An investigation into the relationship between saccadic intrusion and mental workload

AN INVESTIGATION INTO THE RELATIONSHIP BETWEEN SACCADIC INTRUSIONS AND MENTAL WORKLOAD

A Dissertation by

Satoru Tokuda

Bachelor of Arts, University of Illinois Urbana-Champaign, 2004

Master of Arts, Wichita State University, 2008

Submitted to the Department of Psychology
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirement for the degree of
Doctor of Philosophy

December 2009

AN INVESTIGATION INTO THE RELATIONSHIP BETWEEN SACCADIC INTRUSIONS AND MENTAL WORKLOAD

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Psychology.

Alex Chaparro, Committee Chair

Barbara Chaparro, Committee Member

Edgar C. Merkle, Committee Member

Evan Palmer, Committee Member

David W. Downs, Committee Member

Accepted for the Fairmount College of Liberal Arts and Sciences

William D. Bischop, Dean

J. David McDonald, Dean

DEDICATION

To my wife Maja and our daughter Akira

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to the people, organizations, and the countries that gave me the possibility to complete this dissertation.

I want to thank my internship supervisor, Dr. Goro Obinata in Nagoya University, Japan. Without experiencing in his lab, I would not have had this research idea in this dissertation.

I want to thank my supervisors, Dr. Alex Chaparro and Dr. Evan Palmer, for their guidance, suggestions, encouragement, and extensive help on developing the algorithms, theoretically supporting the validity of the research, and writing this dissertation. To my committee members Dr. Barbara Chaparro, Dr. Edgar C. Merkle, and Dr. David W. Downs, I am grateful for their support, ideas, hints, and encouragement.

I want to thank my graduate colleagues from the Human Factors program here at WSU for their exchanging research ideas and participating in my studies as pilot participants. I believe some of the preliminary, crude pilot studies were almost torturous for some of you, mentally and physically. Thank you for the feedback on the studies. I want to show my appreciation toward those who proofread many pages of this dissertation. These grad students include Kelsi Lenz, Justin Owens, Shiva Naidu, Doug Fox, Chris Brown, Nichole Morris, Amanda Smith, Adam Probst, Lukas Hulsey, and Bobby Nguyen.

I want to thank the writing center tutors. Because of the nature of my English as a second language, they helped me a total of hundreds of hours in the last 5 years in WSU. Especially, there were six core tutors. Each of them helped me easily more than ten hours on my writings with their logical feedback and comments. These six were Parker Daniel, Cara Ernzen, Ashley Muench, Andrew Parker, Sharon Zeller, and Katherine Reichenberger.

I want to express my thanks to my wife Maja and our family for helping my research, writing, reading, as well as taking care of the family, and especially undergoing her whole period of pregnancy and raising the baby in a different state without my physical support. Without her determination, I would not have been able to even start my dissertation 16 months ago.

This dissertation is a product of the help by all the people I listed and many other people who indirectly helped the project. Thank you all.

ABSTRACT

This dissertation reports a new method for estimating mental workload (MWL) using a specific type of eye movement called saccadic intrusions (SIs). These studies demonstrate in laboratory conditions that there is a consistent relationship between MWL and SI eye behavior. The experiments manipulated MWL using the auditory N-back task, which quantitatively varies short-term memory load, thus manipulating MWL. While MWL was being manipulated, participants were engaged in a controlled visual task. Eye movements were recorded during the visual task and analyzed to determine whether participants had increased SIs as a function of MWL. The author developed a novel algorithm to quantify SI eye behavior; this algorithm takes time-series eye movement data as input, automatically searches for the characteristics of SI eye behavior in the data, accumulates occurrences of SI eye behavior (accounting for their amplitudes and dwell time), and quantifies the SI eye behavior into a single value, called the SI measure. Besides the SI measure, pupil diameter was used as a second dependent variable because it was already known to reliably reflect MWL.

The analyses revealed that as MWL increased, both pupil diameter and the SI measure increased also. Correlation analyses were conducted across 12 one-minute trials with each participant; an average correlation value across 37 participants was $r = 0.57$ for the MWL-pupil diameter relationship. The other average correlation value was $r = 0.45$ for the MWL-SI measure relationship. Multiple linear regression analyses were conducted to predict MWL from the eye movement data. The R^2 values from these analyses were 0.54 or higher; both pupil diameter and the SI measure significantly contributed to the prediction of MWL.

These results indicate that SI eye behavior reliably reflects MWL. While pupil diameter does not remain accurate in reflecting MWL changes during illumination changes in the environment, SI eye behavior should continue to reflect MWL without being affected by illumination changes. This new measure for estimating MWL may be useful in evaluating a vehicle driver's MWL in real-time despite illumination changes.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

Chapter Page

TABLE OF CONTENTS (continued)

LIST OF ABBREVIATIONS / NOMENCLATURE

DV	dependent variable
FEM	fixational eye movement
FFT	fast Fourier transformation
FT	Fourier transformation
ISI	inter-stimuli interval
IV	independent variable
MWL	mental workload
OKR	optokinetic reflex
SED	saccadic eye deviation
SEM	saccadic eye movement
SI	saccadic intrusion
VOR	vestibulo-ocular reflex

Units

deg	degrees
min	minute (60 min = 1 deg)
s	seconds
ms	milliseconds (1000 ms = 1 s)
mm	millimeter (1000 mm = 1 m; 25 mm = 1 inch)

CHAPTER 1

INTRODUCTION

This dissertation offers a new method to estimate mental workload (MWL) using a combination of two types of eye activity: pupil diameter and saccadic intrusions (SIs). While the relationship between pupil diameter and MWL has been known for at least the last five decades, the relationship between SIs and MWL is a new finding that has not been reported in other literature.

1.1 Experimental Design

This dissertation is the first stage in a planned series of MWL estimation research projects, and thus does not reach the final goal. The final goal develops a measure using pupil diameter and SIs precisely estimates in real-time how much the operator is cognitively overloaded or how dangerous a driver of a vehicle is to themselves and others in maintaining their present activity. Instead, this dissertation focuses on the basic level of human science; the dissertation retrospectively (not in real-time) examined the relation between one kind of MWL task (i.e., the N-back task) and several kinds of physiological measures including pupil diameter, SIs, and others in a laboratory condition (not in a real world driving environment). To be specific, the N-back task was manipulated as the independent variable (IV) into simulating three levels of MWL (low, medium, and high). The physiological measures, treated as the dependent variables (DVs), were analyzed to determine if they changed depending on the manipulated three levels of MWL.

1.2 Mental Workload

Chapter 2 (Literature Review) describes each of the elements in this dissertation, such as MWL and physiological responses. The first half of the chapter explains MWL, especially emphasizing that MWL estimation is important to human-machine interaction systems because it may detect signs of cognitive overload in eye activity before the cognitive overload actually manifests in a physical error. For example, in a driving environment, a driver’s eye activities (such as SIs and pupil diameter) may reveal signs of cognitive overload. The driver’s cognitive performance is now sub-optimal and may be more prone to having an accident. In this case it would be wise for the driver to stop the car or shed other tasks thereby reducing their crash risk. For example, sometimes a driver may focus too much on cell phone conversation, leaving insufficient cognitive resources for safe driving. Recarte and Nunes (2002) showed that cell-phone conversation increased the chance of having an accident by four times. Other activities that increase MWL include text-messaging, navigating to an unfamiliar destination and a lingering mind from a conversation five minutes ago. MWL estimation may be able to detect these mental conditions of the driver before the driver actually makes a dangerous mistake.

This dissertation examines the relationship between cognitively overloaded conditions and physiological responses. In doing so, this study simulated a cognitively overloaded condition by giving a participant a mentally demanding task, the N-back task. The IV was the N-back levels simulating a high MWL condition. The DVs were the physiological responses, which are explained in the second half of Chapter 2.

1.3 Rationale of This Dissertation

The second half of Chapter 2 explains that in this dissertation, SIs served as one of the DVs. SIs are saccadic (rapid and jerky) eye deviations from the fixation baseline that occur during a period of otherwise relatively stable fixation. This dissertation's area of study arose from the hypothesis that SIs would increase as MWL increased. This hypothesis was derived from the author’s observations in a previous study.

Before this dissertation, the author was engaged in studies to estimate MWL using another type of eye movement: vestibulo-ocular reflex (VOR) (Tokuda, Obinata, Usui, Inuzuka & Hamada, 2007; Obinata, Tokuda & Shibata, 2008; Obinata, Tokuda & Fukuda, 2009). VOR is a reflexive eye movement that cancels out head movement by employing a counteractive eye movement, so that the retina maintains fixation on a single point in space. The algorithm in the studies calculated the gap between the actual VOR movements and the predicted VOR movements as a function of the head movements. The studies show that when MWL was high, VOR deviated more from the expected VOR movement as a function of the head movements. The VOR studies are promising because VOR is a reflex movement, and the reaction time of the reflex takes only a few milliseconds (Robinson, 1981). The discrepancy between predicted VOR and actual VOR responses holds promise as a tool for evaluating operator MWL.

While analyzing the VOR data, the author noticed that there were occasionally large amplitude eye deviations among small amplitude VOR deviations. Unlike VOR deviations, the large amplitude deviations were of the saccadic type, with a dwell time of only a few hundred milliseconds. This phenomenon was observed once in a while during a high MWL condition in the VOR studies. It was at this point that the author started having an interest in the relationship between saccadic eye deviations (such as SIs) and MWL.

After these observations, the author learned that there were studies examining the relationships between saccadic eye deviations (SEDs) and cognitive activity. There are two types of SEDs; microsaccades and SIs. Studies on these two types (microsaccades and SIs) show reliable evidence that there are links between SEDs (microsaccades and SIs) and cognitive attention (Engbert, 2006; Gowen, Abadi, Poliakoff, Hansen & Miall, 2006). Since cognitive attention is a concept closely related to MWL, it can lead to the connection between SEDs and MWL.

Chapter 2 explains each of these elements including MWL, SEDs, and the suspected link between these two elements, Chapters 3 and 4 explain how the dissertation examines if SEDs (especially SIs) are related to MWL.

1.4 Algorithm to Quantify Saccadic Intrusions

Since neither MWL or SIs are numerical concepts, they needed to be converted into measurable values for the dissertation experiments. MWL levels were quantified into values using the N-back tasks. The behavior of SIs was quantified into a representative value using an algorithm. The author developed an algorithm that takes raw eye movement data as the input, detects the characteristics of SIs in the eye movements, and quantifies the behavior of SIs into a SI measure. The algorithm is described in detail in chapter (Chapter 3) of this dissertation.

The purposed of the algorithm was to automatically isolate SIs in eye movement recordings. More specifically, the algorithm automatically removed all other types of eye movements but SIs. SIs represent one of ten major types of eye movements, including regular saccades, saccadic intrusions (SIs), microsaccades, slow drifts, tremors, gaze aversion, smooth pursuits, vestibulo-ocular reflex (VOR), optokinetic reflex (OKR), convergence, and some other minor types of eye movements.

The first step of the algorithm was to classify the types of eye movements. This is challenging, but not impossible, since different types of eye movements usually have different amplitudes, angular velocities, duration, dwell time, repeating cycles, behavior, and orientations. The classification of eye movements was made easier, by employing several constraints. First a chin rest was used which effectively eliminated eye movements related to head movements

(VOR and OKR) in the data collected. Second, the visual tasks used in the three experiments were carefully designed so that they constrained different types of eye movements.

1.5 Experiments

This dissertation was conducted using the study settings explained above, implementing the N-back task as the IV and pupil diameter and SIs as the DVs. As explained above, this dissertation represents basic level research, which limits generalizability but increased experimental control. The MWL task was limited to one kind of task, the N-back task. Head movements were restricted during the experiments. The types of eye movements were limited to several kinds in the three studies. These limitations allowed better characterization of the behavioral changes in SIs.

The pupil diameter and the SI measures were found to be the most effective physiological responses for estimating MWL among the seven possible physiological responses. The results and future goals are found in the results and discussion chapters at the end of the dissertation.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to Literature Review

2.1.1 Goal of this dissertation

This dissertation draws from the fields of psychology, physiology, psychophysics, and calculation methods; it is essential for readers to know the terminology and past research upon which it is based. This literature review is divided into two sections: mental workload (MWL) and eye activity.

The first section addresses the concept of MWL. One of the goals in this dissertation is to demonstrate that a new MWL measure (i.e., saccadic intrusion measure) increases as MWL increases. This means that the experiment needs to be able to manipulate MWL levels as an independent variable (IV). The MWL section will explain the nature of MWL, how it is manipulated, and how it is measured physiologically. The second section will elaborate two physiological measures of MWL, saccadic intrusions (SIs) and pupil diameter, which serve as the dependent variables (DVs) in this dissertation.

This research addresses a basic research issue: whether increases in MWL consistently induce specific involuntary physiological responses such as pupil dilation and large SIs across trials and across individuals. At the same time, this dissertation has the long term goal of applying the basic physiological results to the monitoring of MWL in real world environment such as driving. This is to say, if specific eye activities can accurately estimate MWL, the technique can be used to estimate a vehicle operator’s MWL and detect dangerous states where the driver is mentally overloaded. Again, this application to the automotive industry is a future goal.

2.2 Mental Workload

2.2.1 Introduction

The goal of this review section is to illustrate current trends in MWL research. In a nutshell, none of the current MWL measurements are practically useful for accurately estimating a driver's MWL in real-time.

The MWL literature review begins with describing the significance of MWL, which is as a measure of how cognitively burdened an individual is. The next section describes the N-back task, which is used to manipulate MWL. Using the N-back task, the experiments in this project simulated states of increased cognitive load. At the end of the MWL literature section, pre-existing MWL measurement techniques are briefly overviewed in terms of why they are not practically usable to estimate a driver's MWL.

2.2.2 Definition of Mental Workload

In psychology, especially in Human Factors, "mental workload" or just "workload" is defined as a function of two factors: the operator's capacity and the task demands (Gopher & Donchin, 1986). Tsang and Vidulich (2006) stated, "[a] commonly accepted notion is that mental workload is very much a function of the supply and demand of attentional or processing resources" (p. 246). Similarly, Johnson and Proctor (2004) defined "[mental workload] as the difference between available and required capacities" (p. 262). For example, the required mental capacity for driving a car on a highway on a clear, sunny day may be relatively constant. However, available mental capacities for driving may vary depending on the usage of a cell phone or some other multitasking activity.

This example shows that MWL can vary depending on available and demanded mental capacities for a task, as explicitly mentioned in the definition. In addition, implicit in the definition, the same task can give different MWL levels to different individuals at different points in time because there are individual differences in available and required capacities for the task. In other words, MWL is not a measure of the amount of work, but a measure of how close an individual is to cognitive overload at a given moment.

2.2.3 MWL and Common Resource Models

The concept of MWL is directly linked to common resource models (Proctor & Van Zandt, 1993; Johnson & Proctor, 2004) such as the unitary resource model (Kahneman, 1973), the multiple resource model (Wickens, 1984, 1991, 2004) and others. Common resource models are based on some fundamental concepts: our cognitive activities share common mental resources, mental resources are limited in their capacities, time-sharing mental activities can, at times, not have access to enough mental resource, which can cause a sub-optimal level of cognitive performance or behavior.

The concept of the common resource models is generally accepted in Human Factors and related areas (Proctor & Van Zandt, 1993, Wickens, 2007). Research in most areas of Human Factors, cognitive psychology, and human information processing rely on the common resource models. Examples of common resource models include Baddeley and Hitch's (1974) working memory model, Just and Carpenter's (1992) limited resource model, Kahneman's (1973) capacity model, and Wickens' (2004) multiple resources model. All of these models agree that some mental activities share common mental resources and interfere with each other in time-sharing performance.

2.2.4 MWL as Performance Estimation

The common resource models assume that mental resources are shared for all activities that involve the brain's control. This includes the somatic nervous system as it is related to sensory neurons and motor neurons. However, this does not include the autonomic nervous system. When a MWL measure shows that the individual is cognitively overloaded, or when the brain is not processing all necessary information appropriately, that is not just a sign of performance impairment for that single cognitive task, but also a sign that all cognitive task performance is impaired at the moment. All of the operator's cognitive performance is at risk of deteriorating at that moment. This leads to the notion that MWL is a concept that is strongly related to the prediction of human performance (Gopher & Donchin, 1986; Proctor & Van Zandt, 1993; Tsang & Vidulich, 2006).

The definition by Tsang and Vidulich (2006) explicitly states that the difference between the supply and demand of mental resources determines MWL. This definition implies that MWL is said to be heavy when the available cognitive resources are scarce, and therefore, the individual is prone to making a cognitive error due to a cognitive state with low available resources. The term, MWL, itself does not advertise this importance, but the three concepts, MWL, cognitive resource scarcity, and a tendency to make cognitive errors due to limited available cognitive resources are three aspects of the same mental status. They all link to or lead to poor cognitive task performance. MWL in Human Factors is a concept for cognitive performance estimation.

Throughout this dissertation, the term "mental workload," or "MWL," is used for this concept: the hypothetical ratio of the required mental capacity to the originally available mental capacity for the person. A heavy MWL state indicates that the required capacity is high, and the available capacity is low. A low MWL state indicates that the required capacity is low, and there is plenty of available capacity left for that task or additional tasks.

2.2.5 N-back Task

In the experiments in this dissertation, a new MWL measure (the SI measure) was tested using multiple levels of MWL. If the new MWL measure shows consistent change according to the MWL level changes, it could be concluded that the new MWL measure reflected MWL in a significant manner. In order to demonstrate a relationship between SIs and MWL, there needed to be a way to manipulate MWL levels. Since MWL is a hypothetical concept, an experimenter could not directly manipulate MWL – A cognitive task that could impose MWL on participants was needed. The N-back task was chosen to operationalize MWL. This section describes the N-back task and their advantages for systematically varying MWL.

The N-back task is a set of tasks that can quantitatively manipulate MWL. The N-back task is common in psychology research related to working memory, mental workload, cognitive load, and studies of information processing (Murray, McFarland & Geffen, 2005; Jaeggi, Buschkuehl, Etienne, Ozdoba, Perrig & Nirkko, 2007).

The N-back task is a type of probe-reaction-time task. A probe-reaction-time task (Posner and Boies,1971), has participants judge if two stimuli, presented consecutively in a timely manner, were the same or different. Unlike the probe-reaction-time task, the N-back task can control the timing between two stimuli so that the tasks can manipulate memory load or mental effort to recall the stimulus. The letter “N” represents any whole numbers from 0 to infinity, but common N-back tasks are 0, 1, 2, and 3-back tasks (e.g. McElree, 2001; Murray et al., 2005; Jaeggi et al. 2007). In an N-back task, single letter stimuli are presented one by one every few seconds. Participants are instructed to answer if the current stimulus is the same as another stimulus N-events before. Stimulus modality can be auditory or visual. A stimulus set can be single-digit numbers, alphabet letters, colors, locations, or something else. For example, in an auditory 2-back task using a random number stimuli from 1 to 4, the stimuli may be presented as "4, 1, 2, 1, 2, 4, 3," and the correct answers are "blank, blank, different, same, same, different, and different." The MWL of the N-back tasks can be measured by the N levels as an analytical MWL measure (which will be explained later in this chapter). A higher number indicates higher MWL. The N-back task performance can be also measured by reaction time and correct response by percent.

The N-back task is commonly used in MWL studies because of the seven following advantages. First, N-back task monotonically increases MWL as the N increases (Murray et al., 2005). This means that the N-back task can quantitatively manipulate MWL in multiple levels, rather than qualitatively. Quantitative changes mean that difficulty levels change by degrees. On the other hand, qualitative changes mean that difficulty levels are different categorically. For example, these two tasks, one to search for a vertical line and another to search for a horizontal line among visual noise on a computer screen, are qualitatively different because the difference is categorical.

For the 0-back task, the memory load is counted as zero since the current number and the number 0-events previous are always the same. The 0-back task is also called as a simple reaction task. Likewise, the 1-back task requires remembering one other number in order to correctly perform the task. The 2-back task requires remembering two other numbers. The memory load thus monotonically increases as the N increases, and therefore the MWL increases. This aspect of a MWL task is important in a study with a new MWL measure because monotonic changes in the DVs (eye activities) can easily be associated with the monotonic changes in MWL levels.

Second, all levels of N-back task can use the same stimulus set. This ensures that any changes in DVs are due to the MWL level difference, not due to the stimulus differences.

Third, the difficulty levels can be adjusted for each individual by assigning different N's and ISIs (Inter-Stimuli Interval or the auditory stimuli presentation rate). For those who are good at numbers and memorizing, the N can be higher numbers (such as the 4-back task), or ISI can be shorter than the baseline (which is usually around 3000ms). On the other hand, poor N-back performers can use lower Ns or slower ISIs.

Fourth, the N-back task is popular in MWL studies because MWL levels are stable during one trial (Murray et al., 2005). Since each N-back task always requires a specific number of remembered items the workload remains the same within a trial.

Similarly, as the fifth advantage of the N-back task, MWL levels remain the same across the trials in an N-back task for a participant. For example, the difficulty levels can be assumed to be the same between the first trial of the 3-back task and the second trial of the 3-back task, if there are enough practice trials before these experimental trials. This enables the "within-participant" experimental design, by repeating the same N-back task several times.

Sixth, N-back task timing can be control experimentally. Other popular MWL tasks such as arithmetic tasks can vary in answering time by question and participant. Unlike arithmetic tasks, each response of an N-back task takes up to only a few seconds, and trials can be done within a pre-defined time (e.g. about 50 seconds in this research). This makes analysis and comparison easy and strong.

Finally, the N-back task is popular in MWL studies because of the flexibility it affords the experimenter. For example, stimuli can be presented either the visually or auditorially, and operator responses can be obtained using verbal, keyboard, or other response modalities.. This dissertation exploited this flexibility, and used the auditory N-back task to manipulate MWL without affecting the performance of visual tasks. Overall, the advantages of the N-back task matched the needs of the dissertation research for each participant to experience three levels of MWL from light, to medium, to high.

2.2.6 Operationalization

Using the N-back task, MWL was quantitatively "operationalized." In other words, MWL is quantitatively defined to be equivalent to the N-back task levels. Thus the "operationalized" concept MWL can be treated as a quantitatively malleable variable in this dissertation, using the N levels such as 1, 2, 3, and 4.

In an ideal world, one should not conclude that the new MWL measure is correlated with MWL itself, because technically we don't know the exact MWL changes from 1-back to 2-back, and 2-back to 3-back tasks in each participant. However, the field of psychology routinely accepts this type of operationalized assumption. Similar assumptions can be seen in many psychology studies. For example, two separate studies (i.e., Ahlstrom & Friedman-Berg, 2006; Van Orden, Limbert, Makeig, & Jung, 2001) collapsed the results of 7 or 11 participants respectively and took the ANOVA and linear correlation analyses, using the physiological behavior on the Y-axis of the correlation plot, and the task complexity on the X-axis. Their physiological measures correlated significantly with task difficulty levels: R^2=0.697, p=0.001 for shorter blink duration and the number of aircraft in air traffic control operations; R^2=0.232, p=0.01 for pupil diameter increase and target density in visuo-spatial memory tasks. The author of this dissertation is not criticizing these two studies, but emphasizing that psychology is such an elusive area of study that we sometimes need to replace an unquantifiable scale such as MWL with a quantifiable scale such as N-back levels.

2.2.7 Interlude

So far, this mental workload (MWL) review section has explained the definition of MWL and manipulation of MWL as an independent variable (IV) in order to demonstrate MWL measure changes with MWL changes. The remaining section in this chapter focuses on another topic: MWL assessment as dependent variables (DVs).

2.2.8 MWL Measures

As explained above, MWL is a measure of how limited mental resources are, and poor performance on one cognitive task can be a sign of poor performance on other cognitive tasks. This aspect of MWL is used to predict cognitive performance in many MWL tasks including primary- and secondary-task performance measures, and some physiological measures. These are all considered empirical measures because they use empirical (experienced or observed) data instead of analytical measures. Before explaining details of these measures, an overview of MWL assessment techniques is provided.

For the last five decades, many researchers have tried to develop measures of MWL (Proctor & Van Zandt, 1993). There are many MWL measurements, but as of today, there is no practical measurement that can be readily used to measure a driver's MWL under real world conditions. There are several reasons that MWL estimation is difficult. The biggest challenge is that MWL is a hypothetical concept (Johnson & Proctor, 2004; Wickens & Hollands, 2000), and like other mental activities such as love, sleepiness, etc., it is difficult to quantify. Despite this challenge, the needs in industry to accurately measure MWL are high, and many MWL measures have been developed. The following paragraphs provide an overview the pre-existing MWL measures and the criteria to evaluate these MWL measures.

2.2.9 Criteria to Evaluate MWL Measures

Gopher and Donchin (1986) and Boff and Lincoln (1988) offered criteria to evaluate MWL measures. These criteria are sensitivity, reliability, intrusiveness, implementation requirements, and operator acceptance. These criteria help to clarify why the pre-existing MWL measurements are not good at estimating a driver's MWL.

Sensitivity refers to how well a measure can sense the operator's MWL. Reliability asks if the MWL measure can repeatedly produce similar results in similar situations. Intrusiveness asks if the MWL measure interferes with the operator's primary task, such as driving. Implementation requirements ask if the MWL measure is easy to implement in terms of costs, operators' training, or needs of experimenters to run the MWL measure. Lastly, operator acceptance asks if the MWL measure is easy for the operator in terms of psychological and physical easiness.

Using these criteria, the following section evaluates the pre-existing MWL measures. These MWL measures can be categorized using Lysaght, Hill, Dick, et al's (1989) taxonomy into two groups: analytical and empirical methods.

