12-11 17:32:19.941
User: 27 Type: High_Beta Value: 17842.0 Timestamp: 2016-12-11 17:32:19.941
User: 27 Type: Delta Value: 54919.0 Timestamp: 2016-12-11 17:32:19.941
User: 27 Type: Theta Value: 51932.0 Timestamp: 2016-12-11 17:32:19.941
```

Figure 3. Log File Sample

4.3 Parsing the EEG data from the Log File and inputting into database

After the data were collected, a second program parsed out the data from the log file and placed them inside an Oracle database. This program simply read the file line by line and tokenized the line using the Java split method (see Figure 4). Once the data had been parsed, they were persisted in an Oracle database (Oracle 11g Express Edition) in a table called "EEG_OUTPUT" (see Table 2). The table structure is listed below in Table 3.

```java
package FileParser;

import java.io.BufferedReader;
import java.io.FileReader;
import java.io.IOException;
import java.sql.Timestamp;


import DAO.BaseDAO;


public class ParserDriver {

    public static void main(String[] args) {
        //Log//UserID_25_.txt
        String fileName = args[0];

        BufferedReader br = null;
        FileReader fr = null;

        try {

            fr = new FileReader(fileName);
            br = new BufferedReader(fr);

            String sCurrentLine;

            br = new BufferedReader(new FileReader(fileName));

            while ((sCurrentLine = br.readLine()) != null) {
                String [] line = sCurrentLine.split("\\s+");
                int user = Integer.parseInt(line[1]);
                String type = line[3];
                Double value = Double.parseDouble(line[5]);
                String[] dateSplit = line[7].split("-");
                String[] timeSplit = line[8].split(":");
                String[] milliSplit = timeSplit[2].split("\\.");

                Timestamp ts = new Timestamp(Integer.parseInt(dateSplit[0]),
                Integer.parseInt(dateSplit[1]), Integer.parseInt(dateSplit[2]),

                Integer.parseInt(timeSplit[0]), Integer.parseInt(timeSplit[1]),
                Integer.parseInt(milliSplit[0]),Integer.parseInt(milliSplit[1]));

                BaseDAO.insertLogRecord(user, type, ts, value);

            }

        } catch (IOException e) {

            e.printStackTrace();

        } finally {

            try {

                if (br != null)
                    br.close();

                if (fr != null)
                    fr.close();

            } catch (IOException ex) {

                ex.printStackTrace();

            }
        }
    }
```

Figure 4. Parse Driver

COLLUMN_NAME	DATA_TYPE	NULLALE	DATA_DEFAULT	COLLUMN_ID
USER_ID	NUMBER	No	(NULL)	1
DATA_TYPE	VARCHAR(20 BYTE)	No	(NULL)	2
TIME	TIMESTAMP(6)	No	(NULL)	3
VALUE	NUMBER(20,9)	No	(NULL)	4

Table 2. EEG_OUTPUT Table Structure

CHAPTER 5

POST STUDY ANALYSIS

After the data were collected from each study, a secondary Java program ran to parse the data from the log files and place them into the database. After this, both the EEG data and the participant fatigue reports needed to be appropriately transformed so that the data analysis could be conducted. The participant fatigue reports were captured via an excel spreadsheet and were manually loaded into the database table FATIGUE_REPORT (see Table 3).

However, the EEG data required additional extraction, loading, and transformation operations to arrive at the result set desired (the Percent Change for each ratio for each participant). To accomplish this a set of Oracle stored procedures were created in the package EEG_LOAD (see Figure 5) to automate many of these tasks. The flow of the Extraction Transform Load (ETL) process can be seen in Figure 5.

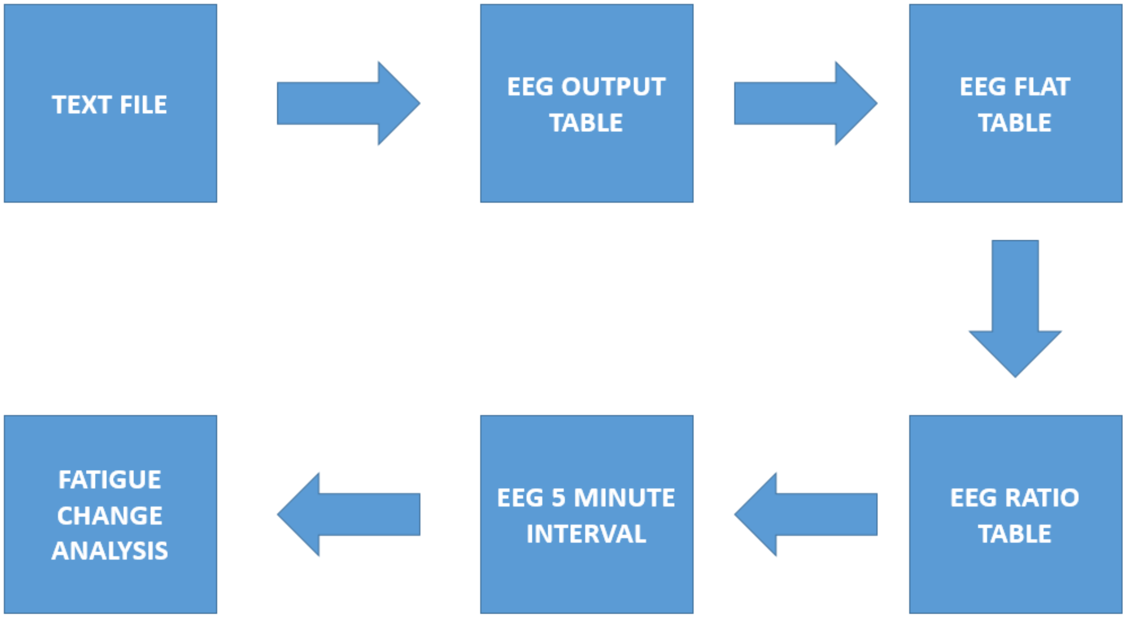

Figure 5. ETL Flow

COLUMN_NAME	DATA_TYPE	NULLALE	COLUMN_ID
USER_ID	NUMBER	No	1
TIME_INTERVAL	NUMBER	No	2
FATIGUE_LEVEL	NUMBER	No	3

Table 3. FATIGUE_REPORT Table Structure

```
create or replace
PACKAGE BODY EEG_LOAD AS

  PROCEDURE loadEEGData(UserId IN    VARCHAR2)
  IS
  BEGIN
    parseData(UserId);
    createRatios(UserId);
    createTimeIntervals(UserId);

  END;

  PROCEDURE parseData(UserId IN    VARCHAR2)
  IS

  BEGIN

     /*
      *Pivots the EEG_OUTPUT table to flaten out the data
      */
      INSERT INTO EEG_FLAT_TABLE
      (SELECT USER_ID,ROUND(Low_Alpha,3) Low_Alpha,ROUND(High_Alpha,3) High_Alpha,ROUND(Low_Beta,3) Low_Beta,ROUND(High_Beta,3) High_Beta,
      ROUND(Delta,3) Delta,ROUND(Theta,3) Theta,DATE_TIME
      FROM
        (SELECT USER_ID,DATA_TYPE,"VALUE","TIME" DATE_TIME
         FROM EEG_OUTPUT
         WHERE DATA_TYPE != 'Raw'
         AND USER_ID = UserId)
      PIVOT(AVG("VALUE")  FOR (DATA_TYPE) IN('High_Beta' AS HIGH_BETA,'Low_Beta' AS Low_Beta,'High_Alpha' AS HIGH_ALPHA,
                                             'Low_Alpha' AS LOW_ALPHA,'Theta' AS THETA,'Delta' AS DELTA)));
      COMMIT;

  END;