2.2.10 Analytical Techniques to Quantify MWL

Analytical techniques are the methods that quantify MWL by calculating theoretical task difficulty and theoretical cognitive capacities. For example, the N levels in the N-back tasks are an analytical MWL measure because they use only theoretical information without looking at empirical data. The biggest advantage of analytical MWL quantification techniques is the validity of quantification. The MWL values estimated by an analytical technique are more meaningful than the other MWL measures with an ordinal scale. For example, the N levels of the N-back tasks represent interval increments from 0 to 1, 1 to 2, 2 to 3, and so forth. This aspect of analytical technique is so important that this dissertation refers to the analytical MWL levels (i.e., the N levels) in the results section.

There is a side note that is less relevant to the dissertation. Usually the words quantification and estimation are used interchangeably. However, the author of this dissertation prefers to use the term “MWL quantification” for analytical techniques because they produce objective numbers for representing MWL. The focus here is to quantify. On the other hand, the author prefers to use the term “MWL estimation” for empirical techniques because they try to make the best estimate of MWL using available measures. Consequently, empirical methods usually offer more accurate prediction of cognitive performance deterioration (Wickens, 2007). In other words, the criterion of “sensitivity” is not always satisfied with analytical techniques.

2.2.11 Empirical Techniques to Estimate MWL

Empirical techniques are the methods used to estimate MWL using operators’ responses. Lysaght’s et al. (1989) taxonomy has four categories under the empirical techniques: primary task performance measures, secondary task performance measures, subjective rating measures, and physiological measures (Figure 2.1).

- Primary task Performance measures
  - Driving lane position
  - Reaction time
  - Correct percent
- Secondary task Performance measures
  - Road-sign search task
- Rating techniques
  - Single scale rating
  - NASA-TLX (Hart, 1989)
- Physiological response
  - Heart rate, respiratory rate, blood pulse rate
  - Pupil diameter, blink rate
  - Eye movements

Figure 2.1. Taxonomy of empirical MWL measure techniques with some examples.

Primary-task performance measures estimate the operator's MWL using a performance score on the primary task in which the operator is engaged. Some examples include reaction time, correct response percent, and driving lane position. None of these, as of today, are practical in estimating a driver's MWL because of either a lag or delayed response to changes in MWL, inaccuracy of estimating MWL, or having not enough samples to measure MWL effectively. These are considered a violation of the "sensitivity" criterion; primary-task performance measures are not sensitive enough to estimate a driver's MWL in real-time, or with enough accuracy, or every time the estimation is needed.

For example, Hancock, Lesch and Simmons (2003) examined the reaction time for a stop light when the participants were either using a cell phone or not. This kind of MWL measure is great for examining the effects of cell phone usage because it is an objective measure. The experiments in this dissertation have two examples of primary-task performance MWL measures: reaction time to the N-back auditory stimuli and the correct percent of the N-back task. Both of these measures were used to confirm the analytical MWL levels (i.e., the N levels). However, the reaction time and percent correct methods allow measurement of MWL only when the primary task shows explicit performance (in this case, at stop lights). Since the future goal of this dissertation is to estimate a driver's MWL continuously, even without explicit performance, these measures are not entirely appropriate.

Secondary-task performance measures estimate an operator's MWL using a performance score on a secondary task. For example, Tokuda, Morris and Chaparro (2007) asked participants to react to roadside alphabet signs that appeared every 3 to 10 seconds, so that the researchers could examine working memory usage in their study. This sort of MWL measure can be useful especially in a laboratory condition where a secondary task can be added and where the interference of the secondary task with the primary task does not lead to an accident. However, outside of laboratory, this measure is not realistically implementable to estimate a driver's MWL.

The third type of empirical measure is subjective difficulty ratings, which is a common way to reveal the person's subjective feeling of a task's difficulty. In these methods, participants themselves rate their own feelings of task difficulty, based on the assumption that people can accurately report their workload (Johnson & Proctor, 2004). This approach is advantageous because the observers themselves may be the best people to relay their cognitive state. This method of asking the operators has great potential to discover something the other methods cannot measure. Also, this is one of the most inexpensive ways to measure MWL, though its major disadvantage is that it cannot be implemented in a real-time quantification since the performer needs to stop the primary task when rating the difficulty (Stanton, Salmon, Walker, Baber & Jenkins, 2005). This is a violation of the implementation requirement criterion.

The two major types of subjective rating systems are uni-dimensional methods and multi-dimensional scales, such as the National Aeronautics and Space Administration Task Load Index (the NASA -TLX) (Johnson & Proctor, 2004, Stanton et al., 2005). The uni-dimensional method asks participants to rate how difficult they felt a task was immediately after the task. This implementation quickly estimates MWL, and the accuracy in predicting cognitive performance is almost as good as the other, more complicated method, the NASA-TLX (Hendy, Hamilton & Landry, 1993). The NASA-TLX, takes longer to estimate MWL. The NASA-TLX is supposed to be able to diagnose each of six dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration level. This can be done by asking participants to rate each of these six scales and judge the dominant influential effects by comparing pairs of the six scales (Hart & Staveland, 1988). With the diagnostic ability to extract only mental demand, the NASA-TLX has potential to estimate MWL more precisely than the uni-dimensional subjective rating method (Hart & Staveland, 1988).

Subjective rating techniques are not good at estimating a driver's MWL, because they cannot be used in real-time. However, they are still very useful where MWL estimation does not require real-time estimation. For example, both a uni-dimensional method and the NASA-TLX are used for the experiments in this dissertation in order to confirm the manipulation of MWL by the N-back task.

The fourth empirical measure is the physiological approach. Physiological measures use human behavioral responses as MWL measures. If a certain physiological response occurs consistently and uniquely under certain mental states, then the physiological response has the potential to represent the cognitive state. Physiological responses include electroencephalogram (EEG) recordings, eye activities, heart rate, or any other kind of physiological reaction. Physiological responses have the best potential to accurately estimate MWL because they give continuous signals to indicate MWL status and may potentially be gathered without interfering with the operator's primary task.

In addition to the physiological measure mentioned above, this dissertation offers a new MWL measure: saccadic intrusions. Heart rate is known to increase or decrease with mental activities. Since the directions are not consistent, heart rate variability is a better MWL measure than heart rate itself. Although heart rate variability seems to be accurate in reflecting MWL levels in laboratory conditions, heart rate variability is affected by many other factors as well, such as respiratory rate and other mental activities. This decreases the sensitivity of heart rate variability as a MWL measure, especially outside the laboratory.

Another type of physiological measure is brain waves as measured by EEG. Brain waves, especially P300 waves, seem to be the most accurate in estimating MWL in laboratories (Wickens, 2007). However, it is not practical to conduct EEG analyses in driving environments, because devices to measure these physiological responses are large and intrusive to the driver.

Other types of physiological measures are eye activities. In the last decade, cognitive research has highlighted eye tracking, because eye tracking devices such as the Tobii eye tracker are getting small enough to mount in a vehicle, and eye tracking data seems to accurately reflect MWL. One of the most often measured eye activities, pupil diameter, is known to increase with

MWL. This dissertation will use pupil diameter while proposing a new MWL measure based on saccadic intrusions. Both of these eye measures show promise in estimating a driver's MWL and thus are used in the dissertation experiments. In the next section, these eye measures are described in more detail than the other MWL measures described above.

2.3 Eye Movements

2.3.1 Introduction

One of the goals of this dissertation was to demonstrate an increase in a specific type of eye movement (saccadic intrusions) with increased mental workload (MWL). This section describes the characteristics of saccadic intrusions (SIs) and other types of eye movements. The eye movement types are essential to this dissertation, specifically in the Algorithm section, because the algorithm detected only particular types of eye movements, specifically SIs, and quantified them in a dependent variable (DV).

2.3.2 Two Systems of Eye Movements

Eye movements can be categorized into two distinct types: saccades and pursuit (Carpenter, 1988; Stone, Beutter & Lorenceau, 2000; Krauzlis, 2005). These two types are distinct not only on the observable level, but also on the physiological mechanism level. Fixations are generated by the pursuit system, whereas saccades are generated by the saccadic system (Robinson, Gordon & Gordon, 1986; Stone, Beutter & Lorenceau, 2000; Krauzlis, 2005).

Saccades are rapid eye movements that change gaze positions from one location to another, whereas pursuit eye movements follow a visual stimulus of interest when the visual stimulus is either stationary or smoothly moving. The two eye movements systems work in a compensatory fashion. Usually, either one of the two types is observed in a single instant and they tend to alternate.

When reading, for example, the eyes' gaze does not smoothly move from left to right at a constant pace, but moves with cycles of jumps and stops. From the beginning of a sentence, the eye gaze jumps to a few words ahead, and stays at the location for 200 to 300ms, then jumps a few words ahead (Rayner, 1984). The stationary period is called a fixation. The jump movement is called a saccade. Each of the fixations and saccades can be subdivided into smaller categories, which are explained later in this chapter.

Figure 2.2 is an example of a 30 s horizontal eye movement obtained while the observer was engaged in a random dot fixation task. Every two to six seconds, the fixation target randomly changed its location, and the observer was instructed to follow the fixation target dot with their gaze. The vertical lines are saccades, and relatively continuous horizontal lines are fixations. Each of the fixations is not a straight horizontal line; fixations entail constant, tiny eye deviations. Since one of the dissertation goals was to quantify eye deviations during each fixation, this section explains the types of these deviating eye movements during fixations.

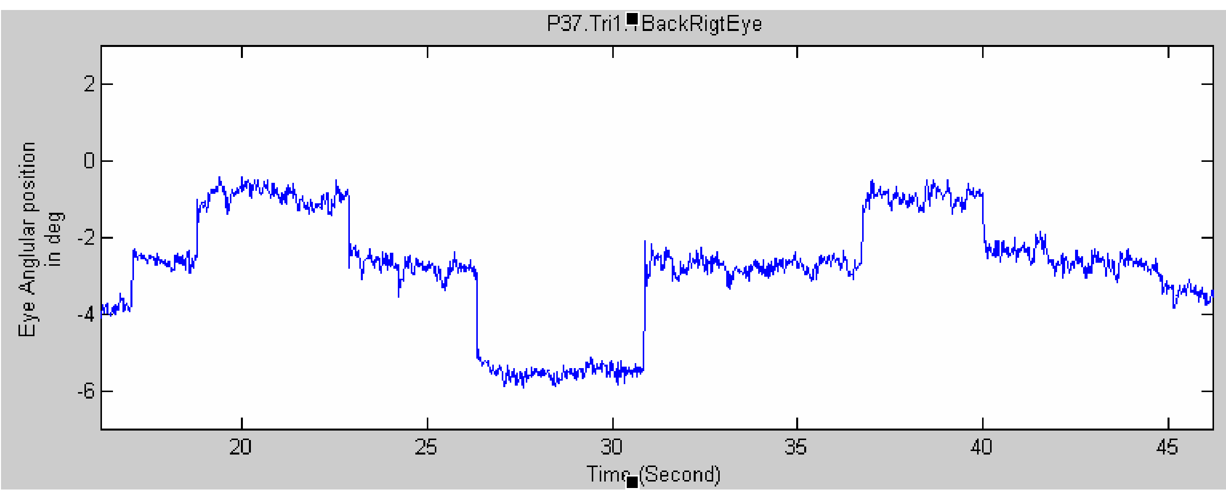

Figure 2.2. Example of eye movements including fixations and saccades, as well as eye deviations during fixations. The research interest was in the deviation during fixations.

2.3.3 Saccadic System

The saccadic system produces several types of eye movements including regular saccades, microsaccades, and saccadic intrusions (SIs). All saccadic eye movements share several common characteristics. They are jerky, ballistic, rapid eye movements that last for a short time. In 1879, Javaland Landolt introduced the term saccade (Leigh & Zee, 2006), referring to eye movements that were “jerky” or a “violent pull”. Thus, all saccadic eye movements share characteristics of jerky movements, in which a saccadic eye movement abruptly starts and abruptly stops. Saccadic eye movements are considered ballistic because the destinations are pre-determined and cannot be modified during the eye movement. Saccadic eye movements are rapid and can have velocities that range from 20 degrees/sec to 500 degrees/sec (Leigh & Zee, 2006). The duration of saccades vary, one episode of a saccadic eye movement lasts can be as short as 20 to 120 ms (Robinson, 1964) while one episode of a pursuit or a fixation (more precisely, a series of fixational eye movements) can last up to a few minutes. All three types of saccadic eye movements – regular saccades, microsaccades, and saccadic intrusions (SIs) – share the characteristics described above.

2.3.4 Three Types of Saccadic Eye Movements

Although there are at least three types of “saccadic” eye movements, the term “saccade” usually refers to regular saccades because the other two types of saccadic eye movements are usually small enough or rare enough to ignore. However, they are all important to this dissertation. To reduce confusion, the normal “saccades” always are referred to as “regular saccades” in this dissertation. Regular saccades are the ones that connect two fixations in different locations. Amplitudes of a regular saccade can be up to 90 degrees (Carpenter, 1988).

The function of a regular saccade is to move the eyes to a new fixation point so that the observer can look at a new visual stimulus.

The other two types of saccadic eye movements – microsaccades and SIs – are also generated by the saccadic system and have characteristics of that system, such as movements that are jerky, rapid, ballistic, and short in duration. However, they differ from regular saccades; their function is not to shift fixation to a new visual stimulus. Their function is not entirely clear, but they apparently serve to destabilize the eye gaze by very small amounts (usually much less than 0.1 degree), so that the eye can stop desensitizing from the fixating visual stimulus, thus, increase visual perception of the fixated stimulus (Martinez-Conde, 2006).

2.3.5 Pursuit System

The pursuit system, on the other hand, seems to have the primary goal of sending strong, stable visual images to the visual perception areas. This seems to be achieved by two seemingly contradicting movements: by stabilizing visual images on the retina and by destabilizing them to a small degree.

The stabilizing part is easy to understand (although it is still amazing how the eye can counteract any motions every few milliseconds). The eye tries to cancel out all kinds of shaking motions by counteracting the eye’s movement to the shaking motions, taking the head movements and the visual target’s movements into account. When head movements are restricted as in the experiments in this dissertation, the pursuit system has two types of eye movement: fixation and smooth pursuit. These types of eye movement can function together to keep the retina image at the same position, so the visual perception system can hold the image longer. Thus, the image can be perceived well. Primate visual systems, including humans, are adapted to coordinate both eye movements and head movements simultaneously. The eye movements automatically cancel out the head movements and sometimes target movements, so that the visual image can be maintained in the same retinal position.

The first type, fixation or visual fixation, is the only one that does not involve either head movement or visual target movement. During a fixation, the eye gaze is not perfectly stable, but is continuously making “ocular drifts” or “fixational eye movements”, seen in Figure 2.2. When an operator stares at a stationary visual target, the eye locks on the stationary target and automatically corrects the small deviation between the seen object and the eye gaze (Leigh & Zee, 2006).

The second type, smooth pursuit, involves a target movement. When a target smoothly moves, the eye can smoothly follow the object, so that the movement of the target falling on the retina is cancelled out by the eye movement. At the same time, the eye automatically corrects the small deviation in the course of smoothly following the moving target. Both of these types of eye movements above are mainly produced by the pursuit system.

2.3.6 Fixational Eye Movements

The second function of the pursuit system is more difficult to understand; the eye purposefully adds destabilizing motions to eye movements. A perfectly stabilized image on the retina will fade due to adaptation (Engbert, 2006; Martinez-Conde, 2006). To prevent this perceptual desensitization, and in order for humans to examine stationary visual stimuli, such as reading materials and emotions on other people’s faces, the eye makes small, seemingly unnecessary movements that prevent desensitization to stationary visual stimuli (Martinez-Conde, 2006). These miniature movements that help visual perception during fixations are called fixational eye movements or miniature eye movements (Carpenter, 1988).

Engbert (2006) states, “The term visual fixation is a misnomer, since there is rich dynamical behavior during each fixation. To capture this built-in paradox, the term fixational eye movements is used most often.” This dissertation follows his view and uses the term fixational eye movements (FEMs) rather than fixation, when movements during each fixation are important.

There are three commonly identified types of FEMs: tremors, slow drifts, and microsaccades (Carpenter, 1988). There is also a fourth type of eye movement that can be observed within FEMs, saccadic intrusions (SIs). All FEMs are involuntary (Carpenter, 1988), despite the intention to fixate the gaze on a stationary target. At least the first three types of FEMs (tremor, slow drift, and microsaccade) serve to help desensitize people to visual stimuli since perfectly stabilized images on the retina will be habituated and not visible after several seconds (Martinez-Conde, 2006; Engbert, 2006).

Tremors or ocular tremors are prevalent and have the highest frequency (up to 150 Hz) among the three types of fixational eye movements (Findlay, 1971). Tremors have very low amplitudes of around 0.25 arcminutes (i.e., 0.004 deg) and have a duration of only a few milliseconds (Yarbus, 1967). Figure 2.3 (left) shows a set of horizontal eye movements during a period of 15 seconds with the horizontal angular rotation ranging approximately from -0.8 to +6.5 deg, recorded by the Tobii eye tracker. The y-axis in Figure 2.3 represents the angular rotation of the eye on the horizontal plane (Figure 2.3 right). The high frequency noise prevalent during the whole period are considered tremors, although the Tobii 1750 eye tracker used to record this data is not accurate enough to measure super-fine angular rotations of 0.004 deg. Tremors are disconjugate (Ratliff & Riggs, 1950), which means that the movements of the two eyes are not synchronized.

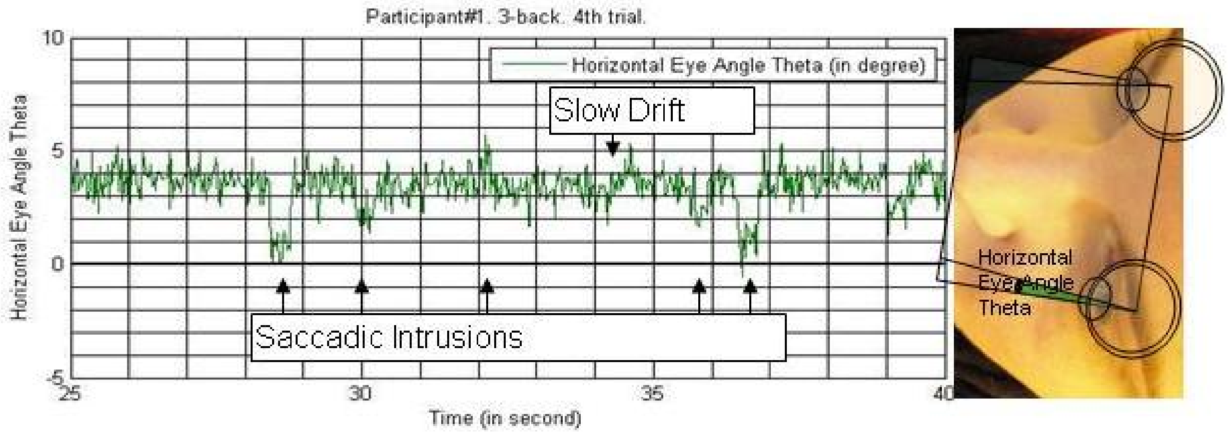

Figure 2.3 (left). Examples of fixational eye movements, including tremors, slow drifts, and saccadic intrusions.
Figure 2.3 (right). Horizontal eye angle theta.

Drifts or slow drifts, which are small angular rotation changes, occur simultaneously with tremors. Each slow drift has a duration of up to a few seconds. In Figure 2.3, among many drifts, a large slow-drift is easily noticed between the 34 and 35 second marks (labeled "slow drift"). Drifts are disconjugate eye movements with a frequency ranging from 0.1 to 3Hz (Ratliff & Riggs, 1950). Amplitudes of drifts are around 0.03 to 0.08 deg (Ratliff & Riggs, 1950).

The third and fourth types of fixational eye movements are microsaccades and saccadic intrusions (SIs). Both are spike-like and have the largest amplitudes among fixational eye movements, seen in Figure 2.3. Since these are the focus of this dissertation, the next several paragraphs review the characteristics of these types of eye movements in more detail.

2.3.7 Saccadic Eye Deviations (SEDs)

Microsaccades and SIs jointly make up saccadic eye deviations (SEDs) because they are saccadic eye movements, and deviations that occur during FEMs. Some characteristics are shared within these two types of eye movements while others are not.

First, there are many common characteristics shared between microsaccades and SIs. Both microsaccades and SIs is eye movements occur during FEMs. Both are types of saccadic eye movements, being fast, jerky, ballistic, and conjugate. For eye velocities of microsaccades and SIs, refer to the velocities of the saccadic system (Section 3.3, The Saccadic System, or Abadi & Gowen, 2004; Engbert, 2006). Both are conjugate eye movements (Abadi & Gowen, 2004; Engbert, 2006; Leigh & Zee, 2006). Therefore, an analysis of one eye is sufficient to account for the behavior of both eyes. Both microsaccades and SIs mainly occur on the horizontal plane (Abadi & Gowen, 2004; Engbert, 2006). This indicates that horizontal eye movement analysis can be sufficient to detect these saccadic eye deviations.

On the other hand, there are some characteristics that are not shared between microsaccades and SIs. While most microsaccades have amplitudes of up to 0.2deg (Engbert, 2006), SIs have larger amplitudes, mainly ranging from 0.1 to 4.1deg or larger (Abadi & Gowen, 2004), though the lower limit of the amplitude varied depending on the types of categorizations and analyses, ranging from 0.1 to 0.5 deg. Since this dissertation focuses on SIs and excludes microsaccades, this dissertation sets the lower amplitude limit of SIs at 0.4 deg. Saccadic eye deviations with an amplitude of 0.4 deg or smaller are considered to be microsaccades. This difference was used in the algorithms to differentiate these two types of saccadic eye deviation. Also, unlike microsaccades, SIs usually have a round-trip eye deviation; the eye gaze deviates from the previous position, dwells at the deviated location for a period of time, and then returns to the original gaze position. The dwelling duration ranges from 10 ms to 880 ms (Abadi & Gowen, 2004). Microsaccades and SIs might be qualitatively the same eye behavior with quantitative differences in the amplitudes (Gowen, Abadi, Poliakoff, Hansen, & Miall, 2007).

2.3.8 The Term “Saccadic intrusions”

The focus of this dissertation is in the type of eye movement that Abadi and Gowen (2004) identified as “saccadic intrusions”. They defined saccadic intrusions as “conjugate, horizontal saccadic movements which tend to be 3-4 times larger than the physiological microsaccades and take the form of an initial fast eye movement away from the desired eye position, followed, after a variable duration, by either a return saccade or a drift” (Abadi & Gowen, 2004). This dissertation adopts the saccadic intrusion definition of Abadi & Gowen (2004) because just two words (saccadic intrusion) or two letters (SI) can represent three characteristics. First, the eye movement is saccadic (or rapid and jerky). Second, it is a deviating eye movement from a fixation position, and the deviation returns to the previous fixation position. Third, the saccadic eye deviation occurs within or intrudes upon a fixational eye movement. These are the exact characteristics of eye movements that this dissertation addresses.

There is concern over the usage of the term saccadic intrusion. In many medical pathology articles, saccadic intrusions are considered as a symptom of a disease (i.e., Clementz, Sweeney, Hirt & Hass, 1990; Zee, Yee, Cogan, Robinson & Engel, 1976). However, this dissertation assumes that SIs would occur in all healthy people, following two studies; Abadi and Gowen (2004) showed the presence of what they called “saccadic intrusions” in healthy adult’s eye movements in 50 out of 50 participants. Similarly, McGivern and Gibson (2006) observed saccadic intrusions in all the 55 “normal subjects” they tested.

Taking into account all the views above except the pathology views, this dissertation defines saccadic intrusions as a type of eye movement with all the following characteristics ; (1) a SI is a saccadic type of eye movement, (2) a SI occurs during a fixation, (3) a SI deviates from the fixation baseline by an amplitude of 0.4 deg or larger, (4) however, a deviation amplitude is not larger than 4.1 deg (this is explained in a later section "*3.12: Gaze aversion*" in this chapter), (5) a SI returns to the fixation baseline, (6) a SI has a deviation dwelling period ranging from 10 to 880 ms, (7) a SI is conjugate, (8) a SI occurs on the horizontal plane, and (9) normal people have SIs. All of these characteristics are used in the algorithm to quantify saccadic intrusions. Also microsaccades can be detected using almost the same methods since microsaccades share many of the characteristics such as (1), (2), (4), (7), (8), and (9). Because of the overlap between microsaccades and SIs, research findings on microsaccades were applied to this dissertation on SIs.

2.3.9 Detection of Saccadic Eye Movements

The characteristics of microsaccades, SIs, and SEDs have been examined by three research groups, Abadi and Gowen (2004), McGivern and Gibson (2006), and Engbert (2006). This section explains how they studied and identified these types of eye movements in their studies.

SIs were extensively examined by Abadi and Gowen (2004). They recorded eye movement data for 50 healthy participants during a fixation task. Abadi and Gowen took the eye movement data into data graphic software, such as Microsoft Visual Basic and Microsoft Excel. They calculated the eye movement's differentiation (such as velocity), and plotted it in a graphic figure, then identified the starting and ending positions of saccades by visual observation. They used the velocity criteria (10 deg/s or above) to identify saccades. They did not use an algorithm to automatically identify saccades. Their eye tracker had a higher resolution than the Tobii eye tracker which was used in this dissertation. The resolution of the eye tracker used in Abadi and Gowen (2004) was 200 Hz (polling every 5 ms) of the temporal resolution and five arcminutes (0.08 deg) of spatial resolution.

McGivern and Gibson (2006) also examined SIs. They recorded eye movements of 55 healthy participants during a fixation task. Their eye tracker's temporal resolution was 1000 Hz (every 1 ms). They calculated velocity using the best fit linear slope of five consecutive eye data (for 5 ms). Following Abadi and Gowen (2004), they used the threshold velocity of 10 deg/s to identify saccades. McGivern and Gibson used average slopes of five consecutive eye data because eyes' spatial data contain noise, and average slopes could minimize spatial noise. They could use this strategy because a duration of 5 samples was as short as 5 ms. Since the 5 ms window was much smaller than a duration of a saccade (around 20 ms or larger), the 5 ms window could detect a saccade.

Engbert (2006) used an algorithm to detect saccadic eye deviations. He used an eye tracker with a temporal resolution of 500 Hz (every 2 ms). Like McGivern and Gibson (2006), Engbert also used 5 samples (10 ms) to reduce spatial noise of eye data. Again, even after using 5 samples, the eye data duration was only 10 ms. If there was a saccade of 20 ms, the 10 ms window could detect the 20 ms saccade.

Unfortunately, this dissertation could not follow these three groups' methods to identify saccade or saccadic intrusions because of the low temporal resolution and low spatial resolution in the Tobii eye tracker. The Tobii eye tracker had a temporal resolution of 50 Hz (every 20 ms) and spatial resolution of 0.25 deg. If the previous method was followed and five samples were used, the duration of 5 samples in Tobii would take 100 ms. If there is a saccade during the 100 ms inspection window, the algorithm would not be able to detect the exact timing of the saccade and the existence of saccade. Also, since the spatial resolution of the Tobii eye tracker was lower than the previous studies by these three groups, the same method would not work in this dissertation. Because of these two reasons, a new algorithm was developed from scratch by the author of this dissertation to detect saccades. The new algorithm to identify SIs is described in Chapter 3: Algorithm.

2.3.10 Ocular Dominance

As Engbert (2006) suggests, detection of microsaccades would be more thorough with an analysis of both eyes. However, this dissertation used only one eye for the analyses for three reasons. First, saccadic eye movements are conjugate, which means every time one eye has a saccadic eye movement, the other eye has an identical one too. Second, this dissertation's goal is to show a general relationship between SIs and MWL. The analysis one eye is sufficient to show the general relationship. However, if more accurate analyses are needed in the future, Engbert (2006) suggests using two-eye interaction analyses. Third, if one eye's movements can predict MWL accurately enough, it may be easier to implement. For these three reasons, the analyses were run using one eye's data.

It does not seem to matter much which eye is analyzed since saccadic eye movements are conjugate. However, the right eye was used for each participant's analyses. The right eye, rather than the left eye, was chosen because of the data on ocular dominance research. Ocular dominance is the degree of the person's dependence on one eye (Papousek & Schulter, 1999). They showed that there were more right eye dominant people (63%) than left eye dominant (32%) after removing the people who were equal ocular dominant. However, upon closer examination of the data, the distribution of ocular dominance seemed to have three modals, left, center, and right ocular dominance. Only around 20 % of sample were left eye dominant. Thus the remaining 80 % use the right eye actively. This shows the right eye may reflect more precisely how the ocular motor system controls the eye.

2.3.11 MWL and Saccadic Eye Movements at the Behavioral Level

It is a reasonable induction that SIs may reflect MWL. Past studies had shown a similar effect at the behavioral level. This section reviews seven studies that show relationships between MWL and eye movements. All show that eye movement patterns change when there was increase in MWL. Some studies measured microsaccades, but due to the strong similarities between microsaccades and SIs as explained before, results on microsaccades are considered applicable to SIs.

Previous studies have shown that microsaccades reliably increased in frequency and amplitude approximately 350ms after visual presentation of an attentional cue (Engbert, 2003; Engbert & Kliegl, 2003; Laubrock, Engbert, Rolfs & Kliegl, 2007), or 300 ms after auditory attention cue presentation (Rolf, Engbert & Kliegl, 2005), or 500 ms after endogenous attention (Laubrock, Engbert & Kliegl, 2005). A series of studies by Dr. Engbert's lab shows that cognitive demand consistently induced microsaccades (Engbert, 2006).

It is unclear if these physiological changes (increased microsaccades with increased MWL) are due to looser physiological control due to cognitive overload of the movement control center. However, the mechanism does not have to be revealed. As long as there is a consistent physiological change with varied MWL levels, the physiological change may be used to reflect MWL.

Similarly, visual fixations were less stable when the participants were engaged in a higher MWL task (Camilli, Terenzi & Nocera, 2007). The behavioral response is consistent with the hypothesis in this dissertation that when cognitive demand is high, fixation is deviated more by having more SEDs such as microsaccades and SIs. All of these studies are consistent with the main hypothesis of this dissertation; SIs would increase as MWL were manipulated to a high level.

2.3.12 Criteria of Eye Movements as MWL Measures

There are several criteria that help to evaluate any measure to estimate MWL. These criteria were proposed by Gopher and Donchin (1986) and Boff and Lincoln (1988) (see the section 2.9, “Criteria evaluate MWL measures”), including sensitivity, reliability, unintrusiveness, and implementation requirements, and operator acceptance. The author of this dissertation argues that it is important to consider three of the criteria outlined by Gopher.

First, a MWL measure should be available almost all of the time. This is part of Gopher’s criterion, “sensitivity” to accurately predict MWL all the time. For example, pupil diameter is usually observable, which is a good characteristic of a MWL measure. This dissertation focuses on SIs and microsaccades, which occur every few hundred milliseconds during a fixation. Given the fixations occur quite frequently, the SI measure seems to meet the first criterion.

Second, a MWL measure should have a predictable behavior under normal circumstances. Without knowing the typical behavior, it is difficult to evaluate if a behavioral response is normal or abnormal. For example, the VOR is a good candidate for showing behavior that is either normal or abnormal. When the VOR works perfectly, eye movements cancel out head movements. In other words, the eye movements would be a perfect mirror image of the head movements. If not, the VOR is not perfectly working, and there must be a reason for its malfunction, including cognitive resource scarcity. The proposed measure using SIs is good because it is known that SIs rarely occur in a normal condition. If SIs are observed more often with larger amplitudes, that is a sign of abnormal behavior.

Third, a MWL measure should be easy to observe. For example, the VOR is very difficult to observe because the head does not shake more than one angular degree per second usually. Consequently, the VOR is usually less than one degree per second. This kind of small eye movement is difficult to extract among the all the other types of eye movements. On the other hand, this dissertation's focus, SIs, is relatively easy to extract. See the Algorithm chapter for how SIs were detected and quantified.

This third criterion for a type of eye movement to be a measure of cognitive resource usage has another aspect. The eye movement type has to be not only big enough to observe, but also distinct enough from other types of eye movements. The following sections describe two types of eye movements that are similar to SIs.

2.3.13 Gaze Aversion

When SIs are detected during FEMs, there are two types of eye movements that can hinder the detection; microsaccades and gaze aversion. For details on microsaccades, please refer the section 3.7: Saccadic Eye Deviation.

Gaze aversion can be defined as an eye movement away from a visual target, so that cognitive resources can be allocated to a non-visual task, rather than to the visual target (Doherty-Sneddon, Bruce, Bonner, Longbotham & Doyle, 2003). Gaze aversions occur even if they are irrelevant to the experimental task. The algorithm tries to distinguish gaze aversions from SIs by using the differences in amplitude, the return window, and orientation of the eye movement (See the chapter 3, Algorithm chapter).

Gaze aversion is typically associated with gaze direction studies such as lateral eye movement, conjugate lateral eye movement, Neuro-linguistic programming, and eye movement rate (Baker, Goldstein, and Stern, 1992). These studies try to reveal "laterality" (i.e., differences between the two hemispheres of the brain) by observing eye movement directions. It is popular enough to show up in movies. One of them is The Negotiator, in which a character played by Samuel Jackson believes a man is lying, saying “If your eyes go up and right, you're accessing the brain's creative centers and we know you're full of shit.” These eye gaze direction studies lost popularity finding a series of inconsistent results. Interested readers can see more in Siegle, Ichikawa and Steinhauer (2008), Baker, Goldstein, and Stern (1992).

2.3.14 Eye Movements as a MWL Measure

So far, we have reviewed the physiological characteristics of eye movements and their links to MWL. The advantage of using eye movements as a MWL measure is the eye’s physiological link to the brain. Since the majority of the brain is used by the visual system, there is a good chance that any cognitive state changes, including MWL changes, can influence eye movements. It is also advantageous that the eye is observable from the outside, unlike brain waves. On the other hand, one of the downsides is that there are many brain regions (more than 30 regions) affecting eye movements (Carpenter, 1988); it is difficult to extract only specific types of eye movements or noises out of many kinds of eye movements. That means that the quality of algorithms to extract specific types of eye movements is the key to the quality of the MWL measure. The algorithm that the author developed is described in detail in , chapter 3.

2.4 Pupil Diameter

Pupil diameter is popularly used as a measurement of a MWL in laboratory conditions (Recarte & Nunes, 2000, 2003; Tsai, et al., 2007). Pupil diameter can reliably and accurately measure MWL as long as the experimental conditions are well-controlled, eliminating other factors that can influence the pupil diameter.

2.4.1 Physiology of the Pupil

The pupil is an aperture in the center of the eyeball. The iris contracts or expands to change pupil size, so that the amount of light coming into the eye is adjusted. Therefore, the pupil controls brightness on the retina, and also at the same time, it changes the number of aberrations that are caused by diffraction of light. The human pupil can change its size from 2 to 9 mm (Guyton, 1977). Both pupil diameter and eye movements are physiological activities that are observed in the same body part. However, the control of these two systems is entirely different. Eye movements are controlled by the more complicated nervous system, taking in account covert attention, many aspects of visual stimuli, and experience. Moreover, the nervous systems responsible for eye movements are the central nervous system and the somatic nervous system, which includes sensory nerves and motor nerves. On the other hand, pupil diameter is a simple reflex controlled mainly by the autonomic nervous system (Andreassi, 2000). Because of the strong link between pupil control and autonomic nervous system, activities such as illumination change reactions and startling can affect pupil diameter (Andreassi, 2000). These activities and cognitive states that affect pupil diameter include fatigue (Kahneman & Peavler, 1969), time of the day (Geacintov & Peavler, 1974), pain (Chapman, Oka, Bradshaw, Jacobson & Donaldson, 1999), and sexual arousal (Chapman, Chapman & Brelje, 1969). All of these conditions are strongly related to the autonomic nervous system.

2.4.2 Workload Measure

In addition to the autonomic nervous system, higher cognitive activities influence pupil diameter as well. Just and Carpenter (1993) found that pupil diameter increased about one second after a sentence was visually presented. This was not because of a stimulus startle effect but because the complexity of the sentence. Pupil diameter is used in a wide variety of laboratory settings to measure workload. For example, Ahlstrom and Friedman-Berg (2006) examined a correlation between pupil size and cognitive workload. They examined six air-traffic control operators and manipulated their cognitive workload by varying the number of aircraft on their radar screens from two to nine by adding one aircraft at a time. The collapsed average pupil diameter varied between 3.5 and 3.6 mm in a positive linear function, reflecting the number of aircraft. The linear regression R^2 value was as high as 0.70. This indicates that MWL can be estimated well by observing pupil diameter.

2.4.3 Pros and Cons of Measuring Pupil Diameter

There are some advantages in using pupil diameter as a measure of MWL. First, as a physiological measure, it can be implemented in objective, real-time, and automatic manner. In terms of real-time, the pupil diameter starts increasing in size about one second after workload is imposed (Just and Carpenter, 1993; Siegle, Ichikawa, & Steinhauer, 2008). This real-time characteristic may be fast enough to estimate driver MWL. Additionally, pupil diameter can be observed from outside the human body without having a physically intrusive device attached to the operator. Considering these advantages, pupil diameter is a viable MWL measure that is reliable, accurate, real-time and unobtrusive.

However, despite these advantages, pupil diameter is not used to measure driver MWL outside laboratory or experimental conditions. There are several disadvantages in using pupil diameter in the real world. Primarily, pupil diameter is overly sensitive, reflecting many aspects of the mind (Wickens, 2007). As the section above shows the factors that change pupil diameter range from light, startle, pain, and sexual arousal to MWL. It is difficult to isolate the influence of MWL alone among these other factors.

Second, pupil diameter is less informative, in contrast to eye movements. While eye movements can be observed on three axes (horizontal, vertical, and torsial movements), pupil diameter can physiologically show its change only on one axis (dilation or constriction). This means eye movements contain more information (such as reflection of environment and reflection of mental condition) than the pupil diameter when being analyzed. In addition, temporal changes in physiological responses seem to be more informative in eye movements than in pupil diameter. The shortest latency for eye movements is as quick as a few milliseconds for VOR (Robinson, 1982). On the contrary, the pupil diameter reflex latency to an illumination change is as slow as 200 ms (Guyton, 1977; Siegle et al., 2008), which is 40 times slower than eye movement latency. Pupil diameter would have been more informative if the latency was a perfect phase shift of 200 ms delay, reflecting illumination changes exact 200 ms ago. Rather, pupil diameter is more likely to have a sluggish delay, gradually losing reflective information during the 200 ms latency. Pupil diameter changes in a few milliseconds are not as functionally important as eye angle changes (i.e. eye movements) in the same period. Therefore, eye movements may have more information in a few milliseconds than does pupil diameter.

There also seems to be more information in the spatial resolution of eye movements than in pupil diameter. As depicted in the eye movement review, the eye integrates many types of eye movements such as fixations, smooth pursuits, saccades, tremor, slow drifts, microsaccades, VOR, OKR, and vergence. These various types of eye movements have a wide variety of amplitudes (from 0.001 deg to 90 deg). Conversely, pupil diameter has not been shown to have multiple levels, like eye movements. It is still possible that pupil diameter contains as much information as eye movements. However, it is not observable, possibly due to the limited motor actions (having only one axis) or due to the slow latency of the pupil diameter.

These characteristics show that pupil diameter does not reveal as much information on different levels in time and space than eye movements. Because of these constraints, it seems the use of pupil diameter as a MWL measure is limited to experiments in controlled conditions, and it is difficult to measure changes of pupil diameter related to MWL in a complex environment.

2.4.4 Pupil Diameter in This Dissertation

This dissertation focuses more on eye movements as a MWL measurement, rather than pupil diameter. Since a pupil measure is oversensitive to illumination change in a driving environment, a pupil measure is not practical in estimating a driver's MWL in real-time in a real world driving environment. To make it worse, a pupil measure is not informative enough to let researchers extract only MWL related pupil change out of many aspects that the pupil would reflect. On the other hand, eye movements are getting easier to measure with much more precise accuracy. Since eye movements have more temporal and spatial information than pupil diameter, there is more potential for further research in eye movement. This dissertation examines eye movements as the main focus, and limits a pupil measure to complementary use to the eye movement measure.

2.5 Conclusion of Literature Review

This chapter reviewed literature on MWL and eye activities such as eye movements and pupil diameter. All the experiments and hypotheses in this dissertation are derived from these past studies on MWL and the expected physiological responses. For example, the core hypothesis in this dissertation predicted that increased MWL would induce more saccadic eye deviations and a larger pupil diameter. The pupil part of the hypothesis has already been established by previous researchers, and this dissertation would confirm those phenomena. The saccadic eye deviation part of the hypothesis was induced from the behavioral level studies on the relationship between microsaccades and attention. Also on the mechanical level, previous research has shown that saccadic eye deviation within fixational eye movements seem to be easy to observe, and therefore easy to use as a MWL measure. Methods for detecting and quantifying eye movements are described more detail in Chapter 3, Eye Movement Analysis Algorithms.

CHAPTER 3

EYE MOVEMENT ANALYSIS ALGORITHMS

3.1 Significance of Algorithms

The majority of psychology studies use experimental designs with independent variables (IVs) and dependent variables (DVs). An experimenter controls the IVs and sees if there is a difference in the DVs. This dissertation did the same, as reported in the following chapter (Chapter 4: Methods). However, in this dissertation research, the DVs were calculated using original algorithms developed by the author. The purpose of this chapter is to describe the eye movement algorithms and their theoretical justification.

The Tobii eye tracker samples eye position every 20 ms (50Hz). The raw data files are large and difficult to interpret without additional post processing. For example, in a 20 seconds period, Tobii produces a data files consisting of 16,000 data points (50Hz x 20 seconds x 16 pieces of eye activities that are being recorded by Tobii). Therefore, the raw data needed to be quantified into a summary format that captures the specific variable of interest. This chapter explains how the algorithms were developed and how the algorithms are theoretically meaningful.

All the algorithms in this dissertation used the same general procedures: the algorithm took a set of eye data from the Tobii eye tracker and summarized it along specific dimensions of interest into a single value for a trial period (usually around 30 seconds). These specific variables of interests included eye gaze deviation, Fourier transformation (FT) power densities, and pupil diameters. The DVs were used in subsequent correlation and linear regression analyses, among others. Changes in these DVs were examined as a function of changes in the IV, mental workload (MWL).

3.2 Saccadic Eye Deviation Algorithm

3.2.1 Algorithm Goals and Constraints

The ultimate goal in this research is to develop an algorithm set that automatically quantifies eye activity (such as eye deviation and pupil diameter) and integrates it to accurately estimate MWL in real-time. Algorithm 1 quantifies saccadic eye deviations, which were hypothesized to correlate with MWL. The development of Algorithm 1 was guided by eight principles that identified conditions that Algorithm 1 needed to satisfy and some caveats that it did not need to achieve.

Principle 1: Algorithm 1 focused on quantifying saccadic eye deviation during fixational eye movement (FEM). Some examples of saccadic eye deviation are shown in Figure 3.1 below. The hypothesis that saccadic eye deviation would predict MWL is reasonable because past studies show that increased attention led to increased saccadic eye movements (Engbert & Kliegl, 2003; Laubrock et al., 2007; see more in the literature review section). It is possible that there are also other kinds of eye deviation that are related to MWL, but Algorithm 1 was designed to detect and quantify saccadic eye deviation.

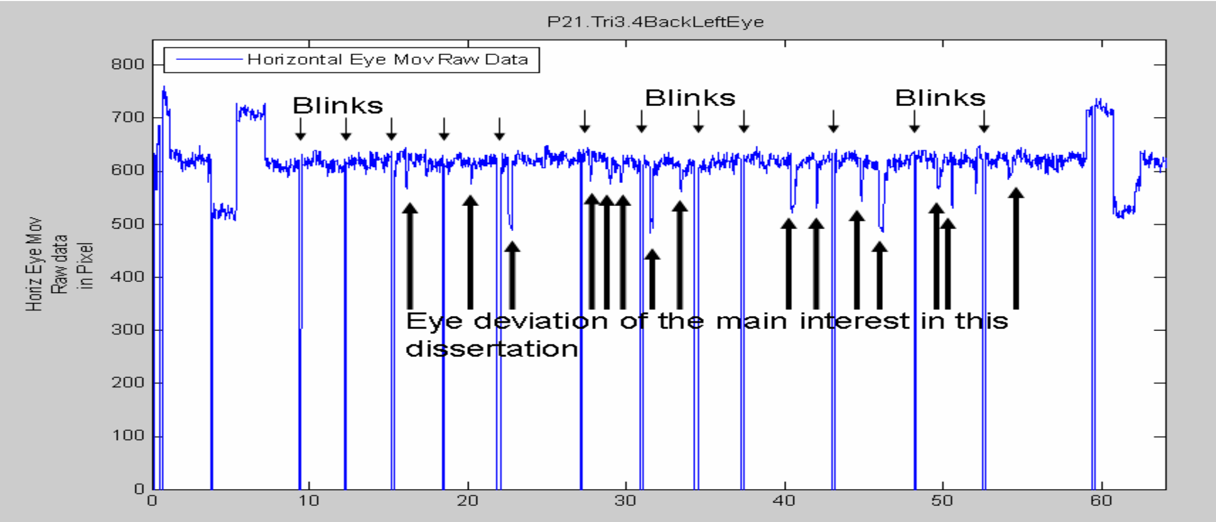

Figure 3.1. Examples of saccadic eye deviation in one experimental trial for 64 seconds. The y-axis of the graph represents the horizontal eye gaze on the Tobii eye tracker screen. Approximately 38 pixels are equivalent to 1 degree of visual angle. The purpose of Algorithm 1 was to quantify only saccadic eye deviation after removing all other eye deviation such as blinks, regular saccades, and slow drifts.

Principle 2: The algorithm quantified the saccadic eye deviations found in three amplitude bands; saccadic deviations below 0.4 degrees which are considered too low to be a saccadic intrusion (SI), between 0.4 and 4.1 degrees which are considered SIs, and over 4.1 degrees which are believed to be too large to be a SI. The amplitude spectra across these three bands were compared as a function of MWL. It was hypothesized that only the amplitude spectra for eye deviations between 0.4 and 4.1 degrees would vary systematically with MWL.

Principle 3: Only one algorithm would be published in the dissertation. Upon implementation of this technology to the real world in the future, it is not practical for different eye movement environments to require different sets of algorithms. One algorithm set should work in any eye movement environment.

Principle 4: The algorithm does not have to work for all kinds of eye movements. Visual tasks were limited to either a dot fixation task, a random dot fixation task, or a photo free-viewing task (see the Methods section). In the future, the algorithm may need to be improved for analyzing eye movements in more complicated visual tasks, but this algorithm can serve as a prototype for future work.

Principle 5: The real-time MWL estimation was limited to retrospective prediction for a period of about 30 seconds. The eye data were pre-recorded and the MWL was retrospectively calculated for the sample period. In the future, it would be nice to estimate MWL every second in real-time using the eye data for the previous one second or so.

Principle 6: The current algorithm used no participant-specific tailoring by the experimenter. Each participant had outlier data, but the algorithm was designed to handle those outlier data without ad-hoc treatment. The algorithm needed to be able to automatically process the raw data and quantify saccadic eye deviations without the help of an operator.

Principle 7: Related to the sixth principle in the dissertation algorithm, all parameter constants were the same for all participants. Despite individual differences, the same algorithm with the same parameter settings was used for everyone. For example, the distance between the two eyes was somewhere between 54 mm and 68 mm based on the participants' data. However, the parameter was set at 59 mm, so that the algorithm did not have to wait for the parameter input by the implementer. Likewise, some participants had more stable fixations while others had more scattered fixations, but the fixation criterion was set to 1 degree of visual angle for all participants. This was because all the parameters were determined by pre-existing theories, which leads to the most important rule, the theory based rule.

Principle 8: Perhaps most important of all, the algorithm was strictly theory-driven. In order to quantify saccadic eye deviation during fixational eye movements (FEMs), there were about ten small steps in the calculation. This represents of integration of many methods derived from several eye movement theories generally accepted in the vision sciences. The algorithm parameters were derived from eye movement demographic data reported by other researchers.

3.2.2 Eye Movements Classified by the Algorithm

Saccadic intrusions are the eye movements of interest in this dissertation, which are a specific type of saccadic eye movements that occur only during fixations (Figure 3.2). The eye usually makes two categorically distinctive kinds of eye movements: fixations and saccades. A fixation is a relatively stable eye gaze at an object, whereas saccades are fast eye movements that change fixation locations.

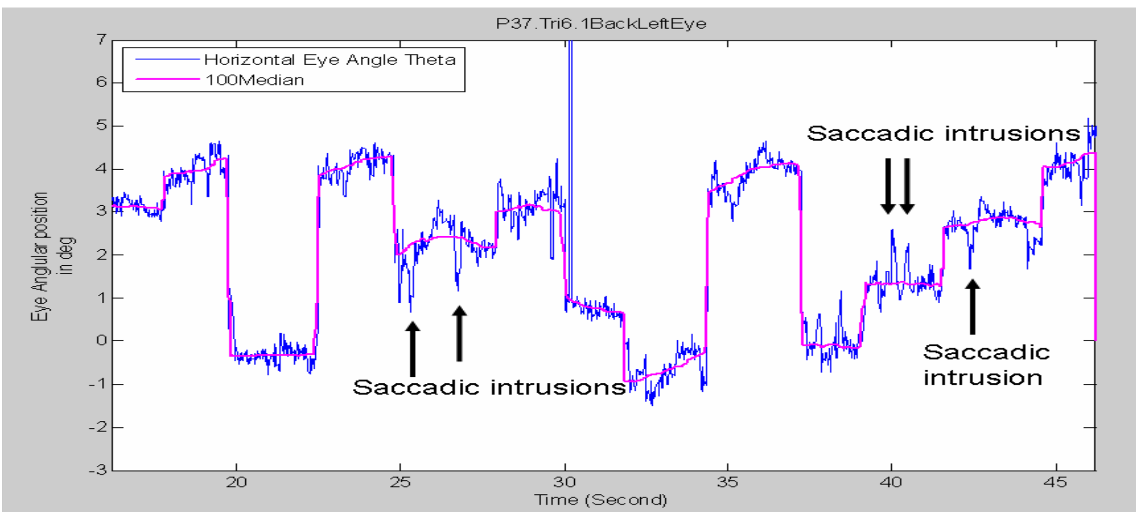

Figure 3.2. Demonstration of two distinct kinds of eye movements: fixational eye movements (FEMs, horizontal lines in pink) and regular saccades (vertical lines in pink). The x-axis depicts time and the y-axis depicts angular eye position. Algorithm 1 quantified saccadic intrusions, which only occur during periods of FEMs.

During a fixation, the eye is not perfectly stable, but rather is constantly jittering around a point. These small movements are called fixational eye movements (FEMs). Among FEMs, relatively large saccadic eye deviations sometimes occur, with the gaze direction changing by up to 4.1 degrees from the initial point in a saccadic manner and usually returning to the previous fixation point within 880 ms. These are called saccadic intrusions (SIs) because a saccadic eye deviation intrudes upon fixation (Figure 3.2).

The terms regular saccades, saccadic intrusions (SIs), saccadic eye movements and saccadic eye deviations are all similar to each other, and can be confusing. All the fast eye movements, including regular saccades and saccadic intrusions, are saccadic eye movements. Regular saccades occur outside a FEM period while SIs occur within a FEM period. When the eye gaze deviates from a fixation point in a saccadic (fast) manner, it is a saccadic eye deviation. Saccadic eye deviations can be categorized as SIs, a microsaccades, or gaze aversions. Microsaccades are smaller than SIs, while gaze aversions are larger. SIs range from 0.4 to 4.1 degrees in visual angle (Abadi & Gowen, 2004).