  PROCEDURE createRatios(UserId IN    VARCHAR2)
  IS

  BEGIN

     INSERT INTO EEG_RATIO
      select
      USER_ID,
      ROUND(((Low_Alpha + Theta) /Low_Beta)  ,2) RATIO_1,
      ROUND((High_Alpha + Theta )/(High_Beta),2) RATIO_2,
      ROUND((High_Alpha + Theta )/(Low_Beta) ,2) RATIO_3,
      ROUND((Low_Alpha  + Theta )/(High_Beta),2) RATIO_4,
      ROUND(((Low_Alpha + High_Alpha) + Theta)/((High_Beta +Low_Beta)) ,2) RATIO_5,
      ROUND(((0.6 * Theta) + (0.4 * Low_Alpha)) /(0.5 * Low_Beta )     ,2) RATIO_6,
      ROUND(((0.6 * Theta) + (0.4 * High_Alpha)) /(0.5 * High_Beta )   ,2) RATIO_7,
      ROUND(((0.6 * Theta) + (0.4 * Low_Alpha))  /(0.5 * High_Beta )   ,2) RATIO_8,
```

Figure 6. EEG Body Package

5.1 Extraction, Loading, and Transformation of EEG Data

The output of the Parser program placed the results in the EEG_OUTPUT table (see section 4.3). The EEG_OUTPUT table was designed with flexibility in mind to store a variety of EEG data types provided by Neurosky. In this case, a new value could be added to the DATA_TYPE column and, as long as the attribute value was a number, the table could easily consume the data. This is known as a horizontal table or an Entity-Attribute-Value (EAV) design (Dinu & Nadkarni, 2007). Alternatively, a flat structure could have been created but would have potentially resulted in null columns which would have been an inefficient use of the table.

That being stated, an EAV design does not lend itself well to data analysis because of the added complexity of searching for specific values in the DATA_TYPE column to perform the analysis on. To accommodate that a flat table structure, EEG_FLAT_TABLE (see Table 4), was created from the EEG_OUTPUT table using only the EEG attributes of interest for this study. To create the flattened out structure the data from EEG_OUTPUT was pivoted using the Oracle Pivot clause in the stored procedure named parseData in the EEG_LOAD package (see Figure 5). The Pivot clause aggregates data and allows for the results to be rotated from rows into columns. The result is one row per sample for all brainwave values at a given point in time per user (see Table 5).

COLUMN_NAME	DATA_TYPE	NULLALE	COLUMN_ID
USER_ID	NUMBER	No	1
LOW_ALPHA	NUMBER(20,9)	Yes	2
HIGH_ALPHA	NUMBER(20,9)	Yes	3
LOW_BETA	NUMBER(20,9)	Yes	4
HIGH_BETA	NUMBER(20,9)	Yes	5
DELTA	NUMBER(20,9)	Yes	6
THETA	NUMBER(20,9)	Yes	7
DATE_TIME	TIMESTAMP(6)	Yes	8

Table 4. EEG_FLAT_TABLE Structure

USER_ID	LOW_ALPHA	HIGH_ALPHA	LOW_BETA	HIGH_BETA	DELTA	THETA	DATE_TIME
1	6520	3655	15406	6677	18541	23112	03-APR-17 06.26.30 PM

Table 5. EEG_FLAT_TABLE EXAMPLE

Once the data had been placed into EEG_FLAT_TABLE the ratio calculations needed to be performed for each of the 15 ratios. This was done in the stored procedure createRatios (see Figure 5) and persisted in the EEG_RATIO table (see Table 6). The result of this is one record for each ratio, per participant, per sample.

COLLUMN_NAME	DATA_TYPE	NULLALE	COLLUMN_ID
USER_ID	NUMBER	No	1
RATIO_1	NUMBER(20,9)	No	2
RATIO_2	NUMBER(20,9)	No	3
RATIO_3	NUMBER(20,9)	No	4
RATIO_4	NUMBER(20,9)	No	5
RATIO_5	NUMBER(20,9)	No	6
RATIO_6	NUMBER(20,9)	No	7
RATIO_7	NUMBER(20,9)	No	8
RATIO_8	NUMBER(20,9)	No	9
RATIO_9	NUMBER(20,9)	No	10
RATIO_10	NUMBER(20,9)	No	11
RATIO_11	NUMBER(20,9)	No	12
RATIO_12	NUMBER(20,9)	No	13
RATIO_13	NUMBER(20,9)	No	14
RATIO_14	NUMBER(20,9)	No	15
RATIO_15	NUMBER(20,9)	No	16
DATE_TIME	TIMESTAMP(6)	No	17

Table 6. EEG_RATIO

Finally, the samples were averaged into 5 minute intervals in the createTimeIntervals stored procedure and placed in the EEG_5_MIN_INTERVAL table (see Table 7). This approach was similar to the method used by (Jap, Lal, Fischer, & Bekiaris, 2009). These 5 minute intervals were used during the data analysis to determine how a user's brainwaves change from the alert state (the baseline) to the fatigue state. The method used in the stored procedure to accomplish this requires additional explanation. To perform the time math on this data, the timestamp for each sample was turned into the exact minute of the day via the following snippet of SQL:

```
60 * extract (hour from DATE_TIME) + extract (minute from DATE_TIME)
```

Once the precise minute of the sample was calculated, it was then divided by 5 to group the data into 5 minute intervals. Next, the time intervals were ordered ascendingly and a numeric time interval was assigned to each one use the ROWNUM() function. This enabled an easy matchup to the participant fatigue reports which were also assigned a numeric value for each 5-minute time interval.

COLLUMN_NAME	DATA_TYPE	NULLALE	COLLUMN_ID
USER_ID	NUMBER	No	1
RATIO_1	NUMBER(20,9)	No	2
RATIO_2	NUMBER(20,9)	No	3
RATIO_3	NUMBER(20,9)	No	4
RATIO_4	NUMBER(20,9)	No	5
RATIO_5	NUMBER(20,9)	No	6
RATIO_6	NUMBER(20,9)	No	7
RATIO_7	NUMBER(20,9)	No	8
RATIO_8	NUMBER(20,9)	No	9
RATIO_9	NUMBER(20,9)	No	10
RATIO_10	NUMBER(20,9)	No	11
RATIO_11	NUMBER(20,9)	No	12
RATIO_12	NUMBER(20,9)	No	13
RATIO_13	NUMBER(20,9)	No	14
RATIO_14	NUMBER(20,9)	No	15
RATIO_15	NUMBER(20,9)	No	16
DATE_TIME	TIMESTAMP(6)	No	17

Table 7. EEG_5_MIN_INTERVAL

After the ratios were averaged into their 5 minute intervals the Percent Change from the participant's baseline was calculated.

5.2 Percent Change Analysis

After the data were transformed into the appropriate format, the Percent Change analysis was performed. To identify the Percent Change from an alert state to a fatigue state, the baseline needed to be established. To do so, an average for each of the first three 5 minute intervals were calculated for each of 15 ratios mentioned in Section 3.7. The result for each ratio served as the baseline for each Percent Change analysis to be calculated (one for each of the 15 ratios).

Once the baseline was established the data taken from the participant's fatigue reports were used to identify the point in time when the participant reported five (the point where a participant begins to lose interest in staying awake) on the Stanford Sleepiness scale. Because the participants reported their fatigue level in five minute increments, the data for the 5-minute time-period in question were averaged for each ratio. Once this point was identified, the Percent Change could easily be calculated using the formula in Equation 24 (Kaplan, 2016).

$$Percent\ Change = \frac{Amount\ of\ Increase\ or\ Decrease\ from\ Baseline}{Baseline} * 100$$

Equation 24. Percent Change

The SQL used to perform the Percent Change analysis was encapsulated in two views V_FATIGUE_CHANGE_ANALYSIS and V_FATIGUE_PERIOD. The SQL used to create the views can be found in Figures 7 and 8.

```
CREATE OR REPLACE FORCE VIEW "SYSTEM"."V_FATIGUE_CHANGE_ANALYSIS"
("USER_ID", "RATIO_1_CHANGE", "RATIO_2_CHANGE", "RATIO_3_CHANGE",
"RATIO_4_CHANGE", "RATIO_5_CHANGE", "RATIO_6_CHANGE",
"RATIO_7_CHANGE", "RATIO_8_CHANGE", "RATIO_9_CHANGE",
"RATIO_10_CHANGE", "RATIO_11_CHANGE", "RATIO_12_CHANGE",
"RATIO_13_CHANGE", "RATIO_14_CHANGE", "RATIO_15_CHANGE") AS
 SELECT FAT_POINT.USER_ID,
 ROUND(((RATIO_1 -BASELINE_RATIO_1)/ BASELINE_RATIO_1) * 100,2) RATIO_1_CHANGE,
 ROUND(((RATIO_2 -BASELINE_RATIO_2)/ BASELINE_RATIO_2) * 100,2) RATIO_2_CHANGE,
 ROUND(((RATIO_3 -BASELINE_RATIO_3)/ BASELINE_RATIO_3) * 100,2) RATIO_3_CHANGE,
 ROUND(((RATIO_4 -BASELINE_RATIO_4)/ BASELINE_RATIO_4) * 100,2) RATIO_4_CHANGE,
 ROUND(((RATIO_5 -BASELINE_RATIO_5)/ BASELINE_RATIO_5) * 100,2) RATIO_5_CHANGE,
 ROUND(((RATIO_6 -BASELINE_RATIO_6)/ BASELINE_RATIO_6) * 100,2) RATIO_6_CHANGE,
 ROUND(((RATIO_7 -BASELINE_RATIO_7)/ BASELINE_RATIO_7) * 100,2) RATIO_7_CHANGE,
 ROUND(((RATIO_8 -BASELINE_RATIO_8)/ BASELINE_RATIO_8) * 100,2) RATIO_8_CHANGE,
 ROUND(((RATIO_1 -BASELINE_RATIO_9)/ BASELINE_RATIO_9) * 100,2) RATIO_9_CHANGE,
 ROUND(((RATIO_2 -BASELINE_RATIO_10)/ BASELINE_RATIO_10) * 100,2)
RATIO_10_CHANGE,
 ROUND(((RATIO_3 -BASELINE_RATIO_11)/ BASELINE_RATIO_11) * 100,2)
RATIO_11_CHANGE,
 ROUND(((RATIO_4 -BASELINE_RATIO_12)/ BASELINE_RATIO_12) * 100,2)
RATIO_12_CHANGE,
 ROUND(((RATIO_5 -BASELINE_RATIO_13)/ BASELINE_RATIO_13) * 100,2)
RATIO_13_CHANGE,
 ROUND(((RATIO_6 -BASELINE_RATIO_14)/ BASELINE_RATIO_14) * 100,2)
RATIO_14_CHANGE,
 ROUND(((RATIO_7 -BASELINE_RATIO_15)/ BASELINE_RATIO_15) * 100,2)
RATIO_15_CHANGE
 FROM
 ( SELECT FAT.USER_ID,FATIGUED_BEGIN_PERIOD, RATIO_1,RATIO_2,RATIO_3,
RATIO_4,RATIO_5,RATIO_6,RATIO_7,RATIO_8,RATIO_9,RATIO_10,RATIO_11,RATIO_12,RATIO_13,RATIO_14,RATIO_15
   FROM V_FATIGUE_PERIOD FAT
   JOIN eeg_5_min_interval EEG
   ON EEG.user_id = FAT.user_id
   AND EEG.TIME_INTERVAL = FAT.fatigued_begin_period)FAT_POINT
 JOIN (SELECT ROUND(AVG(RATIO_1),2) BASELINE_RATIO_1,
ROUND(AVG(RATIO_2),2) BASELINE_RATIO_2,ROUND(AVG(RATIO_3),2)
BASELINE_RATIO_3, ROUND(AVG(RATIO_4),2) BASELINE_RATIO_4,
        ROUND(AVG(RATIO_5),2) BASELINE_RATIO_5, ROUND(AVG(RATIO_6),2)
BASELINE_RATIO_6,ROUND(AVG(RATIO_7),2)
BASELINE_RATIO_7,ROUND(AVG(RATIO_8),2) BASELINE_RATIO_8,
        ROUND(AVG(RATIO_9),2) BASELINE_RATIO_9,
ROUND(AVG(RATIO_10),2) BASELINE_RATIO_10,ROUND(AVG(RATIO_11),2)
BASELINE_RATIO_11,ROUND(AVG(RATIO_12),2) BASELINE_RATIO_12,
        ROUND(AVG(RATIO_13),2)BASELINE_RATIO_13,ROUND(AVG(RATIO_14),2)
BASELINE_RATIO_14,ROUND(AVG(RATIO_15),2) BASELINE_RATIO_15,USER_ID
       FROM EEG_5_MIN_INTERVAL
       WHERE TIME_INTERVAL < 4
       GROUP BY USER_ID) BASE
ON BASE.USER_ID = FAT_POINT.USER_ID;
```

Figure 7. V_FATIGUE_CHANGE_ANALYSIS View