The algorithm first identified regular saccades that did not return back to the previous fixation point. Using regular saccades, FEM periods were determined. Between two regular saccades was a FEM period. If a fixation point suddenly changed to certain amplitudes and returned to the previous fixation point during the FEM period, that eye deviation was considered a SI. SIs, saccadic eye deviation within FEMs, were the special interest of the dissertation. The next several sections elaborate on each step used to quantify SIs within FEM sections.

3.2.3 Algorithm Procedure

Step 1. Raw Data Formatting

Table 1 below displays sample data produced by the Tobii eye tracker. Each row represents eye activity data for a 20 ms period. This sample consists of 10 rows over a period of 204 ms. There are 16 columns, 11 of which were used in the analyses. The timestamp represents the time in milliseconds from the beginning of the recording event. GazePointXLeft and GazePointYLeft are the left eye’s gaze location on the x-axis and the y-axis of the Tobii screen.

CamXLeft is the relative position of the eye in the Tobii video screen from 0 at the very left of Tobii's recorded video frame to 1 at the very right of the video shot. CamYLeft is the relative vertical position of the eye in the Tobii's recorded video. DistanceLeft is the diagonal distance in millimeters between the left eye and the Tobii camera. PupilLeft is the pupil diameter in mm. ValidityLeft shows if the eye activities were recorded validly by Tobii. When the Validity value is 4, all the other eye activity values are recorded in a negative value, indicating that the recording was not valid for that 20 ms (see the first two rows in Table 1). The same set of data is produced for the right eye as well.

TABLE 3.1

TOBII EYE DATA OUTPUT

Timestamp	Number	GazePointXLeft	GazePointYLeft	CamXLeft	CamYLeft	DistanceLeft	PupilLeft	ValidityLeft	GazePointXRight	GazePointYRight	CamXRight	CamYRight	DistanceRight	PupilRight	ValidityRight
24	0	-1280	-1024	-1	-1	-1	-1	4	-1280	-1024	-1	-1	-1	-1	4
44	1	-1280	-1024	-1	-1	-1	-1	4	-1280	-1024	-1	-1	-1	-1	4
64	2	624.4393	488.1239	0.683379	0.265134	595.803	3.951211	0	643.598	471.3026	0.385891	0.274401	600.9084	4.12131	0
84	3	620.0757	463.5478	0.683419	0.265481	593.1836	3.982841	0	632.2959	502.715	0.385833	0.274565	593.7278	3.958246	0
104	4	620.219	463.9774	0.683385	0.265682	592.9802	3.981527	0	633.6733	491.9644	0.385754	0.274794	598.5648	3.989285	0
124	5	619.4338	477.8438	0.683388	0.265764	592.9802	3.971203	0	637.1249	474.8097	0.385618	0.275002	598.5648	3.950633	0
144	6	611.452	462.8654	0.683323	0.265836	592.9802	3.943084	0	635.2385	479.9493	0.385509	0.275096	598.5648	3.953876	0
164	7	620.0853	471.6219	0.683279	0.26586	592.9802	3.940503	0	635.5865	465.5539	0.385449	0.275262	598.5648	3.948311	0
184	8	623.9995	487.8907	0.683223	0.265888	592.9802	3.943117	0	636.2112	484.4063	0.385348	0.275386	598.5648	3.935605	0
204	9	622.0651	477.2586	0.683182	0.265922	596.0291	3.961778	0	636.7228	483.6813	0.385317	0.275483	597.4232	3.927357	0

Sample of Tobii eye data for the first 204 ms.

The eye activity data was in "txt" file form on a Windows computer. The data was read by the numerical computing software Matlab (MathWorks) version 2007a and analyzed using the original algorithm written in the Matlab programming language.

Step 2. Missing Data Identification

Video-based eye trackers such as the Tobii always have missing data due to blinks because the eyelids interfere with the video recording of the eye. Figure 3.3 is a sample data set showing horizontal eye movement data, including missing data, for 64 seconds. The x-axis represents the horizontal pixel where the eye gaze landed from Pixel 1 (very left side of Tobii screen) to Pixel 1280 (the right side of the screen). The y-axis is the time in seconds. The left eye was staring at the center of the screen (at Pixel 640) from the 8 to 58 second marks. The eye gaze mostly landed around Pixel 640. However, there were some periods where the horizontal eye gaze went down to Pixel 0. These were considered missing or invalid data that Tobii could not record, most likely due to eye blinks. These missing data were replaced using the method described in the next section.

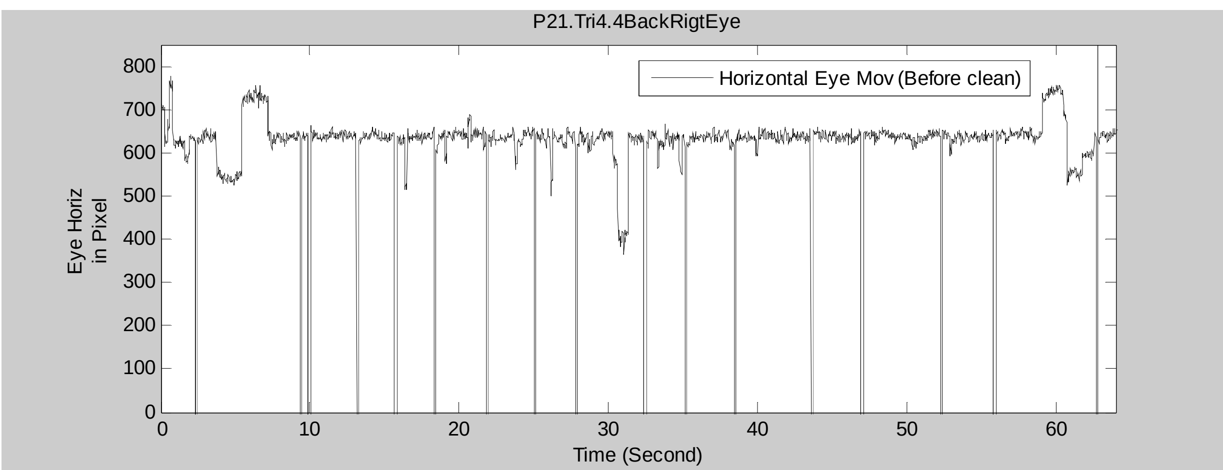

Figure 3.3. Sample raw data of horizontal eye movements including missing data indicated by a horizontal eye position of 0.

Step 3. Missing Data Replacement

The purpose of this clean-up section in the algorithm was to replace these missing data with something else so that the missing data would not affect the quantification of eye deviation.

In order to minimize the effect on the eye deviation calculation, missing data were replaced with the eye data from the previous moment. Missing data were defined as any horizontal eye gaze data that did not land on the Tobii monitor between Pixels 1 and 1280. Missing data were replaced with the average of the previous 10 samples (200 ms). Replacement was done only for the quantification period, which was from the 14 to 58 second marks. Figure 3.4 shows the replacement procedure for a single trial's data. The same replacement method was used for the eye distance data from the Tobii camera as well.

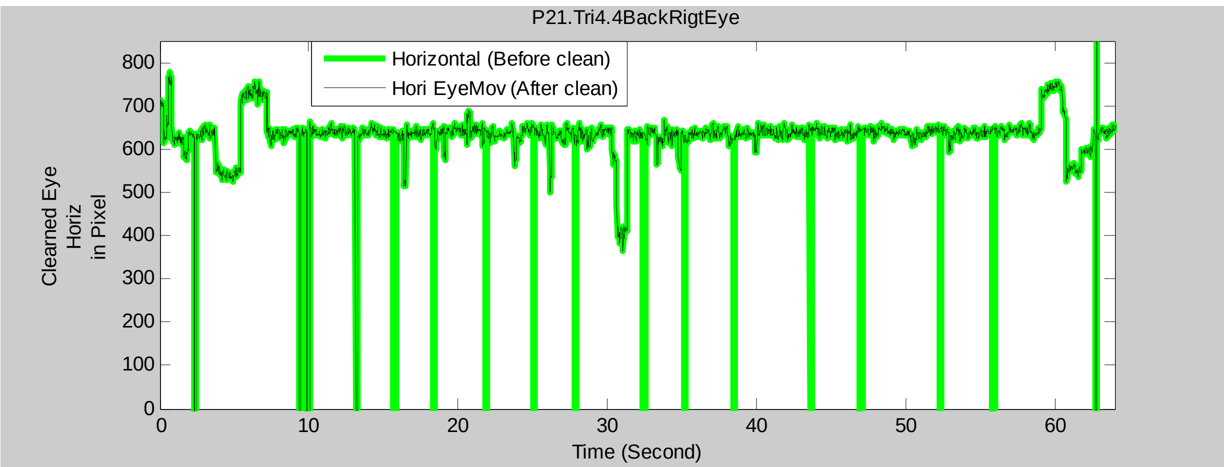

Figure 3.4. Missing data, mainly due to blinks, during the quantification period of 14-58 sec were automatically replaced with the previous moment eye data.

Step 4. Eye Angle Calculation

After cleaning up the raw eye data that Tobii produced, the eye angle theta was calculated. Theta was defined as the eye angle on the horizontal plane in relation to an assumed common face angle (Figure 3.5). Vertical eye angle iota was calculated from the distance data and the vertical eye gaze data using the same methods as those used to calculate theta, described below.

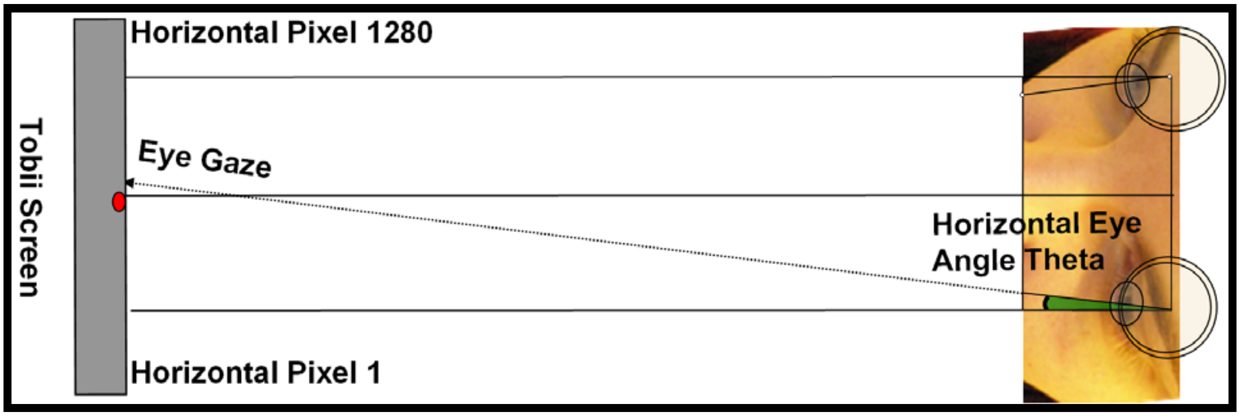

Figure 3.5: Horizontal eye angle theta was calculated in a function of horizontal gaze and distance between Tobii and the eye.

The two concepts of eye angle theta and eye gaze are very similar to each other, and the terms are sometimes used interchangeably, however, the two concepts are distinct in this dissertation. Tobii's eye gaze data needed to be converted into eye angle theta. Gaze or eye gaze is a term to represent what location on the physical computer screen the eye's focus falls onto, measured in pixel location (Figure 3.6, top). On the other hand, the eye angle theta focuses on the eye's direction itself, measured in degrees of visual angle (Figure 3.6, bottom). Since this dissertation is interested in the physiological eye responses, the eye angle theta is more relevant than eye gaze on the Tobii's screen.

There are two reasons for using eye angle theta. This algorithm used only two kinds of eye data, eye gaze (Figure 3.6, top) and the distance between Tobii and the eye (Figure 3.6, middle), to calculate the eye angle theta (Figure 3.6, bottom). First, this clarifies the source of the eye deviation results since the data are plotted in terms of the eye movement alone, not including body or head movements. Second, application of the dissertation results to the real world would be easier with fewer input data. For these two reasons, the face angle was set to the constant angle zero, using the chin rest, and the distance between the eyes was assumed to be 59 mm, using the average from the eight pilot participants.

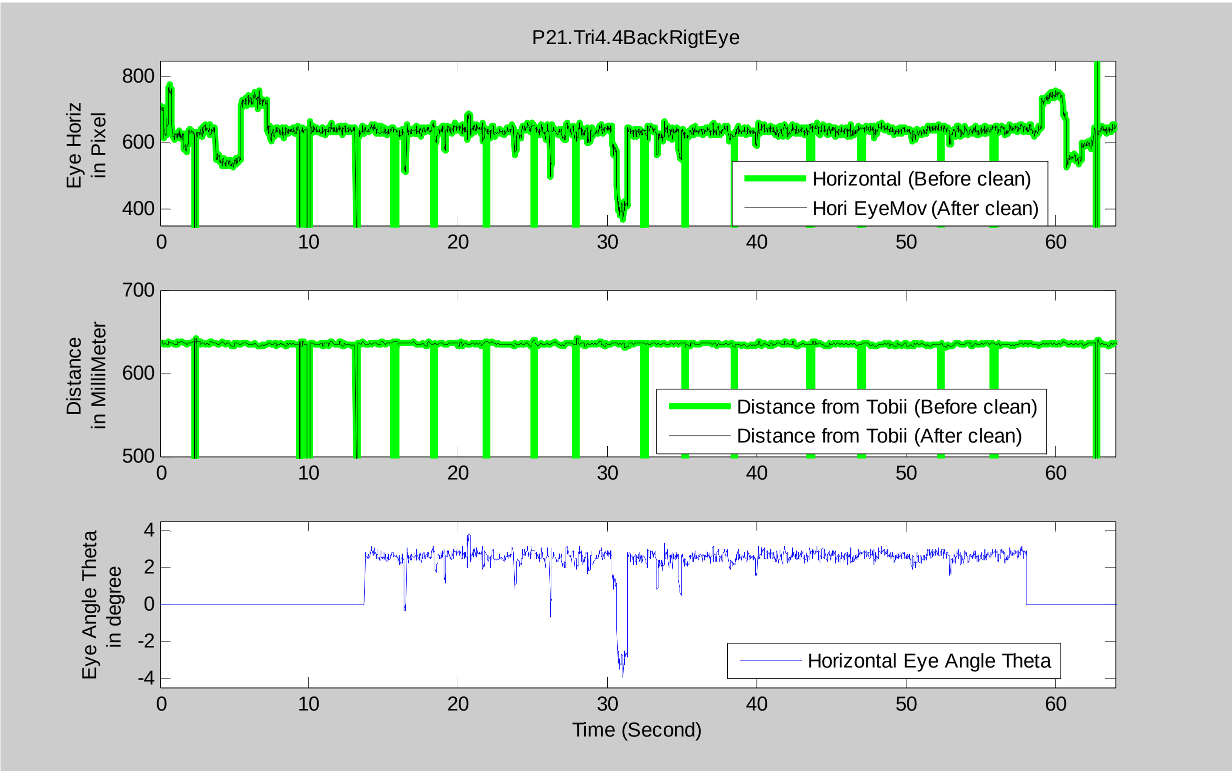

Figure 3.6. The horizontal eye angle theta (bottom) was calculated from the horizontal eye gaze (Top) and the distance (middle).

Step 5. Saccadic Eye Movement Detection

The goal of Steps 5 and 6 was to categorize each 20 ms event as either part of a regular saccade or part of a fixation. In a real world situation, each fragment of eye movement data should be categorized into one of three groups: regular saccades, fixations, or smooth pursuits. However, in this dissertation there were no moving objects and face angle was fixed, so no participants made smooth pursuit eye movements. Given that there were no smooth pursuits, the two other kinds of eye movements (regular saccades and fixations) were mutually exclusive during the dissertation trials. All eye movement data was classified as either being a fixation or a regular saccade. Since SIs occur only during fixation periods, the author's main interest was in the fixation periods (refer to Figure 3.2).

Detection of regular saccades was performed in two steps. First, all the possible candidates of regular saccades were detected using eye angular amplitude as a criterion. Second, the number of the regular saccade candidates was reduced to those more likely candidates using a round-trip criterion. The detail of these two sub-steps is elaborated below.

In the first sub-step, the algorithm detected any eye movements with one degree amplitude or greater within any 40 ms period. The algorithm used a 2-dimensional moving average coordinate method, which computed moving average scores for two samples (40 ms) and compared their moving average scores (Figure 3.7). The first moving average coordinate's temporal window was from 0ms to 40 ms in advance (in the future), while the second temporal window was from -40 ms to -1ms (in the past). The distance between these two moving average coordinates represent the eye angular amplitude on the two-dimensional plane during a period of about 40 ms (see Figure 3.7 below). If the moving average coordinates move more than 1 degree of visual angle, it was considered a candidate of a saccadic eye movement.

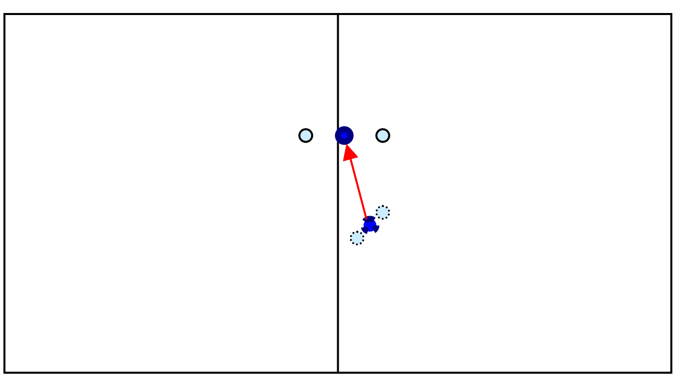

Figure 3.7. The algorithm detected saccadic eye movement when the gaze on the 2-dimensional plane moved more than 1 degree of visual angle in any direction. The distance was calculated using the average of two samples (40 ms) in the immediate past and another average of two samples (40 ms) in the immediate future. Each of the two dark dots represents the median

position of the two adjacent dots. Each of these two adjacent dots represents an eye gaze location for 20 ms.

Multiple samples (two samples or 40 ms), rather than a single sample (for 20 ms), were used to identify saccade candidates because one saccadic eye movement could be recorded by Tobii into two eye data events, separating one saccadic eye movement into two fragments. In that case, one saccadic eye movement is recorded in three 20-ms events. If the temporal windows of the two moving average were 20 ms width each and compare just the two 20-ms events next to next, the two fragments of a saccadic eye movement would not be caught by the algorithm.

Figure 3.8 (top) shows the horizontal eye angle theta and the vertical eye angle iota, and Figure 3.8 (bottom) shows their corresponding eye angular amplitudes during each 20ms interval. The vertical bars at the bottom of the figure indicate the identified saccadic eye movements. There were about 20 candidate saccade eye movements in the trial depicted. The identified saccadic eye movements included both regular saccades and saccadic intrusions. In the next section, these candidate saccadic eye movements were categorized as either regular saccades or saccadic eye deviations.

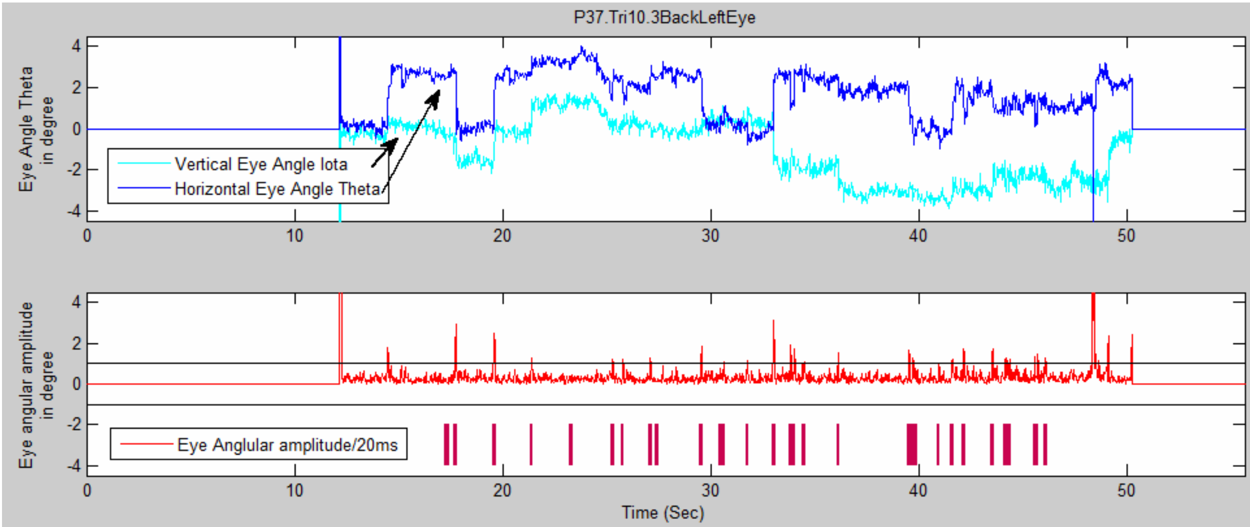

Figure 3.8. (Top) A sample data set of horizontal eye angle theta and vertical eye angle iota during a stable dot fixation task. (Bottom) Regular saccade candidates (the vertical bars) were detected by the criteria of eye angular amplitudes greater than 1 degree. Eye angular amplitude was calculated from theta and iota.

Step 6. Fixational Eye Movement Detection.

The second sub-step to detect regular saccades was to eliminate saccadic eye movements within FEMs from the result of sub-step 1. The definition of a FEM (fixational eye movement) was a series of eye movements that either stayed perfectly within 1 degree of visual angle from the fixation point or stayed mostly within 1 degree except for short temporal deviation more than 1 degree of visual angle. It was considered one FEM if the gaze started fixating, saccadicly deviated more than 1 degree, returned to the fixation point within 900 msec, and resumed a fixation within 1 degree radius from the original fixation point. This can be called saccadic eye deviation, including SIs, microsaccades, and gaze aversion.

Four fixed parameters were used in this section of the algorithm: the fixation criteria of 1 degree, the return acceptance period of 900 msec, the fixation return criteria 1 degree, and the fixation duration 100 ms. The maximum fixation amplitude 1 degree and minimum fixation duration 100 ms were adapted after the criteria in Castelhano, Mack, & Henderson (2009). The maximum SI duration criterion was adapted from Abadi & Gowen (2004), in which they observed SI duration up to 880ms. The criterion was rounded up to 900ms in the dissertation.

The algorithm first identified any saccadic eye movements (Figure 3.9 middle). At this moment, the saccadic eye movements were undistinguishable from regular saccades and saccadic eye deviations. After the algorithm identified a saccadic eye movement, it determined whether the gaze returned to the initial fixation point and started eye deviation within FEM sections.

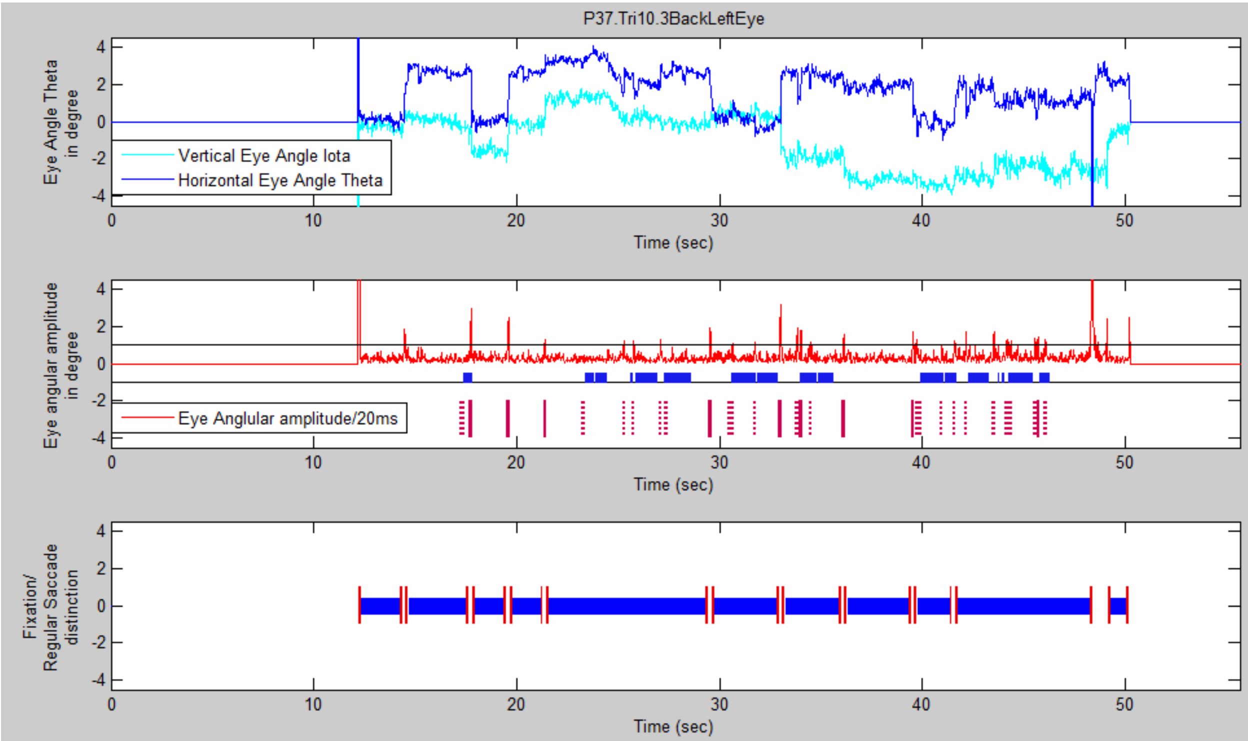

Figure 3.9. (middle) shows saccadic eye movements over 1 deg. If the eye position did not return to the previous position after the saccadic eye movement, that saccadic eye movement was considered to be a regular saccade. The eye movements were categorized as either part of a regular saccade (red vertical lines in the bottom figure) or a fixational eye movement (FEM) section (blue horizontal lines).

Step 7. Eye gaze deviation from fixation baseline

So far, the algorithm has identified regular saccades and FEMs in the eye movement records. In Step 7, the algorithm detects saccadic eye deviations within each of the FEM periods. These saccadic eye deviations during a FEM section were more likely to be the eye movements of interest in this dissertation, saccadic intrusions (SIs). A SI is usually (more than 75% of the time) a round trip (Abadi & Gowen, 2004), which means that eye gaze deviates from a fixational point for a few hundred milliseconds and returns to the fixational point in a saccadic manner. The algorithm used this characteristic of SIs to help identify them using a two boxcar method. The two boxcars had different temporal window sizes. The large boxcar had a window size of 2,000 ms while the small boxcar had a window size of 100ms. The large boxcar served to calculate the base horizontal fixation position while the small boxcar served to identify instantaneous horizontal deviations of eye position. Since SIs occur only on the horizontal plane of eye movements (Abadi & Gowen, 2004), calculation of the horizontal eye deviation was sufficient for the eye deviation quantification.

Figure 3.10 is a sample data set to explain this section of the algorithm. The horizontal eye angle theta was categorized as either part of a regular saccade or part of a fixation. Within each section of the FEMs, the two boxcar moving median method was applied. The result is shown in Figure 3.10 bottom. The black jumpy line is the moving median with the small window. The pink straighter line is the moving median with the large window. Notice that the pink line seems to successfully represent the fixation baseline while the black line seems to successfully represent the short period eye deviation. Also the FEM sections seem to successfully make each section independent from the other FEM sections. Both the black and pink lines are unaffected in each FEM section from the other FEM sections.

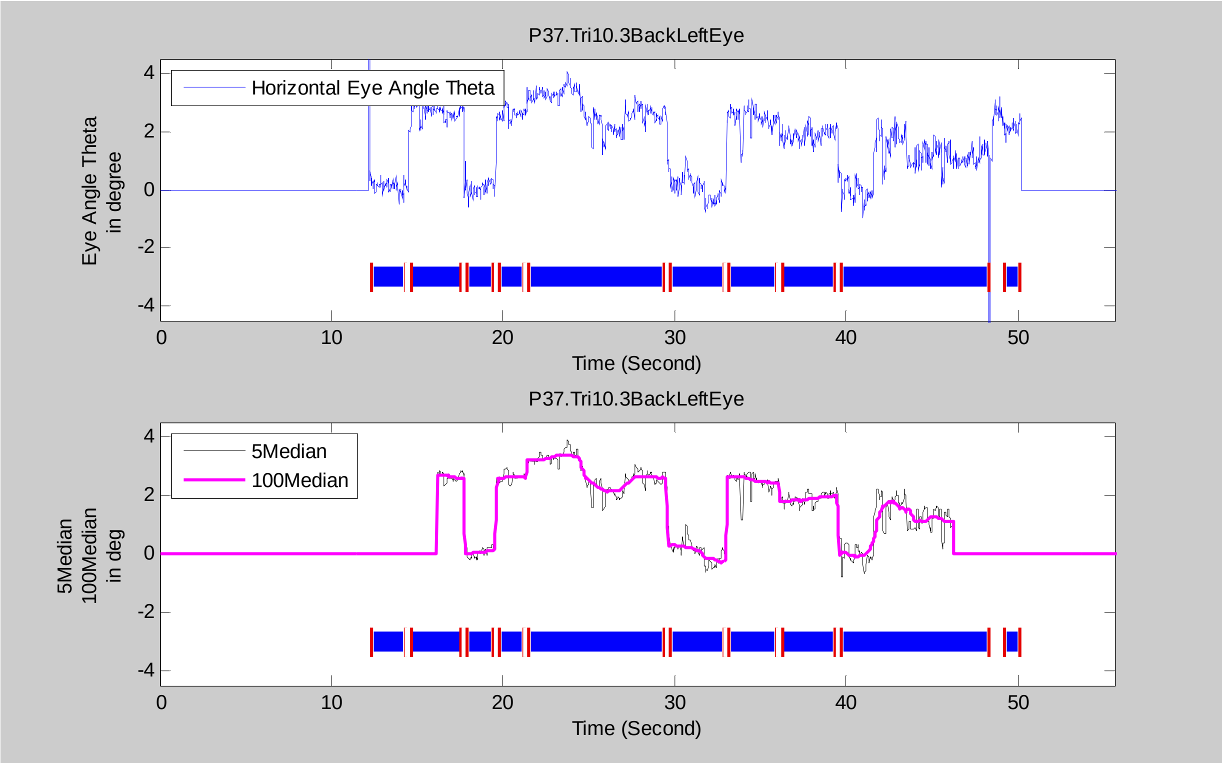

Figure 3.10. the distinction between regular saccades and FEMs.(top) and
the two moving median lines within each of the FEM sections. (bottom)The large window median scores represent fixational eye positions. The small window median scores represent short term eye deviations.

The small boxcar median value took five samples (100 ms) from 50 ms in the past to 50 ms in the future. Since this is a retrospective prediction study, the future event data were available. Almost all the parameters in the algorithms in this dissertation were adopted from the work of other researchers. However, this number, five samples (100 ms), was not adopted from demographic data and was determined from the Tobii's temporal resolution and noise level. Abadi and Gowen (2004) show that SI duration ranged from 10 ms to 880 ms. However, the Tobii eye tracker has a temporal resolution of 20 ms, and therefore does not record events with such a short duration. Also Tobii's relatively coarse spatial resolution produces spatial noise. Considering these two confounding factors, Algorithm 1 required at least three samples to identify them as a legitimate eye deviation. If there was an eye gaze deviation for 60 ms (three samples), the three sample deviation could outnumber the other two samples in the “five” median window, predominate the window, and have the control of the median score of the 100 ms window (five samples), and the algorithm could detect the eye deviation.

The large boxcar had the window span of 2000 ms from 1000 ms in the past to 1000 ms in the future. However, the window was not continuous, having a gap of 500 ms in the middle. This number was adopted from Abadi and Gowen (2004), who showed that most SIs had durations up to 500 ms. The purpose of having the two different size windows was to emphasize the gap between a fixation eye position represented by the large boxcar window and the deviated eye position represented by the small boxcar window. In order to emphasize the gap, the large window did not use the eye data close to the small window area. Figure 3.11 shows an example of the two window sizes. Around the 27 second mark, the large window took eye data for a total of 1500 ms while the small window took eye data for 100 ms. The eye deviations seem to be well detected in this example.

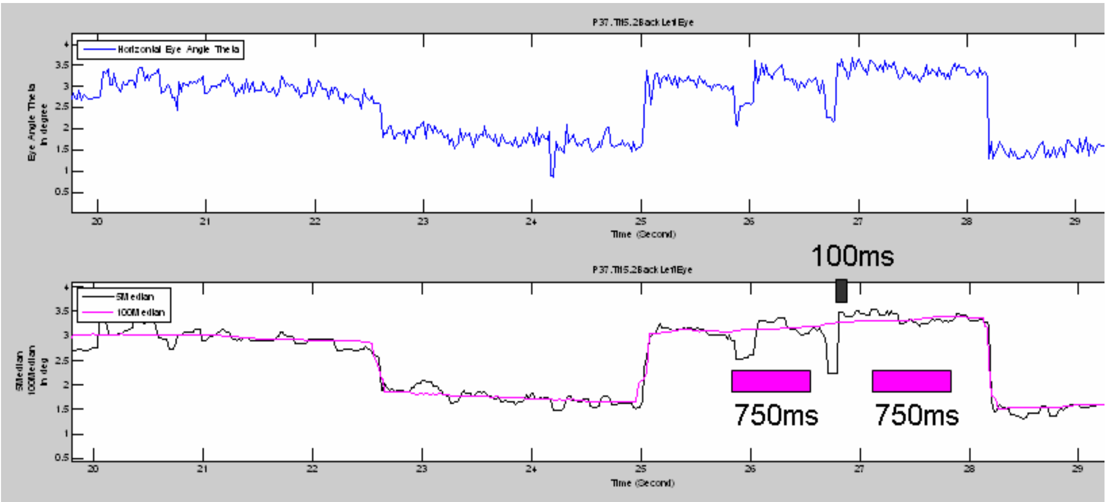

Figure 3.11. Within each FEM section, two moving median scores were calculated for every 20 ms. One moving median had a window span of 2000 ms with a 500ms gap in the middle. Another moving median had a window of 100 ms.

Difference values were calculated between the two moving median positions. The difference is plotted on the red line in Figure 3.12. The red line represents an eye angular amplitude value of the eye deviation every 20 ms. The most useful aspect of Algorithm 1 is that it quantifies saccadic eye deviation from the assumed fixation position. The algorithm does not need to know the actual visual target location, however, since it automatically detects fixation periods and estimates fixation positions, and then calculates saccadic deviation from the estimated fixation position. Because all of the information needed is found in the eye movement data, Algorithm 1 can analyze any eye data as long as there are FEMs.

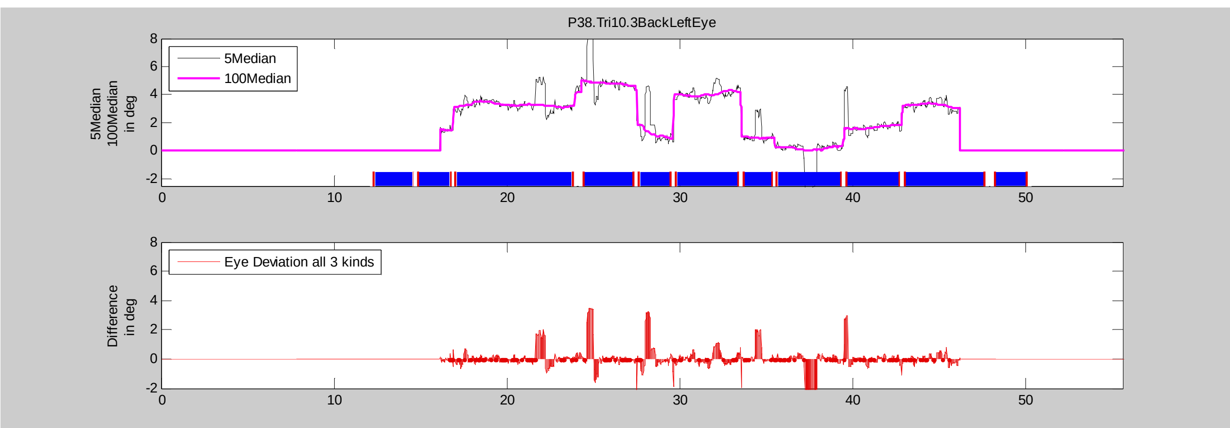

Figure 3.12. Using the two moving median lines (, top), the difference values (, bottom) was calculated. The difference values still containing noise from regular saccades would be cleaned up in Step 8.

Step 8. Data Cleaning

Before the final quantification of the saccadic intrusions as the dependent variable, the difference values were cleaned up. The difference values (the red line in Figure 3.12 bottom) were a mixture of several eye movements, including eye deviation during FEMs and some other values. This step involved cleaning up other kinds of eye deviation, such as regular saccade residual, non-FEM sections, and invalid data.

First, residuals of regular saccades were identified and removed from the calculated difference values (the red line in Figure 3.12 bottom). The residuals of regular saccades came in the red line when the calculated fixation baseline (pink line) and the actual eye gaze (black line) did not perfectly match. These regular saccade residual signals were deleted. Whenever a regular saccade occurred, the difference values were nullified to a value zero at the regular saccade moment and one event before (20 ms) and after (20 ms) the regular saccades. See the difference of before (Figure 3.13, middle) and after the clean-up (Figure 3.13, bottom) of the regular saccade residuals in Figure 3.13. Before cleaning up the residual of regular saccades, there were big red lines due to regular saccades at the 17, 24, and 44 second marks. They were removed in Figure 3.13 bottom, but large saccadic eye deviation within FEMs remained between 30 and 40 second marks.

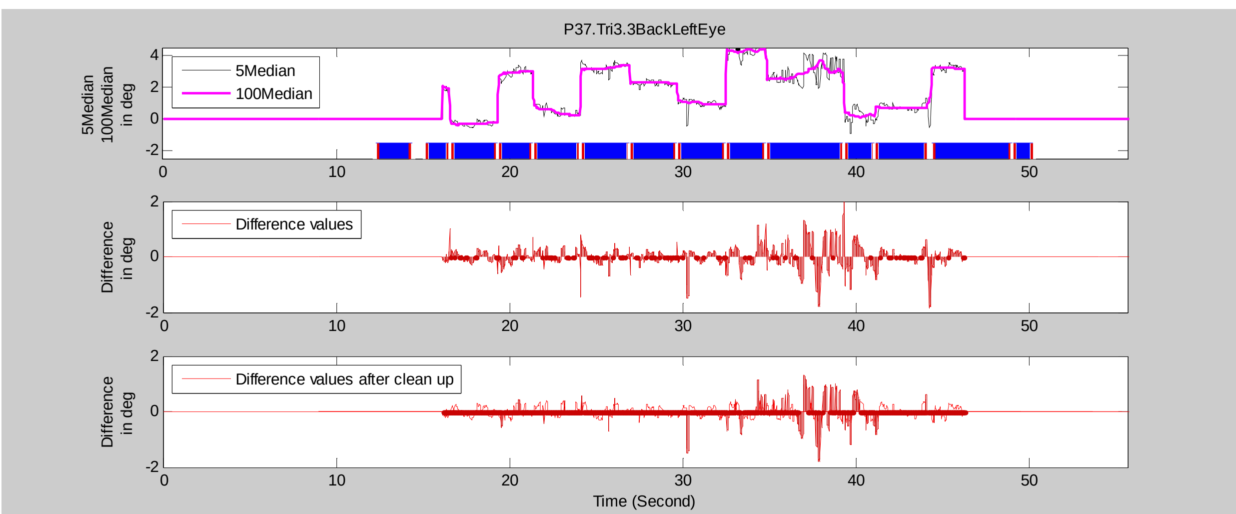

Figure 3.13. The difference values (Figure 3.13, middle) were cleaned up into the saccadic eye deviation within FEMs (Figure 3.13, bottom) after removing the deviation values due to regular saccade at the 17, 24, and 44 second marks.

Second, saccadic eye deviations within non-legitimate FEMs were removed from the difference values (the red line in Figure 3.12 bottom). Fixational eye movement (FEM) sections smaller than 1000 ms were removed because the large median window (See Step 7) could not take an adequate data size from such a short period of time. The current algorithm unfortunately would not work properly with those small FEM areas. Figure 3.14 is an example of an FEM section that was too short at the 21 second mark. The bottom figure shows a big white gap, indicating that the FEM section was not used. All the eye deviation values in this period are not included in the final SI measure.

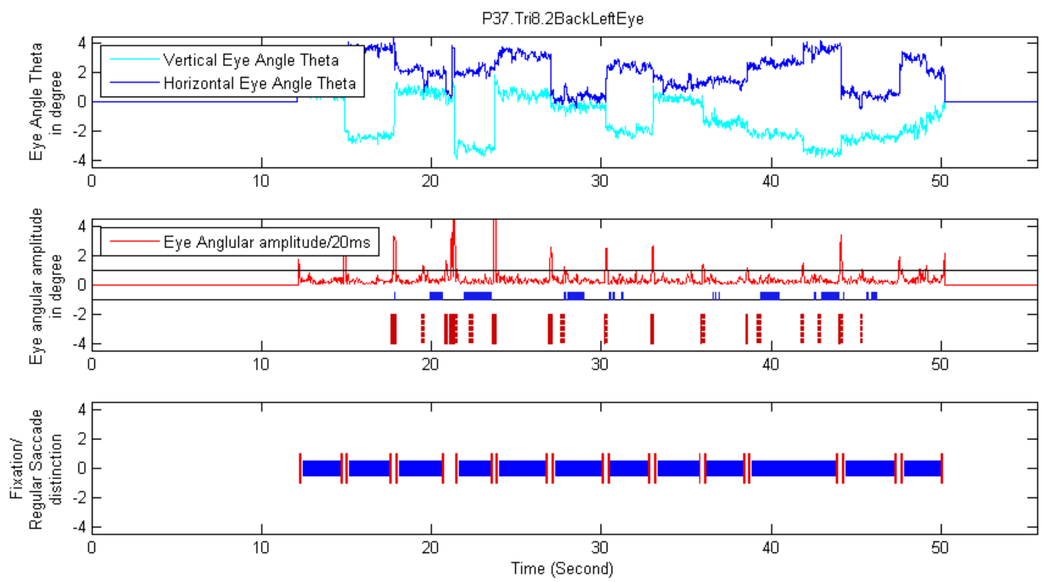

Figure 3.14. An example of a FEM section at the 21 second mark removed from the analysis due to its short window size. Saccadic eye deviation within FEM sections smaller than 1000 mg were not counted in Algorithm 1.

Finally, invalid data were filtered from the eye deviation evaluation. The validity data were from Column 9 in Table 1 and produced by Tobii every 20 ms. If a Tobii recording event for 20 ms was invalid, that data event for 20 ms was removed from the eye deviation analyses.

The saccadic eye deviation values (the red line in Figure 3.13 bottom) should have only eye deviation from the FEMs after removing the three kinds of unwanted eye deviation such as regular saccade residuals, illegitimate small FEMs, and Tobii invalid data events.

Step 9. Quantifying SI for Each Trial

After removing the noise data for eye deviation quantification (Figure 3.15, middle), the eye deviation values occurring every 20 ms were classified into three groups (Figure 3.15, bottom). Abadi and Gowen (2004) show quantitative differences in the eye deviation. Following them, eye deviations only between 0.4 and 4.1 degrees were considered as SIs. The other eye data were labeled as "smaller eye deviation" or "larger eye deviation." Therefore, Algorithm 1 produced three single values for each trial data: an average smaller eye deviation value, an average SI value, and an average larger eye deviation value for the period.

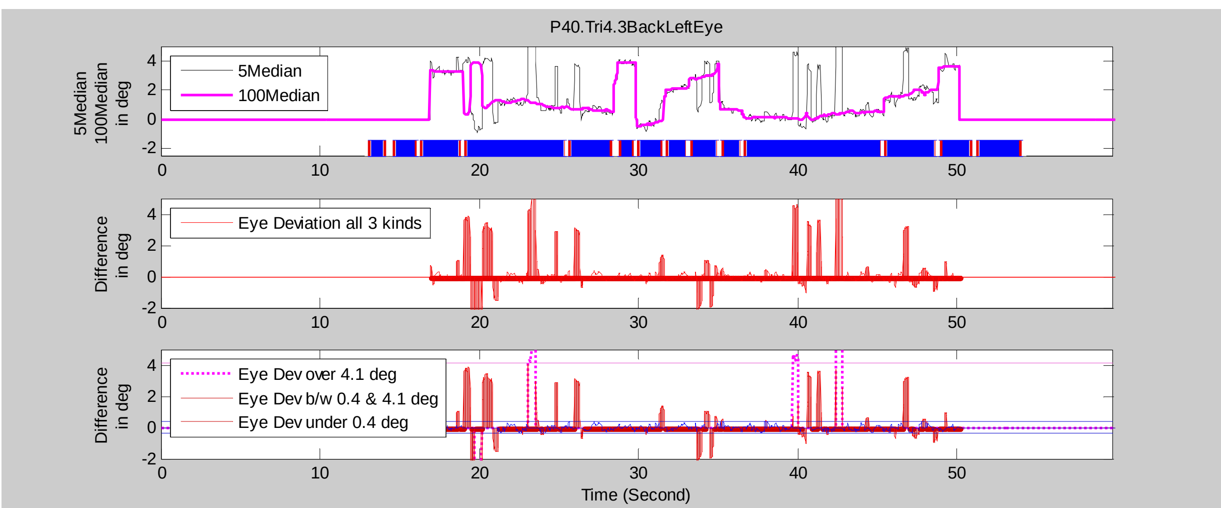

Figure 3.15. After taking the differences between the two median scores (top) and after removing the residual of regular saccades, the eye deviation within FEMs alone remained (middle). The eye deviation within FEMs were categorized into three amplitude levels (bottom).

After classifying eye deviations into the three categories, the algorithm calculated the mean score of each of the three categories of the eye deviation for a typically 30 second trial evaluation period. The mean score was calculated by cumulating all the SI values and by dividing the cumulative SI value by the second. The mean score was used as the dependent variable for the trial. Later, these three single values were also used in the regression analysis to evaluate the relationship with MWL.

3.2.4 Algorithm 1 Conclusion

Each processing step in Algorithm 1 was based on pre-existing methods or parameters of physiological data provided by other researchers. Theoretically, the algorithm was designed to cumulatively quantify both the duration and amplitude of saccadic eye deviation within FEMs during a single experimental trial. This quantification was made by calculating eye deviation from the estimated fixation positions, not from the actual visual stimulus on the computer screen. All the information needed for Algorithm 1 is taken from the computed eye deviation in visual angle. It does not need to know the actual visual stimulus location.

Since the algorithm automatically detects FEMs (fixational eye movements) in eye data and uses the FEMs as part of MWL estimation, it should work on any eye movement tasks as long as a large percentage of the eye movements are fixations, rather than regular saccades or smooth pursuits. Studies 1, 2, and 3 employ tasks that contain a large percentage of fixations.

3.3 Fourier Transformation Algorithm

This section explains Algorithm 2 which is based on the Fourier transformation. Algorithms 1 and 2 are similar. Both algorithms take eye movement data as input, automatically searching for saccadic intrusions (SIs), and then quantifying SIs into a single value called either the SI measure in Algorithm 1 or power density measure in Algorithm 2. It is believed that these two alternative analyses may offer complementary evidence of the effects of MWL on eye movement patterns (SIs). However, there are some differences between these two approaches. While Algorithm 1 quantified the impact of MWL by detecting SIs in the eye movement recordings, Algorithm 2 evaluates the effects of MWL by using the Fourier transformation to quantify changes in the amplitude spectra of the eye movement data. This section explains how the Fourier transformation was used to detect SI behavior.

3.3.1 What is the Fourier Transformation?

The Fourier transformation (FT) is a series of calculations that converts a set of time-domain data into a set of frequency-domain data. It is named after the French mathematician, Joseph Fourier (1768-1830), who first developed the technique. Fourier showed that any complex waveform can be decomposed into a series of component sine waves. For instance, a time varying signal like sound can be decomposed into the component frequencies of power and relative phase information. Likewise, applied to saccadic eye movements, the FT can be used to identify the component frequencies and power. Changes in saccadic eye movements with MWL should be reflected in the frequency and power spectra of the FT.

For example, Figure 3.1 shows six depictions of simulated eye movement data (Figure 3.1, left column) and their FT conversion (Figure 3.1, right column). The simulated data in Figure 3.1a (left) represent a series of horizontal eye movement data for five seconds, during which the eye gaze stays at the center and deviates to the right for a dwell time of 60 ms with an amplitude of 1 deg. An example SI like this can be converted to the frequency-representation such as Figure 3.1a (right), using an equation below.

$$\text{Valley frequency (Hz)} = 1\ /\ \text{Dwell time (sec)} \tag{3.1}$$

In Equation 3.1, valley frequency represents a wave cycle at the end of the fundamental frequency component, identified by the red arrows in the right column of Figure 3.1. The right side of the equation is the reciprocal of dwell time, depicted as the rectangular shapes in the column of Figure 3.1.

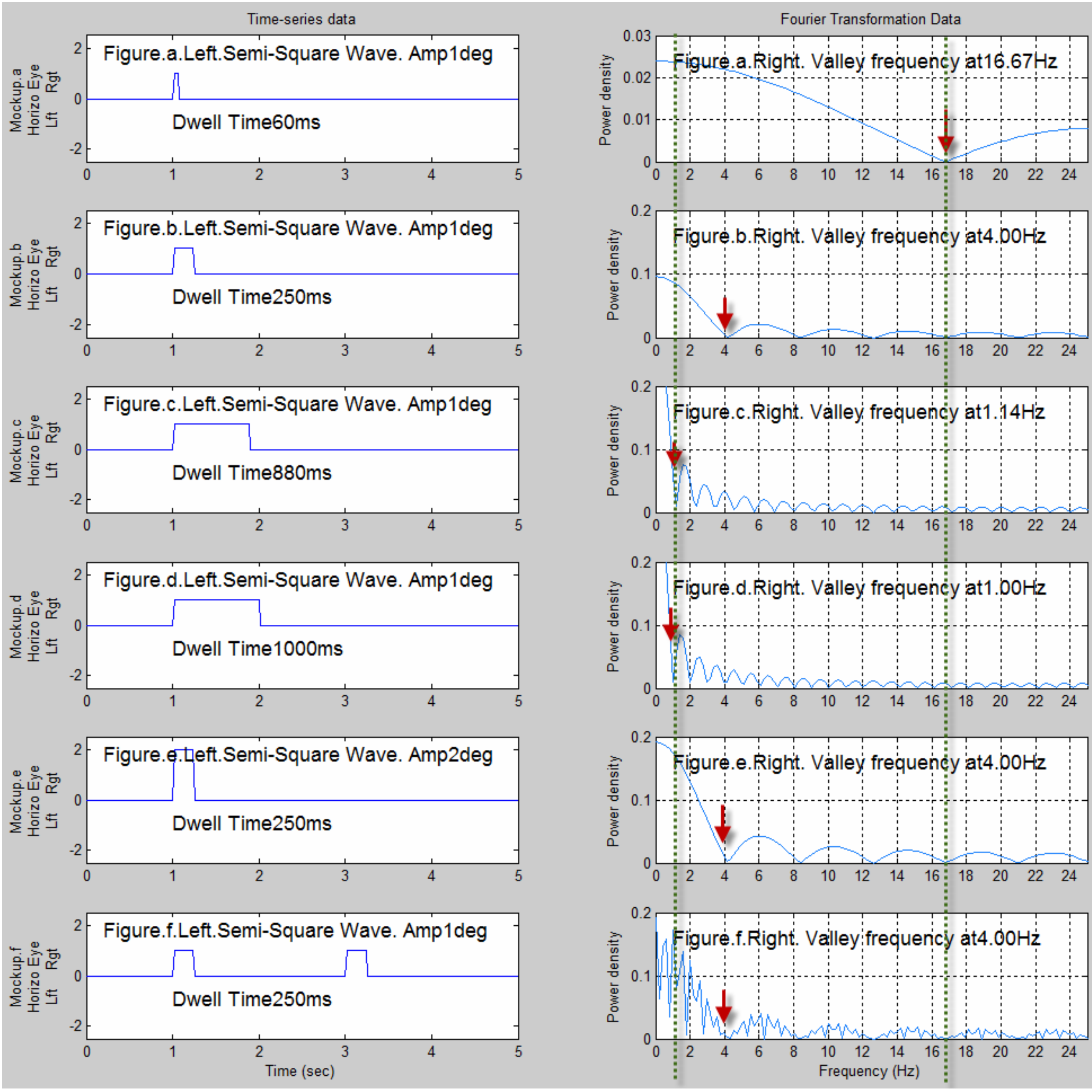

Figure 3.1. Six sets (a, b, c, d, e, and f) of simulated horizontal eye movement data (left column) and their frequency components after Fourier Transformation (FT) analysis (right column).