```
 CREATE OR REPLACE FORCE VIEW "SYSTEM"."V_FATIGUE_PERIOD" ("USER_ID",
"FATIGUED_BEGIN_PERIOD") AS
 SELECT EEG.USER_ID,MIN(EEG.TIME_INTERVAL) FATIGUED_BEGIN_PERIOD
FROM EEG_5_MIN_INTERVAL EEG
JOIN FATIGUE_REPORT REP
ON EEG.TIME_INTERVAL = REP.TIME_INTERVAL
AND EEG.USER_ID = REP.USER_ID
WHERE FATIGUE_LEVEL = 5
GROUP BY EEG.USER_ID;
```

Figure 8. V_FATIGUE_PERIOD View

CHAPTER 6

RESULTS

In this thesis, we utilized Percent Change analysis to determine whether any of the EEG ratios reported in the literature could reliably detect fatigue when a one channel EEG device was used. To accomplish this, the EEG Ratio data were grouped into 5-minute time sections, as previous studies have done, and to coincide with participant fatigue reports. To establish a baseline, the first 3 sections of the data were averaged similar to (Jap, Lal, Fischer, & Bekiaris, 2009). Afterward, the point in time that a participant felt fatigue was identified by locating the time interval that the participant stated they felt a 5 on the Stanford Sleepiness Scale. The data were again averaged for this 5-minute time interval. The Percent Change analysis was done by performing Equation 24 based on the baseline and the 5-minute time interval that the participant felt fatigued.

6.1 Data Normalization

Recall that every ratio was calculated for each data sample returned by the Neurosky device. Because the EEG Ratio Data is widely distributed between the 15 ratios (see Table 8), the data analysis was conducted on a normalized version of the dataset. Normalization was done at the sample level for all ratios by performing Feature Scaling (see Equation 25). This scaled the data to a fixed range between 0 and 1 for all ratios.

RATIO	MIN	MAX	Standard Deviation
Ratio 1	0.29	162.73	12.79473163
Ratio 2	0.29	170.35	12.7824389
Ratio 3	0.28	163.58	12.12017763
Ratio 4	0.19	194.98	13.58932639
Ratio 5	0.3	88.79	5.393878485
Ratio 6	0.3	194.47	14.72352248
Ratio 7	0.28	203.93	15.03395592
Ratio 8	0.2	223.63	15.63656814
Ratio 9	0.24	213.63	15.34259118
Ratio 10	0.27	101.64	6.158363767
Ratio 11	0.12	289.94	22.6935848
Ratio 12	0.19	303.45	26.54314328
Ratio 13	0.17	322.17	24.87672562
Ratio 14	0.11	275.68	23.55482818
Ratio 15	0.08	115.37	10.01159366

Table 8. Un-Normalized Ratio Min and Max

$$Xnorm = \frac{X - Xmin}{Xmax - Xmin}$$

Equation 25. Feature Scaling Normalization

6.2 Percent Change and Frequency Distribution

Out of the 10 participants who participated in the study, only 6 stated that they felt a rating of 5 on the Stanford Sleepiness Scale, despite driving for one hour in the simulator. Therefore, the Percent Change analysis was only performed on those participants, for the purposes of identifying Percent Change from an alert state to a fatigue state. Tables 9, 10, and 11 demonstrate the Percent Change results of the study using the Normalized data set. Table 9 shows the Percent Change for each ratio by participant. Table 10 shows the Percent Change Frequency Distribution in 10 % increments. This was done to try and identify a Percent Change that is common across the participant base to establish if one or more of the ratios showed potential as a fatigue indicator. For example, 3 out of 6 participants (50%) felt a change within the range of 11-20% for Ratios 1 and 3. Similarly, 50% of the participants felt a change within the range of 31-40% for Ratio 5. Table 10 demonstrates that raw statistics for Percent Change per participant for each ratio. Table 11 demonstrates the most common Percent Change for each ratio, the number of participants who felt a Percent Change in that range, and the percentage of participants for each ratio. Based on this analysis, Ratios 1, 3, 5, and 6 appear to be the most likely candidates for fatigue detection based on the Un-Normalized data because 50% of the participants felt a common Percent Change based on this approach.

Participant	1	2	3	4	8	9
RATIO_1_CHANGE	4.977%	15.854%	16.308%	19.637%	26.828%	29.170%
RATIO_2_CHANGE	9.513%	-28.918%	61.147%	40.652%	60.175%	5.896%
RATIO_3_CHANGE	3.506%	17.737%	13.285%	13.762%	26.361%	37.711%
RATIO_4_CHANGE	11.960%	-26.880%	68.497%	48.205%	60.893%	1.234%
RATIO_5_CHANGE	10.286%	-13.387%	35.215%	32.644%	33.896%	16.350%
RATIO_6_CHANGE	5.492%	16.584%	16.086%	21.343%	28.797%	30.247%
RATIO_7_CHANGE	10.542%	-28.531%	64.511%	43.737%	63.826%	3.163%
RATIO_8_CHANGE	12.266%	-27.091%	69.672%	49.150%	64.243%	-0.045%
RATIO_9_CHANGE	-38.179%	-14.705%	13.930%	31.741%	-9.663%	4.827%
RATIO_10_CHANGE	145.476%	34.548%	154.270%	107.527%	205.565%	84.960%
RATIO_11_CHANGE	-28.992%	-12.749%	-41.531%	60.593%	-16.431%	36.888%
RATIO_12_CHANGE	0.409%	-51.161%	-7.917%	93.606%	21.494%	-25.064%
RATIO_13_CHANGE	-50.609%	-69.268%	-59.169%	-5.749%	-48.617%	-53.152%
RATIO_14_CHANGE	-27.824%	-11.944%	-17.513%	136.615%	-8.451%	39.051%
RATIO_15_CHANGE	173.005%	37.859%	105.491%	326.177%	227.599%	133.669%

Table 9. Ratio Percent Change from Alert State to Fatigue State by Participant

RANGE	R 1	R 2	R 3	R 4	R 5	R6	R 7	R 8	R 9	R1 0	R1 1	R1 2	R1 3	R1 4	R1 5
(-70) – (-61)%	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
(-60) – (-51)%	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0
(-50) – (-41) %	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0
(-40) – (-31)%	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
(-30) – (-21) %	0	1	0	1	0	0	1	1	0	0	1	1	0	1	0
(-20) – (-11)%	0	0	0	0	1	0	0	0	1	0	2	0	0	2	0
(-10) – (0) %	0	0	0	0	0	0	0	1	1	0	0	2	1	1	0
1-10%	1	2	1	1	1	1	2	0	1	0	0	0	0	0	0
11-20%	3	0	3	1	1	2	0	1	1	0	0	0	0	0	0
21-30%	2	0	1	0	0	3	0	0	0	0	0	1	0	0	0
31-40%	0	1	1	0	3	0	0	0	1	1	1	0	0	1	1
41-50%	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0
51-60%	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
61-70%	0	1	0	2	0	0	2	2	0	0	0	0	0	0	0
81-90%	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
91-100%	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
101-110%	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
131-140%	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
141-150%	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
151-160%	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
171-180%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
201-210%	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
221-230%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
311-320%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 10. Percent Change Based on Frequency Distribution (Normalized)

	Percent of Change	Number of Participants	Percent of Participants
R1	11-20%	3	50%
R2	1-10%	2	33%
R3	11-20%	3	50%
R4	61-70%	2	33%
R5	31-40%	3	50%
R6	21 -30%	3	50%
R7	1-10, 61-70%	2	33%
R8	61-70%	2	33%
R9	(-40) – (-31)%, (-20)– (-11)%, (-10) –(0)%, 1-10%, 11-20%, 31-40%	1	17%
R10	31-40%, 81-90%, 101 -110%, 141-150%, 151-160%, 201-210%	1	17%
R11	(-20) - (-11)%	2	33%
R12	(-10) – (0)%	2	33%
R13	(-60) - (-51)%, (-50) - (-41)%	2	33%
R14	(-20)- (-11)%	2	33%
R15	31-40%, 101-110%, 131-140%, 161-170%, 221-230%, 311-320%	1	17%

Table 11. Most Common Percent Change for Each Ratio (Normalized)

Simply grouping the data into 10% ranges did not help us with identifying a common Percent Change across the participant base. Fifty percent of the participant base does not seem common enough to implement Percent Change for any of these ratios in a real-time system. Frequency Distribution Analysis provides a skewed view of the data because even though 2 participants may have felt a Percent Change of 19% and 21% the two participants would have fallen into different frequency groups. However, another view of the data may provide better insights.

6.3 Cluster Analysis

As an alternative to Frequency Distribution Analysis, we employed clustering analysis, which allowed us to overcome the boundary problem mentioned above. Clustering is "the process of forming groups of items or entities such that entities within a group are similar to one another and different from those in other groups. The similarity between entities is determined based on their characteristics of features." (Maqbool & Babri, 2007). Hierarchical clustering is the recommended clustering approach for small datasets because it allows you to examine solutions with an increasing number of clusters (Camm, Cochran, Fry, Ohlmann, & Anderson, 2015). Contrast this with a K-means approach where the number of desired cluster needs to be known beforehand (Camm, Cochran, Fry, Ohlmann, & Anderson, 2015). Because the number of desired clusters is not necessarily known it is more advantageous to use Hierarchical Clustering in this case.

Specifically, Agglomerative Hierarchical Clustering was used in this case. Agglomerative is a bottom-up approach where each data point starts out as 1 cluster. Clusters are then combined recursively by equating the distance between 2 clusters until all clusters have been merged into 1 (Xu & Wunsch II, 2008). RapidMiner Studio was utilized to execute this clustering algorithm. RapidMiner Studio is a toolset used to perform a variety of data science and analytical functions (RapidMiner, 2017).

For each ratio, the Agglomerative Hierarchical Clustering algorithm was executed to look for closely related Percent Change data points between the participant groups that felt fatigue. To do this the data were brought into RapidMiner using the "Read Excel" module. This module simply reads an excel file and imports the data for usage within the Rapid Miner tool (RapidMiner, 2017). The output of the Read Excel module was fed into the "Agglomerative Clustering" module, which takes the data and runs the algorithm based on the parameters established for the module (RapidMiner, 2017). The Measures Type parameter was set to Numerical Measures (appropriate because our data is numerical). The module was ran using Single Link mode in conjunction with Euclidean distance. Single Link was appropriate for this study because it simply merges the clusters recursively based on the smallest distance between members of the clusters with the smallest distance (Patra, & Nandi, 2015). Euclidean distance type is simply the "straight line" approach to find the distance between 2 points (Monroe & Chapman McGrew Jr., 2009).

The output of the Agglomerative Clustering Module is a dendrogram. A dendrogram is a tree structure that shows how the clusters merge from N number of objects into 1 cluster (Vahidipoura, Mirzaei, & Rahmati, 2014). This works well for smaller datasets such as this one where the dendrogram can be visually inspected to identify relevant clusters and outliers. However, a tabular result is easier to read and analyze for our purposes (keep in mind that what we were interested in finding a small range of data for a higher percentage of the population). To do that, a small number of clusters needed to be identified for each ratio.

RapidMiner accommodated this by incorporating the Flat Hierarchy module. The Flat Hierarchy module takes the number of desired clusters as an input and outputs a tabular view of what participant is assigned to which cluster. This method was somewhat trial and error to establish a tight clustering of data. Because there is no standard definition of what a tight clustering of data is, an arbitrary definition was established. Only clusters with a range of 20 or less were considered of interest for our study. If there was no cluster with a data range of less than 20 for at least 50% of the participants, then that ratio was not considered as a potential fatigue indicator. Table 12 is a summary of the number of clusters and the range of data for each ratio that had a cluster with a data range of less than 20. A detailed result set can be found in Appendix D for each ratio.

Ratio	No. of Clusters	Range	Range Difference	Number of Participants	Percent of Participants	Standard Deviation
Ratio 1	2	15.85 – 29.17%	13.32	5	83%	6.11
Ratio 6	2	16.08 – 30.24%	14.16	5	83%	6.65
Ratio 13	2	(-69.26)–(-48.61)%	20.65	5	83%	8.33
Ratio 14	3	(-27.82) – (-8.45)%	19.37	4	66%	8.46

Table 12. Summary of Relevant Clusters (Normalized)

No clusters were identified where 100% of the participants fell into a range of 20 or less. However, four ratios had a cluster with a range of approximately 20 or less for at least four out of the six participants (66%). Of these four ratios, Ratio 1 had the smallest range with the most participants which may suggest it as a good indicator of fatigue. However, Ratios 6, 13, and 14 also showed potential as fatigue indicators because their range was relatively small and they have a high percentage of participants.

6.4 Fatigued vs Non-Fatigued

As an additional check to validate the ratios identified in the cluster analysis, a comparison was made to the data from the users who stated that they did not feel fatigue (these users never stated that they felt a 5 on the Stanford Sleepiness Scale) to the users who stated they felt fatigue. For participants who did not feel fatigue, there was no point in time they stated that they felt a 5 on the Stanford Sleepiness Scale, a relevant value for some time period needed to be identified so that Percent Change Analysis could be performed. To identify such a value, the last 15 minutes of the study was averaged for these participants. This was similar to the approach taken to calculate the participant's baseline. Our hope here was that because these participants never felt fatigue then the Percent Change Analysis on the last 15 minutes of the data would be significantly different to that of the participants who felt fatigue. This analysis was conducted for each ratio of interest (listed in Table 12). Essentially, we are looking for conflicts between Percent Change range measure for fatigue and non-fatigue participant groups. Conflict in this context would be when the Percent Change value of a non-fatigue participant fell in the range reported in Table 15 for fatigue participant groups.

For Ratio 1, two of four participants who did not feel fatigue had approximately 4% change (see Table 13). However, 21.4 and 17.5% change were experienced by two participants which conflicts with the Ratio 1 Cluster range in Table 12. For Ratio 6, two of four participants who did not feel fatigue experienced a Percent Change of 20.5 and 18.4, which conflicts with the Ratio 6's cluster range (see Table 14). All participants who did not feel fatigue for Ratio 13 conflict with the Ratio 13 cluster range in Table 15. Finally, for Ratio 14 two of the four participants conflict with the cluster range in Table 16.

R1 Cluster Range 15.85 – 29.17%	
Participant	**Percent Change**
6	4.59
5	21.43
7	4.05
10	17.56

Table 13. R1 Last 15 Minutes of Participants who did not feel Fatigue (Normalized)

R6 Cluster Range 16.58 – 30.24%	
Participant	**Percent Change**
6	4.87
5	20.50
7	4.52
10	18.45

Table 14. R6 Last 15 Minutes of Participants who did not feel Fatigue (Normalized)

R13 Cluster Range (-)69.26– (-)48.61%	
Participant	**Percent Change**
6	-49.22
5	-60.89
7	-59.93
10	-66.77

Table 15. Ratio 13 Non-Fatigued Last 15 Minutes

R14 Cluster Range (-)27.82 – (-)8.45%	
Participant	**Percent Change**
6	15.01
5	28.40
7	-25.41
10	-22.77

Table 16. Ratio 14 Non-Fatigued Last 15 Minutes

6.4.1 Non-Fatigued Anomalies

The Fatigued vs Non-Fatigued participant analysis appeared to discredit the ratios of interest that were identified in the cluster analysis by highlighting conflicts between the participants who stated they had not felt fatigue and those who did. To better understand these conflicts a deeper dive into the sessions of these participants was necessary. Beginning with Ratio 1, Participants 5 and 10 experienced a Percent Change that conflicted with the range of the cluster identified in Table 12. After considering the user fatigue reports of both Participant 5 and 10, a possible explanation for why this occurred became clear. Both participants reported a 4 on the Stanford Sleepiness Scale for their last 3 user reports (the last 15 minutes of the study). Participants 5 and 10 experienced a percent change of approximately 21% and 17% respectively. The lower bound for the Ratio 1 cluster range was approximately 15%. Both Participants 5 and 10's Percent Change crossed into the bottom half of Ratio 1's cluster range indicating that the participants were both approaching a 5 on the scale for their last 15 minutes of driving. Similarly, for Ratio 6, both Participants 5 and 10 felt a conflicting Percent Change with Ratio 6's cluster range. The Percent Change for these participants were 20% and 18% for Participant 5 and 10. Again these Percent Change levels are in the bottom half of Ratio 6's cluster range. These discrepancies could be result of participants misinterpreting the Stanford Sleepiness Scale or under estimating their level of fatigue.

All non-fatigued participants conflicted with Ratio 13's cluster range. Because of this it is likely not a suitable candidate for fatigue detection. For Ratio 14, Participants 7 and 10 experienced a Percent Change of -25% and -22% respectively. Again, the user fatigue reports were analyzed to determine if the participants were close to a fatigue state. When contrasting these Percent Change levels with the participants who experienced fatigue inside of Ratio 14's cluster range, both Participants 7 and 10 were on the bottom end of this range. While Participant 10 stated their fatigue level at 4 for their last 3 fatigue reports, Participant 7 only recorded a 2 for their last 3 reports. Therefore, it cannot be concluded that the Ratio 14 anomalies can be blamed on misinterpretation of the Stanford Sleepiness Scale.

Participant 1 was the only person inside of the Ratio 14 cluster that had a lower Percent Change than Participants 7 and 10. If Participant 1 could be considered an anomaly, and was removed from the Ratio 14 cluster, than neither Participant 7 or 10 would conflict with the cluster at all. However, if this were to be true, then only 3 out of 10 participants (50%) would fall into this cluster. That is simply too low to be considered common for our study. Therefore, Ratio 14 cannot be considered as a good indicator of fatigue.