Since one of the SI characteristics was defined as an eye deviation with a dwell time from 60 ms (Figure 3.1a) to 880 ms (Figure 3.1c), the corresponding power density in Fourier space ranges from 1.14 Hz (Figure 3.1c) to 16.7 Hz (Figure 3.1a). Figure 3.1c (left) is an example of a long dwell time at the edge of the SI range. Fundamental frequency components lower than 1.14

Hz, depicted on the right side of Figure 3.1c and Figure 3.1d, were not included in the measure that represent SIs.

The analyses in this dissertation divide FT frequency components into three bands: the low frequency band ranging from 0 to 1.14 Hz which corresponds to eye deviations too long to be SIs, the middle frequency band ranging from 1.14 to 16.67 Hz which corresponds to eye deviations of the correct duration range to qualify as SIs, and the high frequency band ranging from 16.67 to 25 Hz which corresponds to eye deviations too short to be SIs. Average power density values (such as the y-axes in the right column of Figure 3.1) were calculated for each of these three bands in each trial of the experiments. These three values were then used as the dependent variables for the FT analyses. These frequency differences are just one aspect of the FT analyses, as shown in Table 3.1.

TABLE 3.1

CONVERSION BETWEEN TWO REPRESENTATIONS OF EYE MOVEMENT DATA

	Time-series eye movement data	Fourier Transformation frequency representation	Example
1	Dwell time	Frequency	Fig 3.1a, b, c, d
2	Amplitude	Power density	Fig 3.1b, e
3	Occurrence Rate	Power density	Fig 3.1b, f

Time series eye movement data can be converted into a combination of frequency and power density via the Fourier transformation.

Some of the plots in Figure 3.1 demonstrate another aspect of the FT: power density. If a SI like Figure 3.1b (left) occurred, it would be reflected in the average power density in the middle band ranging between 1.14 and 16.67 Hz. If a SI with an amplitude two times larger like Figure 3.1e (left) occurred, it would cause the average power density of the middle band to be almost two times higher. Similarly, a reoccurrence of SI behavior (Figure 3.1f, left) will affect its power density (Figure 3.1f, right).

3.3.2 Procedures for Implementing the Fourier Transformation

This section explains how the FT calculation was implemented in this dissertation. Algorithm 2 was developed to quantify Fourier transformation power densities of the eye movement data into dependent variables (DVs). FT itself was provided by the Matlab programming language as a function named fft. Algorithm 2 set up the input of the fft function before Matlab did the FT, and Algorithm 2 summarized the output of the fft function.

Step 1. Eye Angle data

The eye angle theta was calculated with the same method as Algorithm 1, using the eye distance and horizontal gaze landing positions. Missing data were replaced with the same method as Algorithm 1, taking the average of the previous 10 values. Also as Algorithm 1 did, Algorithm 2 used only the horizontal eye movement data because SIs occur only on the horizontal plane (Abadi & Gowen, 2004). The requirements of the eye data were the same; the eye data needed the sampling rate of 50 Hz (every 20 ms).

One small difference from Algorithm 1 was the time data. Tobii is supposed to record eye activity data every 20 ms. However, the sampling time fluctuates between 19 and 20 ms by 1 ms. Unlike Algorithm 1, Algorithm 2 ignored this time fluctuation because FT requires equal time intervals. Algorithm 2 assumed that eye events were collected every 20 ms. Figure 3.2 shows sample eye data after replacing the missing data. For this participant the evaluation period was 35.4 seconds from the 19.2 to 54.6 second marks. These eye angle data were the input data to the fft function in Matlab.

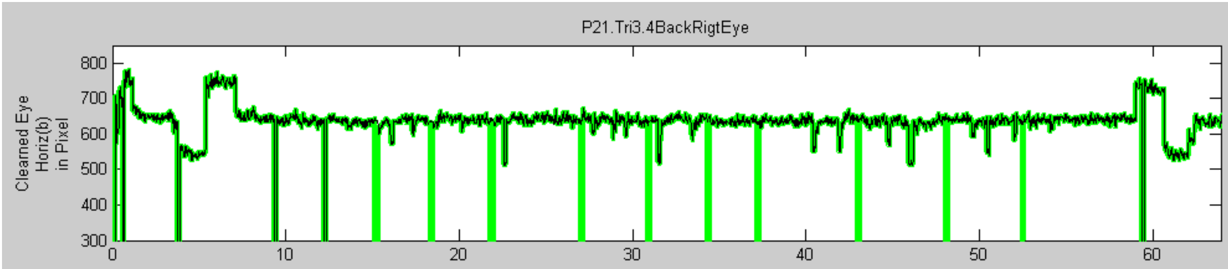

Figure 3.2. Horizontal eye movement record for one trial. The period between the 19.2 and 54.6 second marks was used for the Fourier transformation analysis.

Step 2. FT calculation

FT used the eye movement recorded within the designated evaluation period. The designated evaluation period was the same as Algorithm 1; typically for about 30 seconds in each trial from the onset of the fourth audio stimulus to the end of the 15th audio stimulus. The record consisted of approximately 1500 samples of eye position during this time period. The eye data count is important in FT. Fast Fourier transformation (FFT) is a specific kind of FT; it takes the data size of 2 to the power of n, such as 2, 4, 8, 16, or more because calculation is exponentially much faster with these data sizes. The dissertation uses FFT by padding the data to a size of a power of n. If the evaluation period was 30 seconds, the data length was 1500 counts (50 Hz x 30 seconds). The Matlab FFT function filled out the data with value 0 until the sample size became the next “2 to the power of n” value, which was 2048 with 548 zero replacements (2048 – 1500) in this case the Matlab FFT function returned the power spectrum values for the frequencies. The FFT calculation took less than 1 ms for 1500 sample data. The calculation speed is important to future applications of this research on MWL estimation because one of the goals is to estimate MWL in real-time. Since the FFT calculation is fast enough for real-time calculation, FFT is still an alternative method to quantify MWL in real-time in the future. A sample result of FT is shown in Figure 3.3, which is a conversion of a time-domain representation, Figure 3.2, using the same data set for 35.4 seconds from the 19.2 to 54.6 second marks.

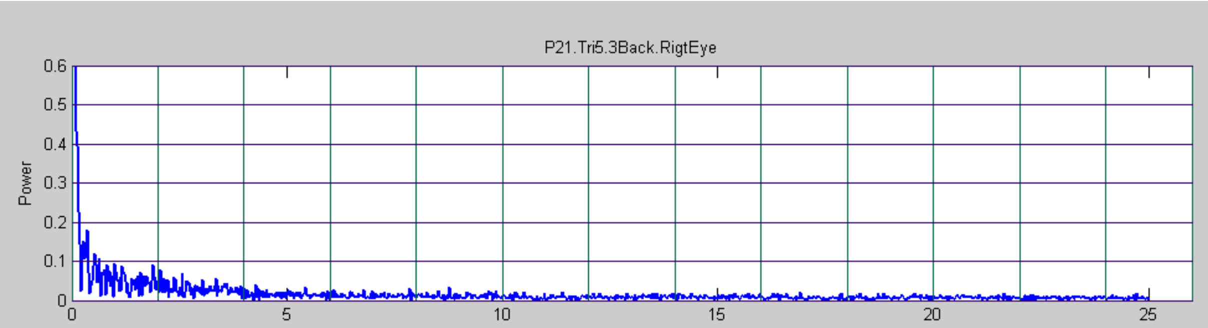

Figure 3.3. A result of FT analysis using the horizontal eye data in Figure 3.2. Power density values were calculated between 0 and 25 Hz. However, the analysis used the range from 1 Hz to 25 Hz. These power density values were compared between experimental conditions.

There is a caveat in Algorithm 2 using FT. That is regarding regular saccades. Algorithm 1 eliminated the unwanted effect from regular saccade eye movements, but Algorithm 2 did not. Algorithm 2 might work better by dividing fixational eye movement (FEM) sections, as Algorithm 1 did, and taking a FT analysis in each of the FEM section. Since this dissertation's goal was to examine if FT analyses of eye movements would contribute to predicting mental workload (MWL), a complicated analysis such as FT analyses with divided FEM sections was not conducted.

Step 3. Processing fft output

Matlab's fft function produces an output file like that shown in Figure 3.3. The algorithm processed the output data set, and quantified them into dependent variables (DVs) of the trial. The eye movement frequencies of interest were between 1.14 and 16.67 Hz as mentioned in the paragraph above. Out of the data obtained from 0 to 25 Hz, the power spectrum data were divided into three temporal frequency bands: the lower band FFT ranging from 0 to 1.14 Hz, which does not relate SIs, the middle band ranging from 1.14 to 16.67 Hz, which is supposed to relate SIs, and the higher band FFT ranging 16.67 to 25 Hz, which does not relate SIs. The average score was calculated for each of the three bands in each trial. The red horizontal lines in Figure 3.4 represent the average values of the power spectrum. These average values were the three dependent variables (DVs) in each trial representing the power of eye movement in the frequency. The DVs were compared later with the independent variables (IVs), such as the N-back task levels in the result section.

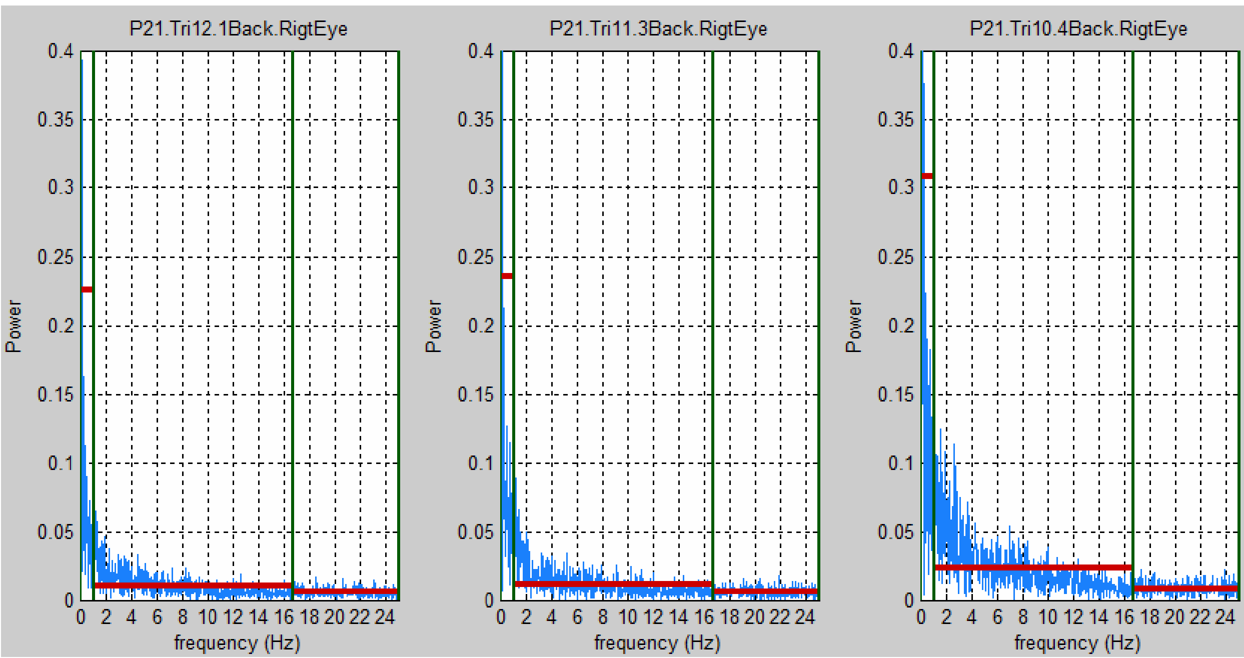

Figure 3.4. Sample FT results in three trials from the three different level MWL conditions. The power densities were statistically compared at each of the three frequency bands.

3.3.3 Conclusion of FT Quantification

Fourier transformation (FT) seems to be a good means of quantifying saccadic intrusions (SIs) because a typical SI makes a round-trip eye movement, which can be converted into wave frequency components; frequency components can be analyzed using FT. The power density measures in the three FT bands were theoretically associated (and dissociated) with some of the

SI characteristics (such as dwell time, round-trip behavior on one plane, and movement amplitude). However, since this dissertation is the first research to use FT to examine SIs, some steps in the analyses might be still crude and need future improvement. This dissertation focused on the primary algorithm (the SI algorithm) and placed the FT algorithm as a secondary measure. Although the FT algorithm (Algorithm 2) in this dissertation may not be as refined as the SI algorithm (Algorithm 1), the FT algorithm was associated enough with the SI characteristics, and the usage of the FT algorithm in this dissertation revealed many aspects in the SI characteristics. See more in the Results chapter.

3.4 Pupil Algorithm

3.4.1 Measurement of Pupil Diameter

Algorithm 3 calculates the DV pupil diameter. The pupil diameter quantification is probably the easiest among the three kinds of DV in this study for two main reasons: first, the pupil diameter behavior is already known and, second, the time-series aspect of the pupil diameter change is not as essential as Algorithms 1 and 2.

First, quantifying a pupil diameter is not difficult because it is known that the pupil dilates as mental workload (MWL) increases (Just & Carpenters, 1993). Therefore, there are already guidelines as to how pupil diameters should be quantified, unlike Algorithms 1 and 2, in which a method to quantify saccadic eye deviation had not been established.

Second, quantifying the pupil diameter is relatively easy because the pupil diameter is directly measurable at any single moment. On the other hand, saccadic eye deviation measured in Algorithms 1 and 2 does not make sense with a single datum at a single moment. Eye movements have to entail a time-series of eye data. Conversely, pupil diameter does not require a sequence of data. The purpose of this study necessitated a means to estimate the average MWL in a certain period of time (usually around 30 seconds) in each trial. Therefore, Algorithm 3 did not require much calculation; a mere average pupil diameter over the evaluation period should be sufficient to represent the pupil diameter for the trial.

3.4.2 Procedures to Measure Pupil Diameter

Tobii recorded the pupil diameter by 0.01 millimeter every 20 ms. There were some missing data, usually because of blinks. Each blink typically takes about 100 to 200 ms (5 to 10 samples of 20 ms events). That means that 5 to 10 consecutive samples could be missed for every blink. Missing data were replaced with the previous value. Figure 3.5 shows both the pupil data before and after cleaning up.

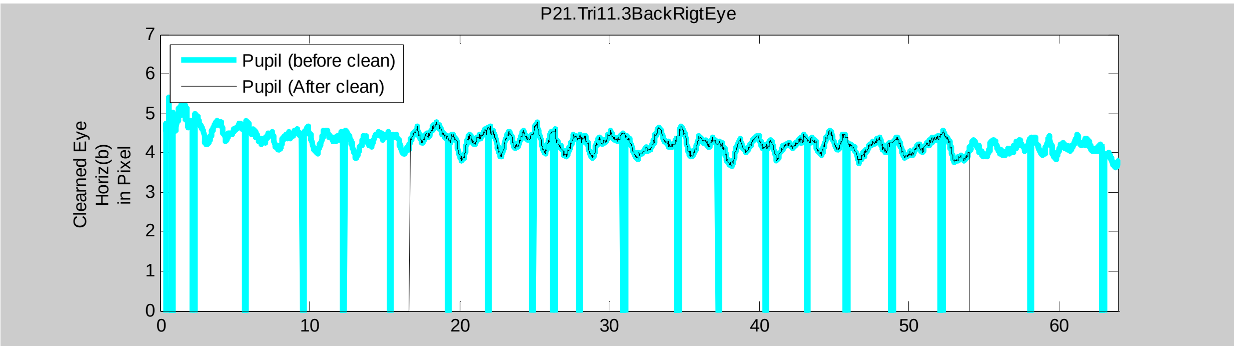

Figure 3.5. Sample pupil diameter data set. The raw pupil data were cleaned by replacing missing data with the previous values. The average pupil diameter during the evaluation period represented the pupil diameter in the trial.

After collecting the data, the average pupil diameter was calculated for the evaluation period. While Algorithms 1 and 2 did the complex calculations to quantify DV to represent the eye deviation, Algorithm 3 quantified the pupil diameter into the DV, by just taking the average pupil diameter size over the evaluation period. The DV was examined in relation to the independent variable, also called mental workload (MWL).

CHAPTER 4

METHODS

4.1 Hypotheses

The experimental hypothesis in this dissertation proposes that saccadic intrusions (SIs) increase as mental workload (MWL) increases. Given the common resource model (reviewed in Chapter 2), higher MWL should lead to fewer mental resources available for controlling fixational eye movements. Less control over fixational eye movements might lead to temporary lost of eye control and therefore an increase in saccadic deviations from fixation – SIs. MWL was manipulated by the N-back task. The independent variable (IV) was the N level. The SIs were quantified by the SI algorithm explained in Chapter 3. The dependent variables (DVs) were the seven physiological behavior variables. One is a SI measure, five are eye movement measures similar to a SI measure, and the last DV was pupil diameter. All of these were calculated based on the Tobii eye tracker data. The main hypothesis stated that the SI measure would increase when the N level increased. The secondary hypothesis stated that pupil diameter would increase when the N level increased.

Three studies were designed to test the hypotheses. In Study 1, the hypotheses were tested in an attentional fixation task, where the eye gaze stayed on a stationary fixation dot. This tested the hypothesis that when an operator's eye fixates at one point all the time, then the eye deviates from the fixation point more often as MWL increases. This study had high experimental control, but was far removed from real life situations. Consequently, Study 2 used a random dot fixation task in which the dot switched between one of 81 possible locations during the course of the trial. Therefore, participants in Study 2 had controlled saccadic eye movements within a certain portion of the display screen. Adding saccadic eye movements to the task improved the study's external validity by making it a step closer to eye movements in the real world. In Study 3, participants freely inspected a naturalistic picture from a driving scene. This study was meant to more closely simulate real world eye movement performance in which the observer could look anywhere within an image at any time. One of the goals of the dissertation was to develop an algorithm that was robust enough to handle the any of the situations from Study 1, 2, or 3.

4.2 Participants

4.2.1 Participants

Forty-two undergraduate students from Wichita State University participated in the study in partial fulfillment of requirements for undergraduate psychology courses. Out of 42, five were discarded and 37 were used for the analyses. The five participants were removed from the analyses due to the following reasons; two of these participants could not perform the N-back tasks on the performance level threshold criteria of 60% accuracy. The eye recordings for two more participants accidentally stopped during the experiment. The cause of the unexpected stop has not been found. Also, while debriefing at the end of the experiment, one of the participants whose eye recording stopped said that she also had known the premise of the study. Another participant's eye tracking calibration was not successful, possibly due to eyelash interference. The analyses in this dissertation used the remaining 37 participants' data, integrated from11, 12, and 14 participants in Study 1, 2, and 3, respectively. The participant identification numbers in the figures in this dissertation were from an arbitrary number (21) because of pilot experiments. Eleven participants (P21 to P31) were for Study 1, 12 participants (P37 to P48) were for Study 2, and 14 participants (P49 to P62) were for Study 3.

4.2.2 Demographic Data of the Participants

The remaining 37 participants were 29 females and 8 males, and their ages ranged 18 to 55 (M = 20.2, SD = 6.2). During the experiment, 14 participants (38%) wore contact lenses, one (3%) wore eye glasses, and 22 (59%) used no correction (See Table A-1 in Appendix). After correction, if needed, all participants had near visual acuity of 20/40 or better. The demographic data for the 37 participants are summarized in Table A.1 in the Appendix.

All 37 participants passed the screening tests described below and were unaware of the hypotheses of the study. Two screening tests, such as a questionnaire and a visual perception screening test, were conducted to screen undesired participants. The questionnaire assessed the participants' current eye conditions based on factors such as recent sleep habits or pre-existing eye conditions (see Table A.2 in Appendix). The visual perception screening test ensured that participants had 20/40 vision or better on the Snellen near acuity test. No participants except for the sixteen and five participants described in the previous section were screened out using these two screening tests.

4.3 Materials

4.3.1 Tobii Eye Tracker

The Tobii 1750 Eye-Tracker was used to collect eye activity data. Tobii is an infrared, non-intrusive real-time eye-tracker, that does not require participants to wear any eye tracking related head gear (a pilot experiment scene is depicted in Figure 4.1). The eye tracking data was sampled at 50Hz. The eye movement data were recorded as gaze landing location on the computer screen on the x axis (the horizontal axis) from 0 to 1280 pixels and the y axis (the vertical axis) from 0 to 1024 pixels on the flat-surface computer screen. The physical dimensions of the display were 340mm wide and 274mm high. Tobii Technology, Inc. reports a reliable spatial resolution of 0.25 deg, which is approximately 10 pixels. The other eye activity measures recorded by Tobii included time in milliseconds, pupil diameter in 0.01 millimeters, distance of the eye from the infrared camera in millimeters, the horizontal eye position in the video (from 0 to 1), the vertical eye position in the video (from 0 to 1), and the validity of the recorded data. All data are recorded for both eyes individually.

Figure 4.1: Experiment scene setup including Tobii monitor, speakers, keyboard, mouse, and chin rest.

4.3.2 Computers

Visual and auditory stimuli for the experiment were programmed and presented using Matlab R2007a and the Psychophysics toolbox 3.0 (Brainard, 1997; Pelli, 1997). Visual stimuli were presented on the Tobii monitor, which is a color LCD monitor 27 cm high and 34 cm wide. The monitor refresh rate was 60Hz (every 16.7ms) and was driven by an ATI Radeon X1300/X1550 Graphic Processor. Auditory stimuli were presented from the computer speakers located next to the Tobii monitor. The computer keyboard was used to collect the participants' responses to the N-back task. The computer mouse was used for the single-scale subjective difficulty rating. All of the above were controlled by a 1.86 GHz Windows PC with 2.0 GB RAM. A chin rest positioned approximately 60 cm away from the Tobii monitor, and approximately 30 cm above the Tobii table was used to minimize head movements beside the eye activities. The NASA-TLX subjective difficulty ratings were collected on another Windows PC in the same room with Tobii (PC2 in Figure 4.2).

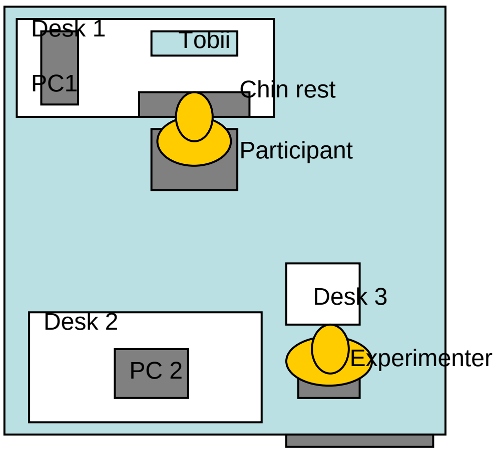

Figure 4.2: Experimental room setting-up. The Tobii screen was approximately 60 cm away from the participants' eyes.

4.4 Tasks

All three studies used a dual-task paradigm; an auditory N-back task imposed mental workload (MWL) on the participants while different visual tasks in each of the three studies served to induce different types of eye movements. Study 1 used the stationary dot fixation task (which is called "fixation task"). Study 2 used the random dot fixation task. Study 3 used the stationary picture free-viewing task. Each of the 37 participants participated in one of the three studies.

4.4.1 Study 1: Fixation Task

Study 1 examined if the pattern of eye behavior changed during a stable-location dot fixation task while participants were engaged in different levels of MWL. A dual-task method was used in the experiment to examine the relationship between eye movements and MWL. A visual task was used to help keep eye movements as steady as possible while auditory N-back tasks were used to systematically vary MWL.

The visual stimulus was a red dot on a black background. The participants were instructed to stare at the red dot and maintain steady fixation . The red dot was four pixels in diameter, or 0.1 deg of visual angle. The position of the red dot was fixed in the center of the computer screen except during the first eight seconds and the last eight seconds in each block. During these two periods, the red dot was shown at one of the four diagonal corners about 96 pixels or 25 mm or 2.3 deg away from the screen center (Figures 4.3 and 4.4). These two eye recalibration periods later served to confirm that the eye data were appropriately collected.