USER_ID	RATIO_14_CHANGE
1	-27.82
3	-17.51
2	-11.94
8	-8.45

Table 17. Ratio 14 Cluster for Fatigued Users

6.5 Conclusions

The Frequency Distribution Analysis did not yield any results that could help identify a commoner Percent Change. This was likely attributed to the boundary problem mentioned in Section 6.2. However, the cluster analysis did appear to present a useful result set. Four ratios were identified that had a cluster with a range of approximately 20 or less for at least for four out of the six (66%) of the participants. Of these four ratios, Ratio 1 had the smallest range with the most participants which suggested it may be the best indicator of fatigue.

While these ratios appeared to show promise, they were significantly discredited based on the analysis of the participants who did not feel fatigue. Each of these ratios displayed at least two conflicts with the non-fatigued users. Despite this, there are possible explanation for these conflicts for each of the ratios except Ratio 13. Because of the potential for errors in the participant fatigue reports it is very likely that Ratio 1, Ratio 6, and Ratio 14 are all good indicators of fatigue.

CHAPTER 7

CONCLUDING REMARKS

Prior to the study conducted in this thesis, there had been no empirical research conducted to determine at what point fatigue begins to occur for a driver. This research has provided a novel method for identifying the point in time when a person begins to feel fatigue, and could potentially be used in a real-time system to prevent accidents caused by fatigue.

By using a clustering approach, we observed that 83% of the participants experienced a similar Percent Change for Ratios 1, 6, and 14. Further analysis revealed that Ratio 1 had the smallest range of data points as well (15.85 – 29.17). With the majority of participants feeling fatigue between this range, it is conceivable that a threshold could be established that would trigger an alarm in a real-time system. For example, the mean or the median of these data points could be used to set the threshold.

However, there were several occasions where data from participants who did not feel fatigue conflicted with these results. While at first glance this appeared to discredit these ratios, a deeper dive into the Non-Fatigued participants appeared to provide an explanation for why this occurred. That being stated, Ratios 1, 6, and 14 all appeared to be good indicators of fatigue.

Additionally, the question of using a 1-Channel EEG to monitor fatigue was addressed in this thesis. Due to a relatively high number of participants demonstrating a common Percent Change, it appears that a 1-Channel EEG is suitable for such a purpose. This would likely increase usability as opposed to using a multi-channel EEG device in a real-time system.

Several factors and limitations could have potentially affected the outcomes of this study. One factor is that the analysis is entirely based on the participant self-reports of fatigue. There are many potential issues with this. First, the participants may have inaccurately reported their level of fatigue based on their interpretation of the Stanford Sleepiness Scale. For instance, a participant may have found the distinction between a 4 (somewhat foggy, let down) and a 5 (Foggy; losing interest in remaining awake; slowed down) somewhat arbitrary. An inaccurate self-reporting here could have potentially affected the study. Second, it is possible that the participant may have falsely reported their fatigue level as a means of trying to keep themselves awake. Either of these may have caused the anomalies that were seen with the Fatigue vs. Non-Fatigued Analysis.

Another potential limitation to this approach is that the study assumes that the participant is not already in a fatigue state when they begin driving. This is because the baseline is a representation of the driver at optimal performance (in our case an awake state). If the participant enters a vehicle nearly ready to fall asleep already, the baseline itself would not accurately represent the driver at an awake state, and furthermore, depending on the driver's level of fatigue, they may not even be able to stay awake for the 15-minute baseline period. However, because of the equations that the participants were asked to perform during the pre-study validation period, it can be reasoned that the participants were in an alert state, which may have mitigated this risk.

Additionally, the sample size of the data is too small to conclude that these results are irrefutable. Because of this we can only draw conclusions from the data that was captured from the 10 participants in this study. Ideally, a larger sample size should be included in future work. Larger sample size could also be helpful in comparing ratios through statistical analysis and identifying Percent Change that may be of statistically significant.

Despite the limitations of this study, the results observed were encouraging. Based on the results of our experiment, it is recommended that future work investigates how the data collected in our study can be used to identify a threshold for when participants begin to experience fatigue. Identifying this threshold will enable the development of a real-time fatigue detection system which could save lives.

In order to develop a real time fatigue detection system, further research is required to identify the best way to embed a single channel EEG into wearable format that is suited for long term usage and not intrusive to drivers. Further research works are also needed for developing efficient software components for presenting fatigue alerts to the driver based on data received from EEG.

REFERENCES

Baars, B. J., & Gage, N. M. (2007). *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience.* Academic Press .

BrainWorks. (2011, December 3). *WHAT ARE BRAINWAVES?* Retrieved June 7, 2015, from http://www.brainworksneurotherapy.com/what-are-brainwaves

Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., & Anderson, D. R. (2015). *Essentials of Business Analytics.* Boston, MA: CENGAGE Learning.

Cochran, W., Cooley, J., Favin, D., & He, H. (1967). What is the fast Fourier transform? *What is the fast Fourier transform?, 55*(10), 1664-1674.

Dinu, V., & Nadkarni, P. (2007, Novemenber-December). Guidelines for the Effective Use of Entity-Attribute-Value Modeling for Biomedical Databases. *International Journal of Medical Informatics, 76*(11-12), 769–779.

Eoh, H., Chung , M., & Kim, S.-H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics , 35*, 307–320.

Graimann, B., Allison, B., & Pfurt, G. (2010). Brain–Computer Interfaces: A Gentle Introduction. In *Brain-Computer Interfaces* (pp. 1-27). Springer Berlin Heidelberg.

Hal, B. V., Rhodes, S., & Dunne, B. (2014). Low-Cost EEG-based Sleep Detection. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4571-4574). Chicago: IEEE.

Hsieh , C.-S., & Tai, C.-C. (2013). AN IMPROVED AND PORTABLE EYE-BLINK DURATION DETECTION SYSTEM TO WARN OF DRIVER FATIGUE. *Instrumentation Science and Technology,, 41*, 429-444.

Hu, H., Guo, S., Liu, R., & Wang, P. (2017, 06 28). *An adaptive singular spectrum analysis method for extracting brain rhythms of electroencephalography.* Retrieved from PeerJ: https://doi.org/10.7717/peerj.3474

Jagannatha , M., & Balasubramanian, V. (2014). Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Applied Ergonomics , 45*, 1140-1147.

Jap, B. T., Lal, S., Fischer, P., & Bekiaris, E. (2009). Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications, 36*, 2352-2359.

K.L. Lal, S., & Craig, A. (2001). A critical review of the psychophysiology of Driver Fatigue. *Biological Psychology, 55*, 173–194.

Kaplan. (2016). *PSAT Math: Percentages - Kaplan Test Prep*. Retrieved 01 10, 2016, from KAPLAN Test Prep: http://www.kaptest.com/study/psat/psat-math-percentages/

Kar , S., Bhagat, M., & Routray, A. (2010, June 28). EEG signal analysis for the assessment and quantification of driver's fatigue. *Transportation Research Part F*, 297–306.

Lal, S. K., & Craig, A. (2002). Driver Fatigue: Electroencephalography and psychological assessment. *Psychophysiology, 39*, 313-321.

Lal, S., & Craig, A. (2000). Physiological indicators of Driver Fatigue. *ROAD SAFETY RESEARCH, POLICING AND EDUCATION CONFERENCE, 2000, BRISBANE, QUEENSLAND, AUSTRALIA* (pp. 489-494). Queensland: Queensland University of Technology.

Lal, S., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology 173–194, 55*, 173-194.

Lamond , N., & Dawson, D. (2002, January 05). Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research, 8*(4), 255-262.

Lebedev, M. A., & Nicolelis, M. A. (2006). Brain–machine interfaces: past,present, and future. *Trends in Neurosciences, 29*(9), 536-546.

Maqbool, O., & Babri, H. (2007, 11). Hierarchical Clustering for Software Architecture Recovery. *Hierarchical Clustering for Software Architecture Recovery*, 759 - 780.

Ming-ai , L., Cheng , Z., & Jin-Fu, Y. (2010). An EEG-based Method for Detecting Drowsy Driving State. *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery* , 2164-2167.

Minnesota Department of Transportation. (2015, January). *Monitoring Motor Vehicle Driver Fatigue.* Office of Transportation System Management. CTC & Associates. Retrieved September 15, 2016, from Minesota Department of Transportation: http://www.lrrb.org/media/reports/TRS1501.pdf

Monroe , C., & Chapman McGrew Jr., J. (2009). *An Introduction to Statistical Problem Solving in Geography: 2nd Edition.* Long Grove , IL: Waveland Press, Inc.

Naqvia , S., Badruddin, N., Malik, A., Hazabbah, W., & Abdullah, B. (2014). EEG Alpha Power: An Indicator of Visual Fatigue. *2014 5th International Conference on Intelligent and Advanced Systems (ICIAS)* (pp. 1-5). Kuala Lumpur: IEEE.

National Highway Traffic Safety Administration . (2011). *Drowsy Driving* . National Highway Traffic Safety Administration , Department of Transporation. Washington D.C.: NHTSA’s National Center for Statistics and Analysis.

National Highway Traffic Safety Administration. (2011, March). *ww.NHTSA.gov.* Retrieved January 17, 2015, from National Highway Traffic Safety Administration: http://www-nrd.nhtsa.dot.gov/Pubs/811449.pdf

National Highway Traffic Safety Adminstration. (2017, March 20). *Drowsy Driving.* Retrieved October 21, 2017, from National Highway Traffic Safety Adminstration: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/12723-drowsy_driving_asleep_at_the_wheel_031917_v4b_tag.pdf

Neurosky. (08, 05 2015). *MindWave Mobile: User Guide.* Retrieved 12 1, 2016, from MindWave Mobile: User Guide: http://download.neurosky.com/support_page_files/MindWaveMobile/docs/mindwave_mobile_user_guide.pdf

NeuroSky. (2014, 05 06). *NeuroSky Developer Program*. Retrieved 12 14, 2016, from ThinkGear Connection Driver: http://developer.neurosky.com/docs/doku.php?id=thinkgear_communications_driver_tgcd

Neurosky. (2014, 09 04). *ThinkGear Serial Stream Guide.* Retrieved from Neurosky Developer Program:

http://developer.neurosky.com/docs/doku.php?id=thinkgear_communications_protocol#thinkgear_data_values
Neurosky. (2015). *Minwave*. Retrieved from Neurosky Store: http://store.neurosky.com/pages/mindwave
Neurosky. (2015, June 19). *What Is BCI and How Did It Evolve?* Retrieved Jan. 25, 2016, from www.neurosky.com: http://neurosky.com/2015/06/what-is-bci-and-how-did-it-evolve/
Occupation Safety and Health Administration. (2015). *Extended Unusual Work Shifts*. Retrieved January 17, 2015, from www.OSHA.gov: https://www.osha.gov/OshDoc/data_Hurricane_Facts/faq_longhours.html
Oracle. (2016). *Introduction*. Retrieved 12 05, 2016, from Oracle Java SE Documentation: http://docs.oracle.com/javase/7/docs/technotes/guides/jni/spec/intro.html
Patra, , B., & Nandi, S. (2015, Jan). Effective data summarization for hierarchical clustering in large datasets. *Knowledge & Information Systems; , 42*(1), 1-20.
Punsawad, Y., Aempedchr, S., Wongsawat, Y., & Panichkun, M. (n.d.). WEIGHTED-FREQUENCY INDEX FOR EEG-BASED MENTAL FATIGUE ALARM SYSTEM. 36-41.
RapidMiner. (2017). *Agglomerative Clustering.* Retrieved 07 22, 2017, from RapidMiner Documentation: https://docs.rapidminer.com/studio/operators/modeling/segmentation/agglomerative_clustering.html
RapidMiner. (2017). *RapidMiner Studio VISUAL WORKFLOW DESIGNER FOR DATA SCIENCE*. Retrieved 07 22, 2017, from RapidMiner: https://rapidminer.com/products/studio/
RapidMiner. (2017). *Read Excel - RapidMiner Documentation.* Retrieved 07 22, 2017, from RapidMiner Documentation: https://docs.rapidminer.com/studio/operators/data_access/files/read/read_excel.html
Ratti, E., Waninger, S., Berka, C., Ruffini, G., & Verma, A. (2017, Aug 3). Comparison of Medical and Consumer Wireless EEG Systems for Use in Clinical Trials. *Frontiers in Human Neuroscience, 11*, 398.
Rocha, A., Thomaz, C., Rocha, F., & Vieito, J. (2017). *A Practical Guide to Brain Data Analysis.* Singapore: World Scientific Publishing Co. Pte. Ltd.
Rubin, M. D. (2009). *Clinical Neurophysiology.* New York, NY: Oxford University Press.
Shier, D., Butler, J., & Lewis, R. (2009). *Hole's Essentials of Human Anatomy & Physiology.* New York , NY: McGraw-Hill Higher Education.
Siddiqui, U., & Shaikh, A. (2013, November). An Overview of "Electrooculography". *International Journal of Advanced Research in Computer and Communication Engineering, 2*(11), 4328=4330.
Sigari, M.-H., Pourshahabi, M.-R., Soryani, M., & Fathy, M. (2014). A Review on Driver Face Monitoring Systems for Fatigue and Distraction Detection. *International Journal of Advanced Science and Technology, 64*, 73-100.
Stanford Sleepiness Scale. (2017). (Stanford, Producer) Retrieved from Stanford.edu: https://web.stanford.edu/~dement/sss.html

Systems Technology, Inc. (2016). *STISM Drive*. Retrieved 12 05, 2016, from www.stismdrive.com: http://stisimdrive.com/products/
Teplan, M. (2002). Fundamentals of EEG Measurement. *MEASUREMENT SCIENCE REVIEW, 2*.
Transparent Coroporation. (2013, May 2). *Gamma Brain Waves Infographic*. Retrieved June 7, 2015, from http://www.transparentcorp.com/research/gamma-brain-waves.php
Vahidipoura, S., Mirzaei, A., & Rahmati, M. (2014). Comparing weighted combination of hierarchical clustering based on Cophenetic measure. *Intelligent Data Analysis*, 547-559.
Vallabhaneni, A., Wang, T., & He, B. (2005). *Neural Engineering.* Chicago: Springer US.
Veeraraghavan , H., & Papanikolopoulos, N. (2001). *Detecting Driver Fatigue Through the Use of Advanced Face Monitoring Techniques.* Intelligent Systems Transportation Institute, Department of Computer Science and Engineering, Minneapolis.
Wascher, E. B. (2014, February). Frontal theta activity reflects distinct aspects of mental fatigue. *Biological Psychology* , 57-65.
Xu, R., & Wunsch II, D. (2008). *Clustering.* Piscataway, NJ: Wiley.
Yasui, Y. (2009). A Brainwave Signal Measurement and Data Processing Technique for Daily Life Applications. *Journal of Physiological Anthropology*, 145-150.
Yuvaraj, R., Murugappan, M., Ibrahim, N., Omar, M., Sundaraj, K., Mohamad, K., . . . Satiyan, M. (2014). *On the analysis of EEG power, frequency and asymmetry in Parkinson's disease during emotion processing.* Universiti Malaysia Perlis, School of Mechatronic Engineering,. Arau: BioMed Central.
Zhaoa , C., Zhaoa , M., Liua, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulatorElectroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis and Prevention , 45*, 83-90.

APPENDIX A

Email Script to contact Professors

Dear Professor,

My name is Lucas Coffey and I am a graduate student at the University of North Florida. For my master's thesis, I'm conducting research on fatigue detection using an EEG headset. We are looking for UNF students to take part in our study. We anticipate that study participation might consume 1 hour and15 minutes of their time. Participation in the study is completely voluntary and students will not be compensated for their participation. We do not foresee any risks for taking part in the study. However, we anticipate that study results will be beneficial in determining a method for detecting driver fatigue.

With this email, we request your assistance in recruiting participants for our study. UNF Students are the ideal participants for this study. We would like you to post below provided study participation information in your Blackboard course pages. This study has been approved by UNF IRB.

If you have any questions or concerns about this study, please contact me

or my professor, Dr. Umapathy

Thank you very much for your assistance.

Sincerely,

-Lucas Coffey

Course Management System post message start -----

Hello,

My name is Lucas Coffey and I am a graduate student at the University of North Florida. For my master's thesis, I'm conducting research on fatigue detection using an EEG headset. To access the detection algorithms, we are requesting that you simulate driving via a driving simulator while wearing an EEG headset. This should only take approximately 1 hour and 15 minutes of your time, and it would greatly help us with our research. As a part of the study, you will simply be required to drive for approximately 1 hour via the simulator. Periodically during the study, we will ask you to report your level of fatigue.

Please note that your participation is completely voluntary and your response will be anonymous. In order to participate in the study, you must also be 18 years or above. This study is approved by UNF IRB (IRB # 475514-4)

Although there are no direct benefits or compensation for taking part in this study, others may benefit from the information we find from the results of this study. Additionally, there are no foreseeable risks for taking part in this project. There are no penalties for not responding to a question or ceasing participation. If you choose to withdraw from this study, there will be no penalty or loss of benefits to which you would otherwise receive.