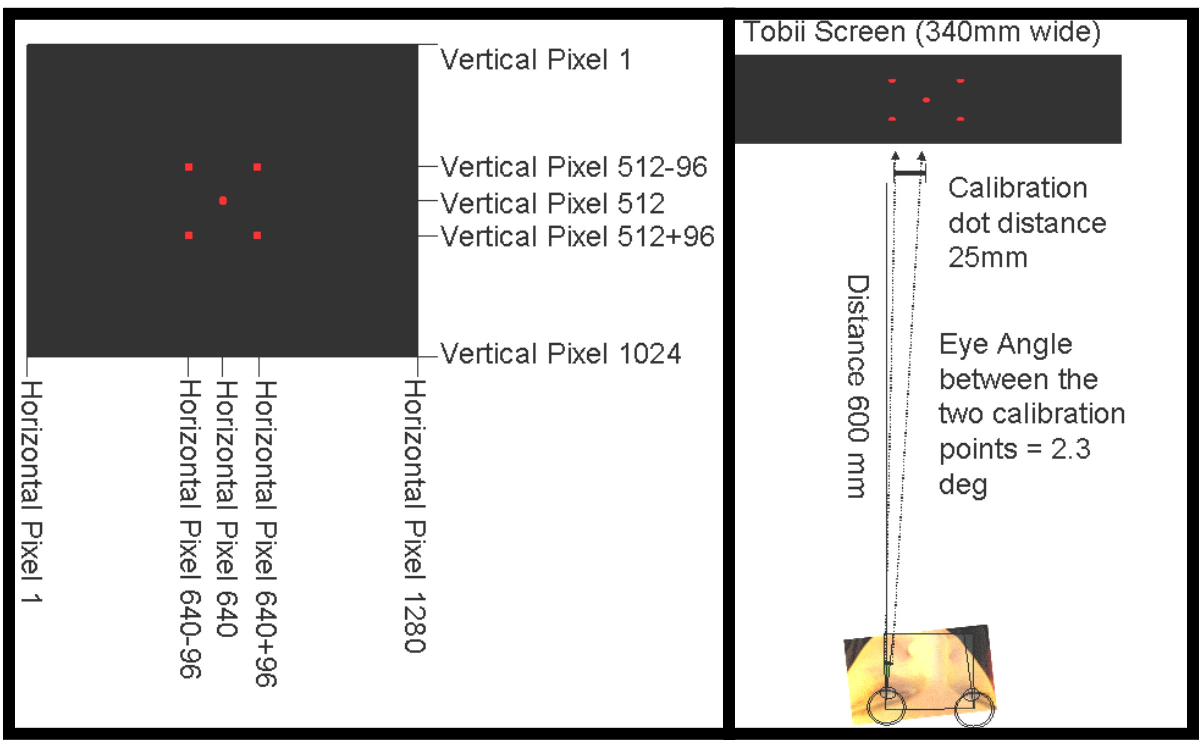

Figure 4.3 (Left) and 4.4 (Right). The participants were instructed to follow the red dot before and after the dual task. The red dot locations were 2.3 deg from the center on the horizontal plane and vertical plane.

Between these two eye calibration periods was a dual-task period originally set to 38.4 seconds. During this time, participants performed both the visual fixation task and the auditory

N-back task. While engaged in the visual task described above, a random number between 1 and 4 was presented every 2.4 seconds. The numbers, pre-recorded by a native English speaker, were played from the computer speakers. The auditory stimuli were set to about 70dB. Each of the 2.4 second periods consisted of 0.5 seconds of voice and 1.9 seconds of silence. Each trial had 15 auditory stimuli and a 2.4 second silence period. The total time for one trial was 54.4 seconds, consisted of (15+1) * 2.4 = 38.4 seconds of the dual task period and two 8 second eye recalibration periods before and after the dual task period. The inter-stimuli-interval (ISI) of 2.4 seconds was tailored for each participant from 2.2 to 4.6 seconds depending on their N-back performance. See the Section 3.8: "Difficulty adjustment" in this chapter. Eleven (P21 to P31) out of 37 participants participated in Study 1.

4.4.2 Study 2: Random Dot Task

Twelve (P37 to P48) out of 37 participants were assigned in Study 2. The method was the same as Study 1, except for the visual stimuli between the two eye calibration periods. Instead of keeping the fixation target at the center of the screen the whole time, the fixation target shifted its location every 2 to 4 seconds. The instructions were the same: to fixate the gaze at the fixation target. Location and timing of the dot position were random. Timing was randomly picked from 11 choices which ranged from 2 to 4 seconds by 0.2 seconds for every dot presentation. A location was randomly selected from 81 alterations which was a 9-by-9 grid. The physical dimensions of the grid were 192 x 192 pixels at the center of the screen. The 81 location candidates are shown in Figure 4.5, which were equally distributed in the 192-by-192-pixel grid, having 24 pixel gaps from the other candidates (see Figure 4.5). Since the locations of the fixation target suddenly shifted every few seconds, rather than moved gradually, this task invoked regular saccades, rather than smooth pursuit. This task would produce pairs of a fixation and a saccade; each pair would consist of a fixation for 2 to 4 seconds and a saccade for about 100 ms. The alternation between a fixation and a saccade would be discrete (qualitatively distinguishable). This task was designed to examine the effectiveness of the SI detection algorithm to extract only the behavior of SIs, removing regular saccades that would occur every 2 to 4 seconds. Since a fixation and a saccade discretely alternate in a slow cycle of 2 to 4 seconds in Study 2, it was easier for the SI detection algorithm to detect fixations and SIs during fixations.

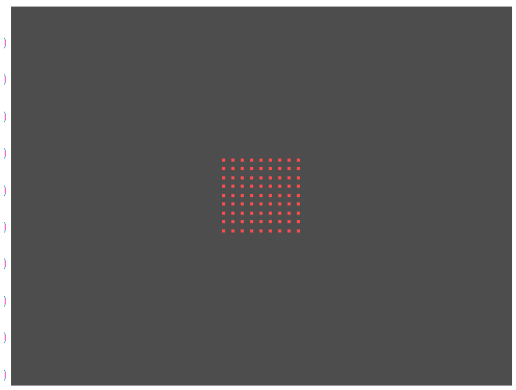

Figure 4.5 (left). Visual stimuli locations in Study 2, Random dot task. Only one dot was presented at a time to invoke a combination of saccades and fixations.

4.4.3 Study 3: Free-viewing Task

Fourteen (P49 to P62) out of 37 participants participated in Study 3. The method was the same as Study 1 and Study 2, except for the visual stimuli between the two eye calibration periods. In a free-viewing task, a stationary picture in a driving scene (Figure 4.6) was displayed until the onset of the second eye calibration period. There were a total of 21 kinds of pictures prepared from the same category, which was a driving scene viewed from a car on a sunny day in Boston, retrieved from the LabelMe project at the MIT computer science and artificial intelligence laboratory (Russell, Torralba, Murphy & Freeman, 2008). There were three criterions to choose visual stimulus pictures in the Boston folder in LabelMe; (1) it was a highway scene, (2) at least one car's rear end was seen, and (3) there were no readable letters in the scene (see Figure 4.7 for an example).

The same three pictures across participants were alternatively and repeatedly used in the practice trials. The other 18 pictures were used randomly but not repeated in the experimental trials. Since some trials would fail for random reasons, the number of pictures (18) needed to be bigger than the number of trials (12). Each of the pictures was gradually cropped in a circle of a size of a 360-pixel diameter (Figure 4.6), which was about 9 degrees in visual angle from the participant's eye, positioned at around 600 mm away from the computer screen. Participants were instructed to freely view the stationary image, but not to look outside of the circular picture.

This task was designed to induce free eye movements with constraints such as no smooth pursuits and no gaze outside of the 9-degree circular picture. Since there was no fixation target, fixations were not as stable as in Study 1 and Study 2. The analyses examined if the SI detection algorithm could extract SI eye data out of many other eye types when fixations were not stable.

Figure 4.6 (left). An example of visual stimuli in Study 3, Free-viewing task, using a driving scenery. There was no fixation target.
Figure 4.7 (right). Magnification image of the circularly cropped picture on the left

4.4.4 N-back Task

N-back tasks quantitatively vary mental workload by requiring that a participant judge whether two designated numbers are the same or different. The two numbers to compare are the most recent number and another number N events before. Participants were instructed to respond to the N-back task by using the right arrow key for the “SAME” answers and the up arrow key for the “DIFFERENT” answers. N-back task performance was evaluated by the percent of correct answers. An experimental trial consisted of 15 auditory stimuli. For example, a 3-back task had 12 answering opportunities since no responses were required for the first N stimuli (i.e. three stimuli). If there were eight correct responses and four incorrect responses (or missed answers), the correct percent would be 8/12 = 66.7%.

4.4.5 N-back Local Rules

In addition to the general rules for the N-back tasks described above and the literature review section, two additional constraints were added. First, the number stimuli were not purely randomized. In a normal N-back task, number stimuli would be chosen randomly. The chance of the “SAME” stimuli would be as low as 25% using the stimuli 1, 2, 3, and 4 since the chance of each number appearance is 25%. Conversely speaking, participants can achieve 75% accuracy by just responding “DIFFERENT” without actually performing the task. In order to avoid this possibility, the chance of the “SAME” stimuli appearance was set to 50% so that participants had no way to cheat on the task.

The second local rule was the auditory feedback. During practice trials participants were also provided auditory feedback indicating whether their response on the N back task was correct or incorrect. Auditory feedback was not provided during the data collection trials.

4.4.6 Subjective Difficulty Rating

A single-scale subjective difficulty rating was collected after the second eye re-calibration period ended on every trial by using a green gauge bar linked to the horizontal mouse movements (Figure 4.8 below). The rating scale was labeled with two words: easy and difficult. The participant rated their subjective feeling of the difficulty level on the dual task by clicking the mouse button once. The rating was coded as an integer between 0 and 100. The screen remained the same until the participant clicked on the mouse, so that there was plenty of time to determine the rating. This usually took less than 5 seconds.

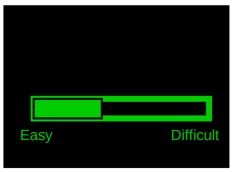

Figure 4.8: A sample image of the Single-scale subjective difficulty rating. Participants were instructed to rate their feeling of difficulty level toward the dual task.

4.4.7 NASA-TLX

The computer-based NASA-TLX subjective rating system (Hart & Staveland, 1988) was used to measure subjective difficulty feelings in addition to the single-scale subjective difficulty rating. While the single-scale subjective difficulty rating was implemented after every N-back task trial, the NASA-TLX was conducted only once for every N after the last trial of the N; once for the 1-back task, once for the 2-back task, and once for the 3-back task for each participant. Participants completed the subscales of Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level. One NASA-TLX rating took usually 90 seconds.

4.4.8 N-back Difficulty Level Adjustment

The goal of these studies was to systematically vary MWL levels for each participant and measure changes in eye activity. Since there are individual differences in remembering numbers, the exact same task for every participant would not systematically vary MWL. During the practice block, task difficulty levels were tailored for each participant by adjusting the N and inter-stimulus interval (ISI), so that all the participants could receive three distinct difficulty levels such as light MWL, medium MWL, and heavy MWL within the person. A flow-chart algorithm was made to determine the appropriate N and ISI (see Figure 4.9).

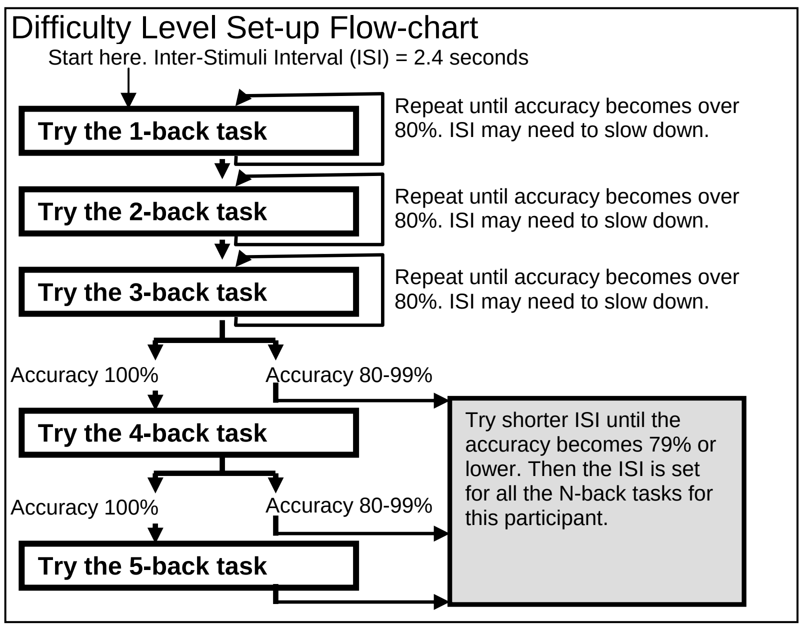

Figure 4.9: Difficulty Level determination flow-chart. Most participants performed up to the 3-back task. Some did the 4-back task.

Practice trials started with a 1-back task and an ISI of 2.4 seconds. The 1-back task was repeated until the participant achieved 80% accuracy or better. If the N-back task performance of the trial was 80% accurate or better, the N increased to 2-back. Likewise, after 80% accuracy for the 2-back task, the N was increased to 3-back. The 3-back task was the typical highest MWL level for the average participants. For those average participants, the ISI was adjusted somewhere between 2.2 and 4.6 seconds to the point that the participant could do the 3-back task between 80% and 99% accuracy. The determined ISI was used for all the N-back tasks for that participant.

There were seven participants (19%) who could do the 3-back task with 100% accuracy. The 4-back task was given to those seven participants. The ISI was adjusted for the 4-back task to the performance level somewhere between 80 and 99%. These participants had three distinctively varied levels of the N-back tasks such as 1, 3, and 4-back tasks.

4.5 Procedure

4.5.1 Overall Procedure

The experiment was carried out in six stages. First, the SONA system was used to recruit participants. Second, participants completed the demographics questionnaire. Third, the visual perception assessment was used to screened participants on the basis of visual ability. Fourth, participants practiced the dual task and mental workload levels were tailored for each participant. Fifth, experimental trials were performed. Sixth, the participants were debriefed about the study. The second and third stages, the questionnaire and the visual perception battery, are detailed above in the Participants and Screening section. This section will elaborate the other stages in the chronological order.

4.5.2 SONA

Participants were recruited using the SONA recruiting system. The title, “Eye movements and brain activities,” and description of the study was posted on SONA. In addition participants were advised that the study involved the recording of eye movements while you engaged in a mentally demanding task”. Participants were also advised to “get at least 6 hours of sleep the night before you arrive”.

4.5.3 Questionnaire

The participants completed out both the consent form and the background survey form (attached as Appendix A). The background survey form asked for age, eye correction (None, Glasses, Contact lens), handedness (left-handed or right-handed), and three questions mentioned in Screening section. The consent form and the survey form took about 3 minutes to fill out.

4.5.4 Visual Acuity

Near acuity was assessed using the Snellen near acuity test. Participants were asked to perform the dual task without their glasses if they had near visual acuity of 20/40 without their habitual correction.

4.5.5 Task Practice

The participants received the instructions for the dual task procedures and practiced them. Each trial consisted of the dual task the visual task and the auditory N-back task, followed by a single-scale subjective difficulty level rating. There was also an original algorithm to determine the most appropriate MWL levels for each participant by changing two factors: the N in the N-back task, and ISI. The practice block was computerized in one Matlab file so that the practice menu was almost self-navigatable (See the practice block figure 4.10 below), the experimenter was seated behind the participant and directed them of the procedures throughout.

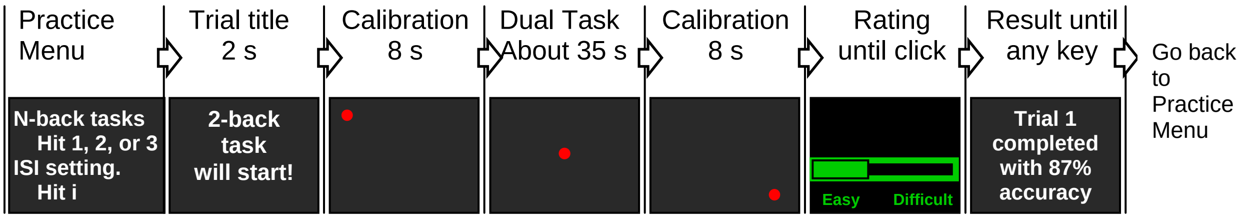

Figure 4.10: Practice block consisted of the practice menu, practice title, calibration, dual task, calibration, subjective rating, and the performance result display.

The practice trials started with the 1-back task, then the 2-back task, and then the 3-back task. Each practice trial was followed by a single-scale subjective difficulty rating, followed by the correct percent result display. If the accuracy percentage was 79% or lower, the practice task was explained again, and the trial was repeated again. If the accuracy percentage was 80% or better, a more difficult task (either a higher N-back task or the same N-back task with faster ISI) was given to the participant in the next practice trial.

All of the practice block activities were done without eye recording although the participants were practicing the dot fixation task during the practice block. The NASA-TLX instrument was not administered during practice. After the practice trials and tailoring of the N-back task difficulty levels (see the detail in the previous section, the Section 3 "Tasks" in this chapter), the practice block was terminated. Participants could spend 5 to 30 minutes for the practice and difficulty adjustment trials depending on how fast the N-back tasks were mastered. To avoid eye strain, the practice duration was limited to 30 minutes maximum, so that the total time could be limited to 50 minutes after adding another 20 minutes of the experiment trial stage.

4.5.6 Eye Recording

After participants practiced the N-back tasks, Tobii recording calibration was conducted. Participants sat approximately 85 cm away from the Tobii monitor, leaned forward, and put the chin on the chin rest, which was approximately 60 cm away from the Tobii monitor. This amount 60 cm was determined because the Tobii eye tracker is supposed to work the best with 60 cm eye distance from the Tobii camera. The nine point eye calibration tailored the eye gaze recording to each participant. The eye calibration was repeated until Tobii said there were no more miss-calibration points. Also the eye calibration was verified using the Tobii verification tool. The calibration took about 1 to 5 minutes. If the calibration did not succeed, the participant was dismissed from the study!

After calibration, the experimenter started the Tobii recording and manually started the Matlab experimental stimuli presentation program. One Tobii recording included the whole trials by one participant, which took about 20 minutes.

4.5.7 Experimental Trials

While Tobii was recording the eye activity, data for 12 trials (three levels of N-back tasks four times each) were collected for each participant unless a total experiment time 60 minutes passed before the 12 trials. Each trial was exactly like a practice trial; the dual task followed by a subjective difficulty rating task. The only difference from the practice trials was that there was no sound feedback to each N-back key response. Also trials with 60% accuracy or less were considered as bad trials and discarded. Additional trials were run to replace the bad trials. This happened usually once at most for each participant since the difficulty levels were already tailored to the individual. However, there were five participants who needed to do a replacing trial twice or three-times.

The 12 trials were grouped into four blocks. The first block was ascending from the light MWL, to medium MWL, to heavy MWL N-back task (either 1,2,3 back tasks or 1,3,4 back tasks depending on the person’s practice performance). After the first block, there was a 20-second break. If everything was fine, the second block descending from heavy, to medium, to light MWL N-back tasks was conducted.

After the second block (after 6 trials), participants were encouraged to close their eyes for 30 seconds to rest. If something was too uncomfortable, the experiment would discontinue. However, none of the participants claimed it to be uncomfortable and discontinued. The third and fourth blocks were the repetition of the first and second blocks except for the NASA-TLX in the fourth block.

The NASA-TLX was conducted in the fourth block (i.e., the 10, 11, and 12^{th} trials). After each of the trials in the fourth block, the computer-based NASA-TLX subjective difficulty rating was conducted. The participants were instructed to rate the overall difficulty level for the specific N-back task (the 3-back for Trial 10, the 2-back for Trial 11, and the 1-back for Trial 12) as a whole, not for the immediate trial only.

4.5.8 Debrief

After 12 trials and three NASA-TLX ratings, the Tobii recording session was ended by the experimenter. The experimenter thanked the participant for the participation and started debriefing. First, the experimenter asked if the participant could guess what the purpose was. Second, the experimenter asked if the participant noticed that he/she had specific eye movements during the trials. If yes, the participant was asked what eye movements she noticed she had. After these two questions, the experimenter revealed to the participant that the study was trying to find specific eye movement patterns that might consistently occur when the person is thinking deeply; if the pattern is consistent across time and across people, the pattern might become a good predictor to indicate heavy brain usage. After the debrief, the participants were dismissed.

CHAPTER 5

RESULTS

5.1 Introduction to the Results Chapter

Three experiments in this dissertation examined if mental workload (MWL) manipulation by the N-back task would produce consistent physiological changes in the pupil diameter and in saccadic intrusions (SIs). All three studies used the dual-task paradigm; the auditory N-back tasks imposed MWL on the participants, while different visual tasks in each of the studies served to induce different types of eye movements. Study 1 used a stationary dot fixation task (which is called “fixation task” in this dissertation). Study 2 used a random dot fixation task. Study 3 used a stationary picture free-viewing task. See the Methods chapter in this dissertation for the details of the experiments.

Data were collected from at least eleven participants for each of the three studies. Typically, each participant performed four trials on each of the three MWL level conditions (for a total of 12 trials). Each trial lasted approximately 60 seconds. The MWL levels (or the N levels) served as the independent variables (IVs). The dependent variables (DVs) were the physiological responses, such as a SI measure and a pupil measure, quantified into a single value for each trial by original algorithms (for details, see the Algorithm chapter). This Results chapter examines whether the IVs statistically influenced the DVs, such as the SI measure and the pupil measure.

5.2 Descriptive Statistics

5.2.1 Participants

The analyses used the 37 participants' data, integrated from11, 12, and 14 participants in Study 1, 2, and 3, respectively. The participant identification numbers in the figures in this chapter were from an arbitrary number (21) because of previous and the pilot experiments. Eleven participants (P21 to P31) were for Study 1, 12 participants (P37 to P48) were for Study 2, and 14 participants (P49 to P62) were for Study 3.

5.2.2 Missing Data Ratio

Since the Tobii eye tracker is a video-based eye tracker, eyelids, eyelashes, and and hair sometimes block the Tobii's observation of the pupil and the cornea. Therefore, eye activities are not always recorded. When Tobii does not record, the data is marked as invalid in the output file. Figure 5.1 shows the invalid data ratio of 436 trials in 37 participants. Each of the 37 panels represents a participant's results. Within each panel, there are up to 12 vertical bars, representing 12 trials' invalid data ratios in percentages. The three horizontal lines represent the average invalid data ratio within each MWL conditions (low, medium, or heavy MWL) within each participant. Overall, invalid data ratios ranged from 1% to 89% (Figure 5.1) in each trial. The trial with 89% invalid data (Subj.#59, 12th trial) was the only one trial that was removed from analyses due to a high invalid ratio. The second highest invalid data ratio was 57% (Subj.#30, first trial of the 3-back task), followed by the other trials with similar invalid data ratios by the same participant (Subj.30). Despite the high invalid data ratios, these trials were used in the analyses. .

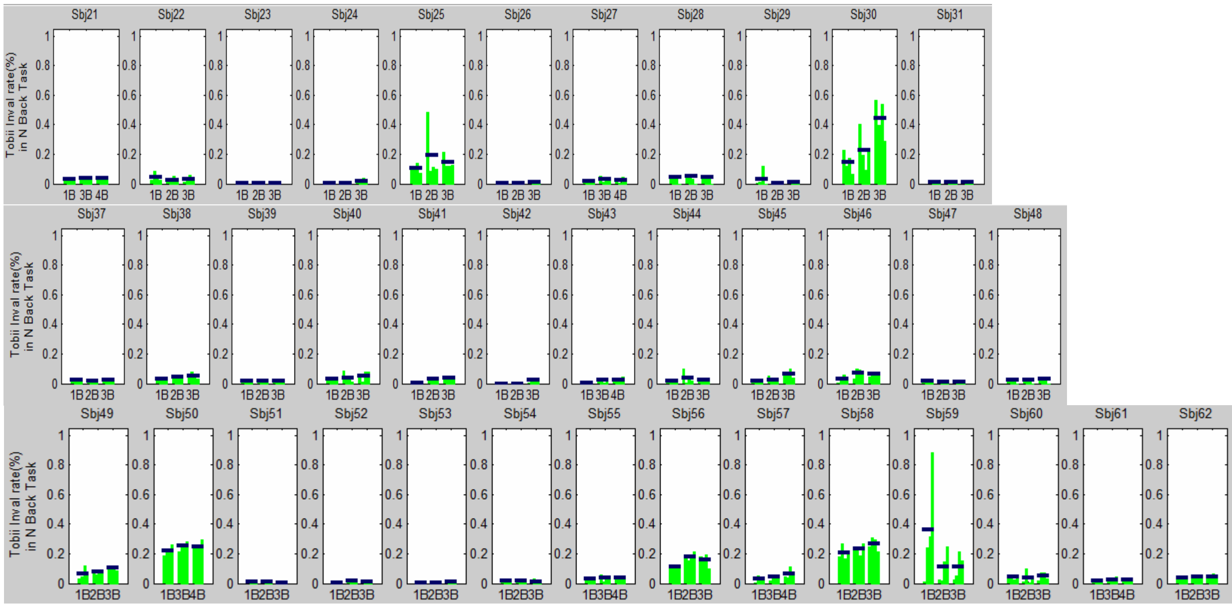

Figure 5.1. In each of the 37 panels, up to 12 vertical bars represent each trial's invalid data ratios. Three horizontal bars represent the average scores in the same N-back condition trials. Generally low invalid ratios show that the eye activity data were recorded almost all the time in the majority of the 37 participants. This indicates that there were enough samples of eye activity data that could be analyzed.

The invalid ratio of 57% (Subj.#30, first trial of the 3-back task in Figure 5.1) was not the optimal amount for the analyses, however, the data replacement method (see the Algorithm chapter) seemed to work well. Figure 5.2 shows the results of an "Invalid Run 200 ms" test. The test was developed by the author. This test first counted the "run" or a sequence length of invalid data that lasted 200 ms or longer. If the eye tracker could record eye data once in 200 ms or more often, the algorithm could theoretically detect a typical SI (see the Algorithm chapter). The "invalid run 200 ms" test could estimate the slip of this scope by calculating the ratio of the evaluation period that was occupied with invalid runs of 200 ms or longer. This test (Figure 5.2) always returns a smaller number than the previous "invalid data ratio" test (Figure 5.1) by removing all the invalid data that lasted their consecutive sequences of less than 200 ms. Both tests can evaluate missing data ratios.