If you have any questions or concerns about this study, please contact me

) or my professor, Dr. Umapathy

If you would like to participate, then visit below website and follow the instructions presented to you:

[URL HERE]

Your participation is an immense help to us, and we greatly appreciate your help.

Sincerely,

-Lucas Coffey

Course Management System post message end ---

APPENDIX B

Instructions Provided to Participants

Upon the participant's arrival: "Good Evening. My name is Lucas Coffey I will be administering the study today. Before we get started I would like to explain to you the purpose of the study. We are attempting to test three different algorithms as potential fatigue predictors. We will do this by collecting brain waves via an electroencephalograph. During this process, I will ask you to wear this device while you simulate driving. The simulator will be controlled by a steering wheel, a gas pedal, and a brake pedal. Before this session begins I will ask you to report your level of fatigue based on the following scale:

1- Feeling active, vital, alert, or wide awake

2- Functioning at high levels, but not at peak; able to concentrate

3- Awake, but relaxed; responsive but not fully alert

4- Somewhat foggy, let down

5- Foggy; losing interest in remaining awake; slowed down

6- Sleepy, woozy, fighting sleep; prefer to lie down

7- No longer fighting sleep, sleep onset soon; having dream-like thoughts

8- Asleep

Afterward I will inform you that the session is beginning and will last 60 minutes. After every 5 minutes I will ask you to verbally report your level of fatigue according to the same scale. This will continue until the session is complete. After 60 minutes have elapsed I will inform you that the session is complete, you may remove the EEG headset. While driving please attempt to maintain a speed of 50mph and refrain from talking during the simulation other than when you are indicating your level of fatigue. Please put the headset on, centering the electrode on your forehead."

Once the participant has placed the EEG on their head and the program headset sinks with the program the participant will be asked to verbally report their fatigue level and begin driving.

Load configuration file cloudy sky and boring run lane 1. Turn off second screen when running.

APPENDIX C

IRB Approval Letter

UNF
UNIVERSITY of NORTH FLORIDA.

Office of Research and Sponsored Programs
1 UNF Drive
Jacksonville, FL 32224-2665
904-620-2455 FAX 904-620-2457
Equal Opportunity/Equal Access/Affirmative Action Institution

MEMORANDUM

UNF IRB Number: 475514-4
Amendment Approval: 11-1-2016
Expiration Date: 11-1-2017
Processed on behalf of UNF's IRB EB

DATE: November 1, 2016

TO: Dr. Dong-Yuan Wang
Psychology

FROM: Dr. Jennifer Wesely, Chairperson
On behalf of the UNF Institutional Review Board

RE: Amendment review by the UNF Institutional Review Board
IRB#475514-4: "Driving and Music"
Original approval: 10/18/2013

This is to advise you that the proposed modifications to your existing project, "Driving and Music" were reviewed and approved on behalf of the UNF Institutional Review Board.

Although, your project was originally approved as "Exempt" Category 2 on 10/18/2013 (declared to be exempt from further review by the UNF IRB), the IRB member who reviewed your amendment request determined that your project should be reclassified as "Expedited" Categories 4, 6 & 7. The need to elevate the review type stems from recent information gleaned from the Office of Human Research Protection (OHRP), explaining that "Exempt" Category 2 cannot include interventions (i.e., manipulations). Now that your project is "Expedited," please note that your project must undergo continuing review every year.

Your reviewer approved your amendments on behalf of the UNF Institutional Review Board to include the following:

1. Removing Mr. Zachary Jimison from the protocol
2. Changing the Principal Investigator from Mr. Zachary Jimison to Dr. Dong-Yuan Wang
3. Adding two additional personnel: Dr. Karthikeyan Umapathy and Mr. Lucas Coffey, and submitting CITI Completion Reports for each

4. Adding a new phase to the project: Phase 3
5. Updating all applicable documents with the proposed modifications

Your study has been approved for a period of 12 months as of 11/1/2016. Please submit a Status Report and other continuing review information to the UNF IRB prior to **10/1/2017** if you would like your study to continue past the 1-year anniversary of the approval date. *We ask that you submit your status report and other continuing review information at least 30 days before the expiration date as noted above to allow time for review and processing*. When you are ready to close your project, please complete a Closing Report Form. The status report and updated documents or closing report will need to be added via a new package in IRBNet. All records relating to this research shall be retained for at least 3 years after completion of the research.

This approval applies to your project in the form and content as submitted to the IRB for review. You may use the electronic *and* the signed informed consent procedures for participants as outlined in your approved documents. Only the approved versions of the consent information should be used during this research. All participants must receive a stamped and dated copy of the approved informed consent document when possible.

Any variations or modifications to the approved procedures or documents must be cleared with the IRB prior to implementing such changes. For *example*, if you plan to make changes to your stamped and dated informed consent form, it will be necessary to submit a copy of the revised form via an amendment so that it can be reviewed and approved prior to use. Once approved, a new stamp and date will be included on the revised consent form so that it can be used. To submit an amendment, please complete an Amendment Request Document and submit it along with any updated documents affected by the changes via a new package in IRBNet. Any unanticipated problems involving risk and any occurrence of serious harm to subjects and others shall be reported by completing this Event Report Form and sending it promptly to the IRB within 3 business days.

CITI Training for this Project:

Name	CITI Expiration Date
Dr. Dong-Yuan Wang	2/26/2018
Dr. Karthikeyan Umapathy	7/12/2018
Mr. Lucas Coffey	6/20/2019

CITI Course Completion Reports are valid for 3 years. The CITI training for renewal will become available 90 days before your CITI training expires. Please renew your CITI training when necessary and ensure that all key personnel maintain current CITI training. Individuals can access CITI by following this link: http://www.citiprogram.org/. Should you have questions regarding your project or any other IRB issues, please contact the research integrity unit of the Office of Research and Sponsored Programs by emailing IRB@unf.edu or calling (904) 620-2455.

This letter will be retained within UNF's records. All records shall be accessible for inspection and copying by authorized representatives of the department or agency at reasonable times and in a reasonable manner.

UNF IRB Number: 475514-4
Amendment Approval: 11-1-2016
Expiration Date: 11-1-2017
Processed on behalf of UNF's IRB

APPENDIX D

Informed Consent Form

This research is being conducted by faculty and students at the University of North Florida (UNF) to study driver fatigue. We will do this by collecting brain wave data via an EEG headset. You, as a participant, will be asked to drive via a driving simulator and verbally report your level of fatigue every 5 minutes. However, we will not attempt to individually identify you and you will not be identified in any reports or publications that comes from this research. We expect that participation in this study will take about 1 hour and 15 minutes of your time.

Participation is voluntary and there are no penalties for deciding not to participate. You may choose not to participate in this research without negatively impacting your relationship with UNF or your instructor. You may, at any time, decline to participate in this study. This study involves a driving simulator. If you have a history of photosensitive epilepsy or some other condition that may be aggravated by using a video game-like interface for an extended time, you are advised not to participate in this study. Please inform the experimenter(s) if you have such a condition.

Although simulated crashes (i.e. approaching an object, crashing noises being heard, and the screen showing a shattered windshield) are not a necessary part of this study, they may be involved in any one participant's experience. As such, if you feel that this may be traumatic for you (e.g., you or a loved one have been involved in an automobile accident), you are advised not to participate in this study. No other potential risks other than loss of time are anticipated.

No specific personal benefits other than the possible novelty and entertainment of using a driving simulator and/or wearing an EEG headset are anticipated. If you have any questions or concerns about this study, please contact Lucas Coffey at

or Dr. Umapathy at .

Any questions or concerns about a research-related injury may also be directed to the chair of the UNF Institutional Review Board (IRB) at 904-620-2498 or by e-mailing IRB@unf.edu.

You must be at least 18, either licensed to drive a vehicle or licensable.

I, ___________________________________, understand and agree to the above.

________________________Date

APPENDIX E

Agglomerative Hierarchical Clustering

12.1 Ratio 1 with 2 Clusters

Ratio	Percent Change	User	Cluster
R1	15.85	2	cluster_0
R1	19.63	3	cluster_0
R1	26.82	4	cluster_0
R1	16.30	5	cluster_0
R1	29.17	6	cluster_0
R1	4.97	1	cluster_1

Table 18. Normalized Ratio 1 with 2 Clusters Results

12.2 Ratio 1 with 3 Clusters

Ratio	Percent Change	User	Cluster
R1	15.85	2	cluster_0
R1	19.63	3	cluster_0
R1	16.30	5	cluster_0
R1	4.97	1	cluster_1
R1	26.82	4	cluster_2
R1	29.17	6	cluster_2

Table 19. Normalized Ratio 1 with 3 Clusters Results

12.3 Ratio 2 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R2	9.51	1	cluster_0
R2	40.65	3	cluster_0
R2	60.17	4	cluster_0
R2	61.14	5	cluster_0
R2	5.89	6	cluster_0
R2	-28.91	2	cluster_1

Table 20. Normalized Ratio 2 with 2 Clusters Results

12.4 Ratio 2 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R2	40.65	3	cluster_0
R2	60.17	4	cluster_0
R2	61.14	5	cluster_0
R2	-28.91	2	cluster_1
R2	9.51	1	cluster_2
R2	5.89	6	cluster_2

Table 21. Normalized Ratio 2 with 3 Clusters Results

12.5 Ratio 3 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R3	3.50	1	cluster_0
R3	13.28	5	cluster_0
R3	13.76	3	cluster_0
R3	17.73	2	cluster_0
R3	26.36	4	cluster_0
R3	37.71	6	cluster_1

Table 22. Normalized Ratio 3 with 2 Cluster Results

12.6 Ratio 3 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R3	17.73	2	cluster_0
R3	13.76	3	cluster_0
R3	26.36	4	cluster_0
R3	13.28	5	cluster_0
R3	3.50	1	cluster_1
R3	37.71	6	cluster_2

Table 23. Normalized Ratio 3 with 3 Clusters Results

12.7 Ratio 4 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R4	11.96	1	cluster_0
R4	-26.87	2	cluster_0
R4	48.20	3	cluster_1
R4	60.89	4	cluster_1
R4	68.49	5	cluster_1
R4	1.23	6	cluster_0

Table 24. Normalized Ratio 4 with 2 Clusters Results

12.8 Ratio 4 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R4	-26.87	2	cluster_2
R4	1.23	6	cluster_1
R4	11.96	1	cluster_1
R4	48.20	3	cluster_0
R4	60.89	4	cluster_0
R4	68.49	5	cluster_0

Table 25. Normalized Ratio 4 with 3 Clusters Results

12.9 Ratio 5 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R5	-13.38	2	cluster_1
R5	10.28	1	cluster_0
R5	16.35	6	cluster_0
R5	32.64	3	cluster_0
R5	33.89	4	cluster_0
R5	35.21	5	cluster_0

Table 26. Normalized Ratio 5 with 2 Clusters Results

12.10 Ratio 5 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R5	-13.38	2	cluster_1
R5	10.28	1	cluster_0
R5	16.35	6	cluster_0
R5	32.64	3	cluster_2
R5	33.89	4	cluster_2
R5	35.21	5	cluster_2

Table 27. Normalized Ratio 5 with 3 Clusters Results

12.11 Ratio 6 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R6	16.58	2	cluster_0
R6	21.34	3	cluster_0
R6	28.79	4	cluster_0
R6	16.08	5	cluster_0
R6	30.24	6	cluster_0
R6	5.49	1	cluster_1

Table 28. Normalized Ratio 6 with 2 Clusters Results

12.12 Ratio 6 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R6	5.49	1	cluster_1
R6	16.08	5	cluster_0
R6	16.58	2	cluster_0
R6	21.34	3	cluster_0
R6	28.79	4	cluster_2
R6	30.24	6	cluster_2

Table 29. Normalized Ratio 6 with 3 Clusters Results

12.13 Ratio 7 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R7	-28.53	2	cluster_0
R7	3.16	6	cluster_0
R7	10.54	1	cluster_0
R7	43.73	3	cluster_1
R7	63.82	4	cluster_1
R7	64.51	5	cluster_1

Table 30. Normalized Ratio 7 with 2 Clusters Results

12.14 Ratio 7 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R7	43.73	3	cluster_0
R7	63.82	4	cluster_0
R7	64.51	5	cluster_0
R7	10.54	1	cluster_1
R7	3.16	6	cluster_1
R7	-28.53	2	cluster_2

Table 31. Normalized Ratio 7 with 3 Clusters Results

12.15 Ratio 8 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R8	12.26	1	cluster_0
R8	-27.09	2	cluster_0
R8	-0.04	6	cluster_0
R8	49.14	3	cluster_0
R8	64.24	4	cluster_0
R8	69.67	5	cluster_0

Table 32. Normalized Ratio 8 with 2 Clusters Results

12.16 Ratio 8 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R8	-27.09	2	cluster_2
R8	-0.04	6	cluster_1
R8	12.26	1	cluster_1
R8	49.14	3	cluster_0
R8	64.24	4	cluster_0
R8	69.67	5	cluster_0

Table 33. Normalized Ratio 8 with 3 Clusters Results

12.17 Ratio 9 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R9	-38.17	1	cluster_1
R9	-14.70	2	cluster_0
R9	-9.66	4	cluster_0
R9	4.82	6	cluster_0
R9	13.92	5	cluster_0
R9	31.74	3	cluster_0

Table 34. Normalized Ratio 9 with 2 Clusters Results

12.18 Ratio 9 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R9	-38.17	cluster_1	cluster_1
R9	-14.70	cluster_0	cluster_0
R9	-9.66	cluster_0	cluster_0
R9	4.82	cluster_0	cluster_0
R9	13.92	cluster_0	cluster_0
R9	31.74	cluster_2	cluster_2

Table 35. Normalized Ratio 9 with 3 Clusters Results

12.19 Ratio 10 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R10	34.54	2	cluster_0
R10	84.95	6	cluster_0
R10	107.52	3	cluster_0
R10	145.47	1	cluster_0
R10	154.27	5	cluster_0
R10	205.56	4	cluster_1

Table 36. Normalized Ratio 10 with 2 Clusters Results

12.20 Ratio 10 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R10	34.54	2	cluster_2
R10	84.95	6	cluster_0
R10	107.52	3	cluster_0
R10	145.47	1	cluster_0
R10	154.27	5	cluster_0
R10	205.56	4	cluster_1

Table 37. Normalized Ratio 10 with 3 Clusters Results

12.21 Ratio 11 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R11	-41.53	5	cluster_1
R11	-28.99	1	cluster_1
R11	-16.43	4	cluster_1
R11	-12.74	2	cluster_1
R11	36.88	6	cluster_0
R11	60.59	3	cluster_0

Table 38. Normalized Ratio 11 with 2 Clusters Results

12.22 Ratio 11 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R11	-41.53	5	cluster_0
R11	-28.99	1	cluster_0
R11	-16.43	4	cluster_0
R11	-12.74	2	cluster_0
R11	36.88	6	cluster_2
R11	60.59	3	cluster_1

Table 39. Normalized Ratio 11 with 3 Clusters Results

12.23 Ratio 12 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R12	-51.16	2	cluster_0
R12	-25.06	6	cluster_0
R12	-7.916	5	cluster_0
R12	0.40	1	cluster_0
R12	21.49	4	cluster_0
R12	93.60	3	cluster_1

Table 40. Normalized Ratio 12 with 2 Clusters Results

12.24 Ratio 12 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R12	-51.16	2	cluster_2
R12	-25.06	6	cluster_0
R12	-7.916	5	cluster_0
R12	0.40	1	cluster_0
R12	21.49	4	cluster_0
R12	93.60	3	cluster_1

Table 41. Normalized Ratio 12 with 3 Clusters Results

12.25 Ratio 13 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R13	-69.26	1	cluster_0
R13	-59.16	2	cluster_0
R13	-53.15	4	cluster_0
R13	-50.60	3	cluster_0
R13	-48.61	8	cluster_1
R13	-5.74	9	cluster_1

Table 42. Normalized Ratio 13 with 2 Clusters Results

12.26 Ratio 13 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R13	-69.26	2	cluster_2
R13	-59.16	5	cluster_0
R13	-53.15	6	cluster_0
R13	-50.60	1	cluster_0
R13	-48.61	4	cluster_0
R13	-5.74	3	cluster_1

Table 43. Normalized Ratio 13 with 3 Clusters Results

12.27 Ratio 14 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R14	-27.82	1	cluster_0
R14	-17.51	5	cluster_0
R14	-11.94	2	cluster_0
R14	-8.45	4	cluster_0
R14	39.05	6	cluster_0
R14	136.61	3	cluster_1

Table 44. Normalized Ratio 14 with 2 Clusters Results

12.28 Ratio 14 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R14	-27.82	1	cluster_0
R14	-17.51	5	cluster_0
R14	-11.94	2	cluster_0
R14	-8.45	4	cluster_0
R14	39.05	6	cluster_2
R14	136.61	3	cluster_1

Table 45. Normalized Ratio 14 with 3 Clusters Results

12.29 Ratio 15 with 2 Clusters Results

Ratio	Percent Change	User	Cluster
R15	37.85	2	cluster_0
R15	105.49	5	cluster_0
R15	133.66	6	cluster_0
R15	173.00	1	cluster_0
R15	227.59	4	cluster_0
R15	326.17	3	cluster_1

Table 46. Normalized Ratio 15 with 2 Clusters Results

12.30 Ratio 15 with 3 Clusters Results

Ratio	Percent Change	User	Cluster
R15	37.85	2	cluster_2
R15	105.49	5	cluster_0
R15	133.66	6	cluster_0
R15	173.00	1	cluster_0
R15	227.59	4	cluster_0
R15	326.17	3	cluster_1

Table 47. Normalized Ratio 15 with 3 Clusters Results

VITA

Lucas B. Coffey is currently working as a Software Engineer in the IT Department of one of the leading transportation companies in Jacksonville, Florida. He is a U.S. Navy veteran of five years, and has three years of civilian experience in the IT industry, mainly in Java/J2EE Applications. He holds a Bachelors of Science in Electronic Systems Technologies from the University of Southern Illinois. He is a self-motivated, team player as well as an independent worker, with a background in leadership and technology. He is always ready to learn and adapt to new tools and technologies. He is originally from Tennessee and currently lives in Jacksonville, Florida.