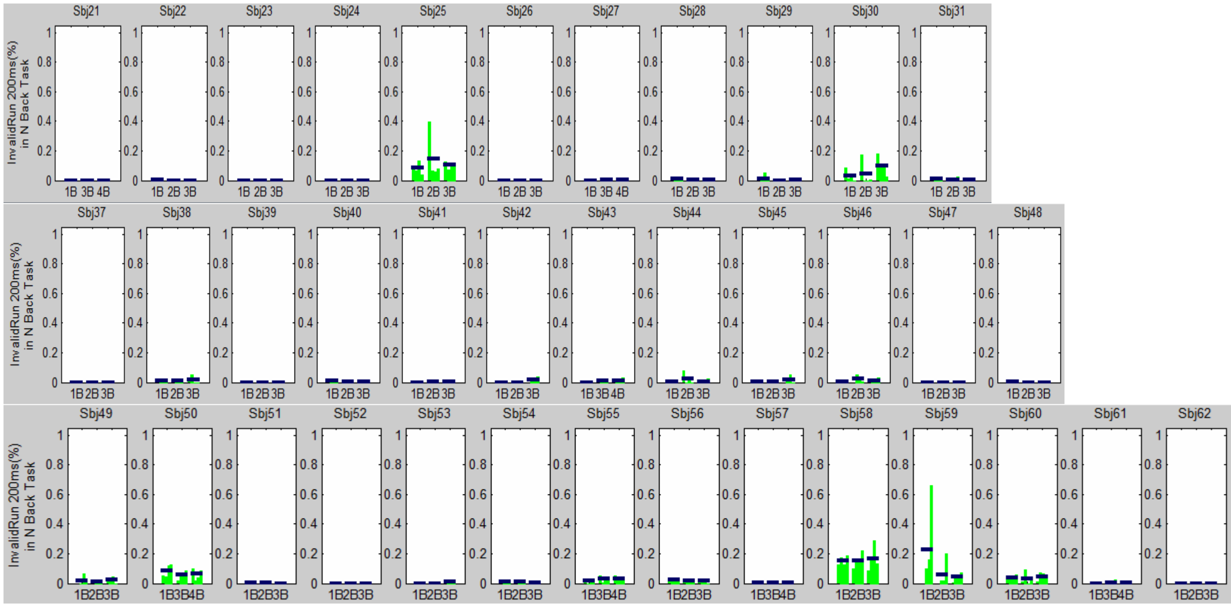

Figure 5.2. Results of the “invalid run 200 ms” test. This test shows the percentage of time that was missing valid data for 200 ms or longer in the trial. If the invalid run is longer than 200 ms, the algorithm may not be able to detect a typical SI, the interest of this dissertation. Except one trial in Subj.#59, all trials had invalid data ratios lower than 50%. This trial was removed.

In Figure 5.2, most trials show shrinkage of their missing data ratios from Figure 5.1. The two trials that still had a large missing data ratio in Figure 5.2 were one for Subj.#25 and one for Subj.#59. The latter one, the trial for Subj.#59 (4^{th} trial of the 1-back task), was removed because both of the missing data ratio tests (Figures 5.1 and 5.2) showed that the invalid data ratio was higher (66%) than the majority line (50%). The trial in Subj.#25 (1^{st} trial of the 2-back task) was included in the analyses because the result of the “invalid run 200 ms” test was 40%. That means there were 60% of time that eye data were recorded at least once in 200 ms in this trial, which is the majority.

Overall, the valid ratios show that the Tobii eye tracker recorded the eye activities most of the time, and the data contained enough valid data to analyze. Only one trial in one participant (Subj. 59) was removed. The other 435 trials’ data in 37 participants were used in the analyses.

5.2.3 Parameters of Participants

Each of the 37 participants practiced the N-back tasks before the experimenter started recording their eye activities. Participants practiced 7 to 18 N-back trials, which took up to 35 minutes. During the practice trials, the N levels and ISIs (Inter-Stimuli Interval) were personalized for each participant, based on the participants' own memory capacity. The baseline N-back sets consisted of 1, 2, and 3-back tasks. MWL manipulation seemed to be adequate with the baseline sets for 30 participants. Seven other participants performed extremely well on the N back task, and the experimenter needed to make the task more difficult for those seven participants. They received the 1, 3, and 4-back tasks for their three distinct levels of MWL. See details in the Methods chapter.

During the practice trials, the ISIs between auditory stimuli were also personalized for each participant. The ISIs for the 37 participants ranged from 2400 ms to 4000 ms. Most of the 37 participants completed 12 experimental trials (four trials for each of the three MWL level conditions), while others completed at least nine experimental trials (three trials for each of the three levels of MWL). These participants had slower ISIs (longer than 3500 ms), therefore, the experiment took longer, and the desired 12 trials were not completed. However, the nine trials were included in the analyses.

5.3 Confirmation of MWL Manipulation

The N-back tasks themselves were designed to impose distinct levels of MWL by varying the N levels; the higher the N was, the more MWL was imposed (see detail in "Analytical methods" in Chapter 2 and "Operationalization" in Chapter 3). This theoretical assumption was confirmed by using four sets of data: N-back performance scores (Figure 5.3 and Table 5.1), N-back reaction times (Figure 5.4 and Table 5.2), uni-scale subjective rating scores (Figure 5.5 and Table 5.3), and the NASA-TLX subjective rating scores (Figure 5.6 with no table).

5.3.1 Correct Percent in the N-back Tasks

In each trial, there were 11 to 14 chances to respond to the N-back auditory stimuli. For each of these responses, the response was judged correct or incorrect. After one trial, the percentage correct was calculated. Figure 5.3 shows the percentage correct in each trial for each of the 37 participants. Each green vertical bar shows the correct percent in a trial. Black horizontal lines represent the average score (%) of four trials in each N level condition for each participant.

Figure 5.3 shows two things. First, the N-back tasks were well understood by the participants. Without understanding the tasks, one cannot keep performing the N-back tasks with 60 % accuracy or better. Second, a majority of the participants showed decreasing trends of the correct percentages as the N level increased. This shows that the performance was worse for the higher N levels (such as 3 or 4-back tasks). This indicates that the higher N-back tasks were more difficult, and imposed more MWL.

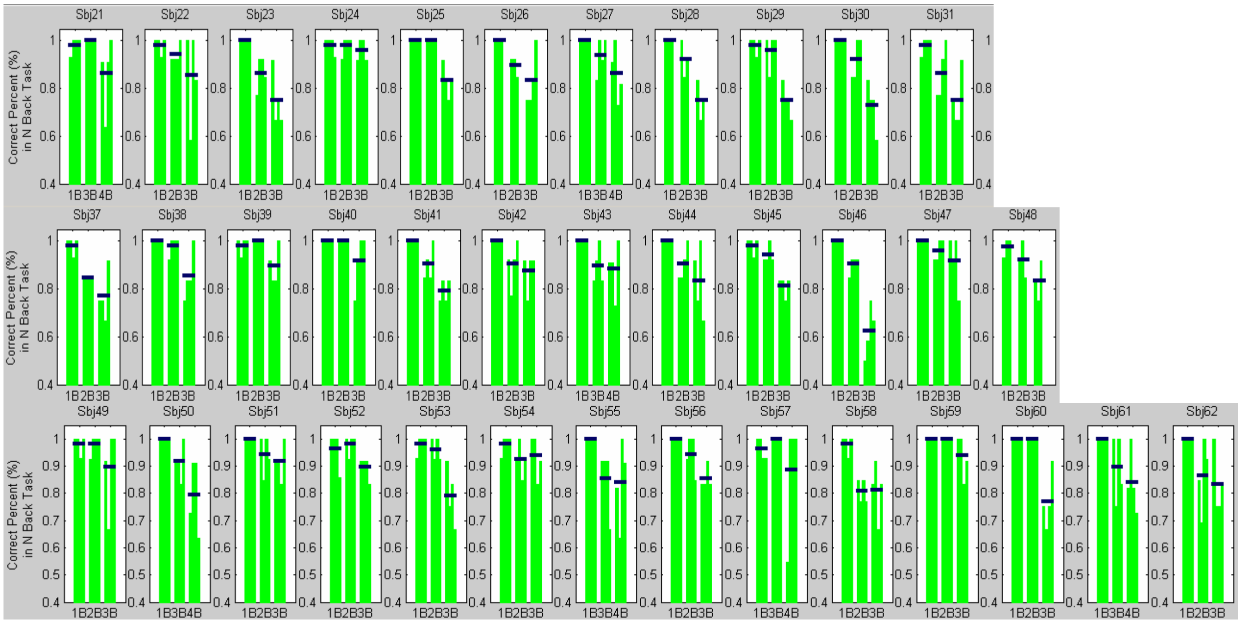

Figure 5.3. Correct percentages of the N-back tasks for 37 participants show that higher N-back tasks were more difficult. This validates the MWL manipulation by the N-back tasks. Eleven participants labeled Subj21 to 31 were for Study 1. Twelve participants from Subj37 to 48 were for Study 2. Fourteen participants from Subj49 to 62 were for Study 3.

Table 5.1 below shows the summary of the relation between the manipulated MWL levels and the correct percentages. A linear regression analysis was conducted for each of the three studies. The predicted variable was the N-back levels. There were two predictors. One was the correct percentages. The other was a categorical variable of the participants' identification numbers.

The goal of this section (*Section 3: Confirmation of MWL manipulation*) is to show that the MWL levels were manipulated well, by showing the relationships of MWL levels with several measures (such as N-back performance, reaction time, and subjective ratings). There might be a better statistics method (such as binomial or non-linear regression) to show the relation for each of the pair than a linear regression. However, the bigger goal of this chapter (the Results Chapter) is to compare predictabilities of the N-levels using nine different measures (such as N-back performance, reaction time, subjective ratings, and seven physiological measures). It is convenient to use the same method (the linear regression) so that the comparison (later shown in Table 5.10) across the nine measures (three confirmation measures and seven physiological measures) is easier.

The R^2 values were 0.47, 0.50, and 0.40 for the three studies respectively, to predict the N-levels using the correct percent and the categorical participants' identification number. This indicates that the higher N-back tasks were more difficult; therefore, the higher N-back tasks imposed more MWL. In other words, these results suggest that low, medium and high MWL levels were appropriately manipulated for each participant.

Table 5.1

SUMMARY OF THE N-BACK PERFORMANCE

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.47**	**0.50**	**0.40**
Adjusted R^2	0.37	0.40	0.28
Standard error	0.76	0.68	0.85
F-statistics	4.46	5.07	3.45
p-value	0.001	0.001	0.001

Summary of linear regression analyses predicting the N levels using the N-back performance (correct response in percent) and the participant identification numbers as a categorical variable.

5.3.2 Reaction Time in the N-back Tasks

Likewise, reaction time (Figure 5.4) confirmed the effects of MWL manipulation. Each vertical bar in Figure 5.4 represents the median reaction time in seconds to the 15 N-back stimuli in each trial. Horizontal lines represent the average score for four trials in each N level condition for each participant. The reaction times were slower for the higher N- back tasks because higher N-back tasks require more mental resources.

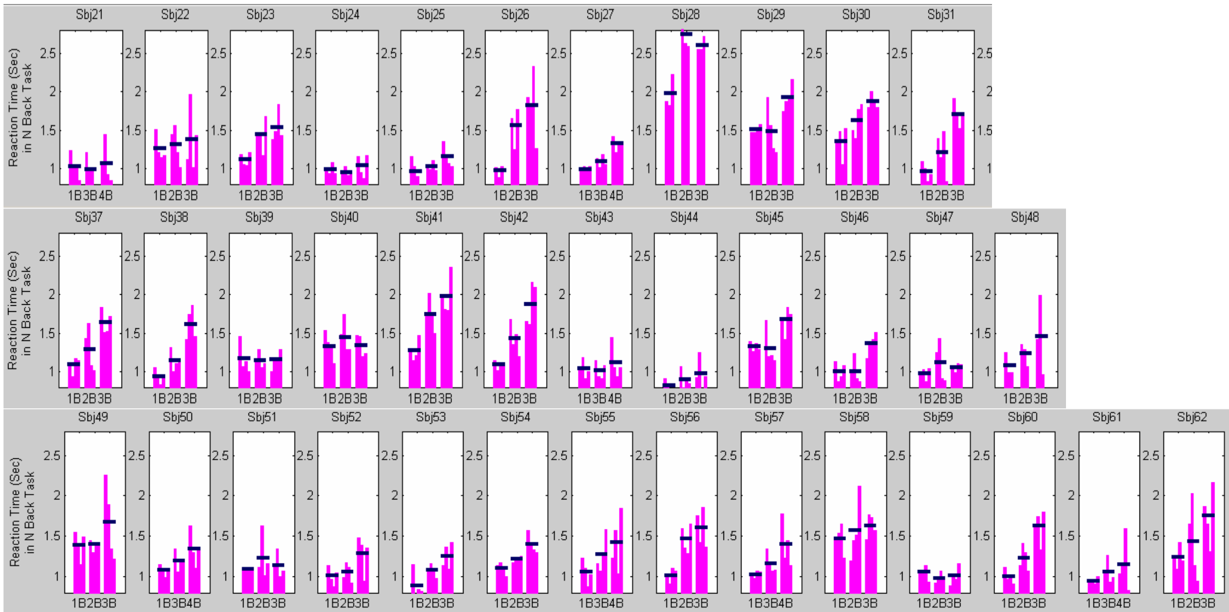

Figure 5.4. Higher N-back tasks yielded longer response times (the y-axis), indicating participants needed more time to think. This validates the MWL manipulation by the varied levels of the N-back tasks.

The R^2 values were 0.44, 0.36, and 0.41 for the three studies respectively, to predict the N-levels using the reaction time and the participants' identification number as a categorical variable (Table 5.2). This indicates that the higher N-back tasks required more time, consistent with the notion that higher N-back tasks imposed more MWL. This also confirms that MWL levels were manipulated well from the low, to medium, and up to heavy levels for each participant.

TABLE 5.2

SUMMARY OF N-BACK REACTION TIME

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.44**	**0.36**	**0.41**
Adjusted R^2	0.33	0.23	0.30
Standard error	0.78	0.78	0.85
F-statistics	3.90	2.80	3.60
p-value	0.001	0.001	0.001

Summary of the linear regression analysis predicting the N levels using the N-back reaction time (in seconds) and the participant identification numbers as a categorical variable.

5.3.3 Uni-scale Subjective Ratings

The manipulation of MWL was also validated via uni-scale subjective ratings. If MWL increased, the participant should report the task was more difficult, and the subjective rating by the participant should increase. The dual task difficulty level was rated after each trial by the participant on one scale with two labels: “Easy” on the left side and “Difficult” on the right side. The rating was quantified into an integer from 0 to 100. Each of the vertical bars in Figure 5.5 represents a rating. The horizontal bar represents the average score for each of the N levels. The vast majority of the participants rated higher N-back tasks more difficult.

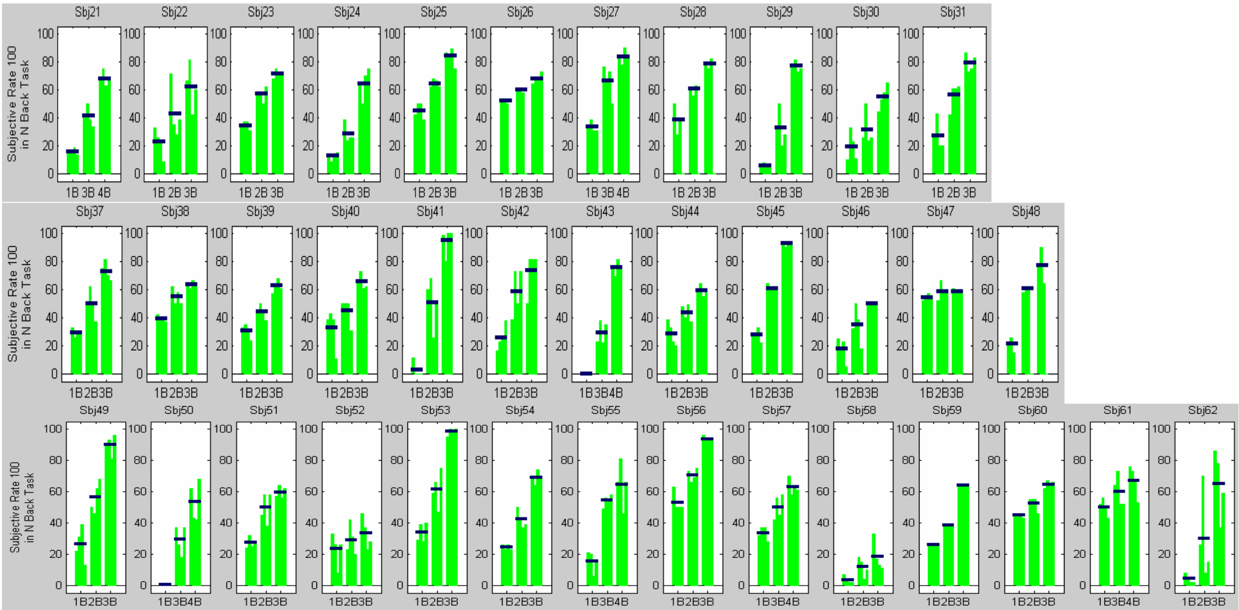

Figure 5.5. Need to first describe the figure. Subjective rating confirmed that the three levels of the N-back tasks imposed three levels of the mental workload (MWL).

The R^2 values for the three studies were 0.88, 0.81, and 0.78 respectively. These results show that the N-level was accurately predicted by the uni-scale subjective rating and the participants' identification number as a categorical variable (Table 5.3). This indicates that the participants felt that the higher N-back tasks were more difficult, therefore, the higher N-back tasks imposed more MWL. This confirms that MWL levels were manipulated in three levels from light, medium, to high MWL for each participant.

TABLE 5.3

SUMMARY OF UNI-SCALE SUBJECTIVE RATING

	Study 1 Fixation Task	Study 2 Random Dot Task	Study 3 Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.88**	**0.81**	**0.78**
Adjusted R^2	0.85	0.77	0.74
Standard error	0.36	0.42	0.51
F-statistics	35.79	21.41	18.82
p-value	0.001	0.001	0.001

Summary of the linear regression analysis predicting the N levels using the uni-scale subjective rating and the categorical participant identification numbers.

5.3.4 NASA-TLX Subjective Ratings

The NASA-TLX subjective rating was conducted only once for each N level for each participant. Therefore, one participant conducted the NASA-TLX three times in the 12 trials. The NASA-TLX produces seven sub-category workload values and one overall workload value. The overall values were used in this dissertation. The results (Figure 5.6) were similar to the uni-scale rating (Figure 5.5); the majority of the participants had increasing values as a function of increasing N level in the N back task. This validates that the MWL levels were manipulated well in the experiments by using the three levels of the N-back tasks.

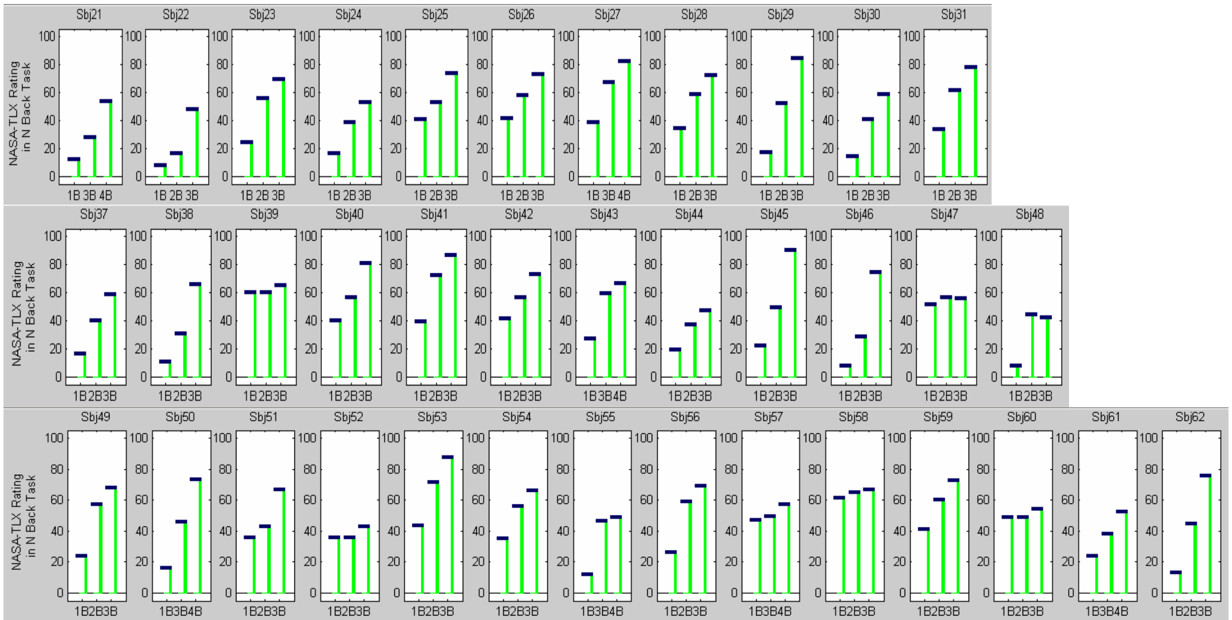

Figure 5.6. Results of the NASA-TLX subjective ratings: Higher N-back tasks were rated as more difficult. This means that the MWL levels were controlled well in the experiments.

5.3.5 Section Summary

This section showed that the four measurements confirmed the manipulation of MWL. The four measurements were the percentage of correct answers to the N-back tasks, reaction time to the N-back tasks, uni-scale subjective ratings, and the NASA-TLX subjective ratings. All of these showed that higher N-back tasks were more difficult. Therefore, if the behavioral responses showed consistent changes as the N-back levels changed, the link between the behavioral change and the MWL change would be strongly suspected. The next section examined behavioral changes, such as pupil diameter and saccadic eye deviations.

5.4 Predicting MWL Using Single Predictor

This section examines relationships between MWL and each of the seven predictors. The seven predictor variables are (1) pupil diameter, three kinds of saccadic eye deviation measure (SED measure), such as (2) saccadic intrusion (SI) measure, (3) microsaccade measure, and (4) gaze aversion measure, and power density of Fourier transformation measures (FT measure) in the (5) lower frequency band, (6) medium frequency band, and (7) higher band. The purpose of this section is to examine the overall effect of MWL on each of the predictors. Usually, correlation analyses would reveal simple relationships between two sets of variables (such as MWL and one predictor), but simple correlation analyses are insufficient because they cannot take individual differences into account. Instead, this dissertation used linear regression analyses, with the N levels as the predicted variables, and having two kinds of predictors: one behavioral variable and one categorical variable of the participant identifications.

When examining these seven predictors, they were evaluated with three criteria: (1) predictability, (2) sign consistency by percentage of participants, and (3) sign consistency by the distribution curve. First, this dissertation has the future goal of estimating a driver's MWL automatically. In doing so, the measurement needs to have high accuracy to estimate the MWL. This is called predictability, which was evaluated by an R^2 value in a linear regression. If an R^2 value is 1.00, then the physiological measure has perfect predictability of MWL. This criterion can evaluate within-subject consistency. An R^2 value can reach 1.00 even with enormous individual differences, and even if two people have two opposite physiological reactions. This leads to the second and third criterions, which examined between-subject consistency.

The second criterion was sign consistency. If a machine and an algorithm automatically estimate MWL using a physiological measure, the relationship between MWL and the physiological measure needs to be consistent across people. This is different from consistency within a person, which is part of the first predictability criterion. If a specific condition makes person A's heart rate consistently faster and person B's heart rate consistently slower, there is inconsistency between the people on the heart rate measure. Inconsistent measurements like this are not desired. The consistency was evaluated by the percentage of participants who had the same sign (i.e. plus or minus) of correlation coefficients. If the percentage of the same sign is 100%, the physiological measure has a consistent relationship, and therefore has a great chance that the measure will consistently show the same pattern in the rest of the population. This dissertation specifically defined that a good MWL measure required having a consistent sign correlation (consistently positive correlation or consistently negative correlation) in 84% of participants. This particular percentage, 84 %, comes from a normal distribution curve. The percentage of the normal distribution above -1 standard deviation is 84.1 %. This dissertation defined 84.1 % as the majority required for consistency in the results in order to reliably generalize the results to the rest of the population. In this dissertation, the number of participants was 37 people. The criterion line, which is 84 % of 37 participants, was 31.1 persons. Given this, a good MWL measure should have consistently positive correlations (or consistently negative correlations) in at least 32 of 37 participants in this dissertation.

The third criterion to evaluate physiological MWL measures was sign consistency in a distribution. This is almost the same as the second criterion. However, the third criterion is evaluated not based on the number of participants, but based on a fitted normal curve distribution. The same threshold line as the second criterion (such as -1 and +1 standard deviation) was used. A good MWL measure should have consistent correlations in individuals of a sample population; when their correlation r values are plotted on a histogram and are fitted to a normal distribution curve, the majority (84.1 % or greater) should be on one side of the zero value (on the positive side or the negative side).

The following sections show all the three criteria (R^2 value and two kinds of sign consistency criterions) for each of the seven physiological measures. The first one, R^2 value, is for evaluating the physiological measure's predictability of MWL. This analysis was conducted with linear regression models. Both of the second and the third criterions together evaluate both reliability and generalizability to the rest of the population. These two criterions were examined in simple correlation analyses.

5.4.1 Pupil Diameter and MWL

First, pupil diameter was analyzed in its relation to MWL. Figure 5.7 shows the changes of the pupil diameter during each of three trials using the three MWL level conditions (1, 3, and 4-back tasks) in one participant (P27). The x-axis represents the time in seconds. The y-axis represents the pupil diameter in millimeters. These results show the pupil diameter differences across the MWL conditions. Pupil diameter was smaller in the 1-back task (light MWL condition) shown in Figure 5.7 (left) compared to the other conditions.

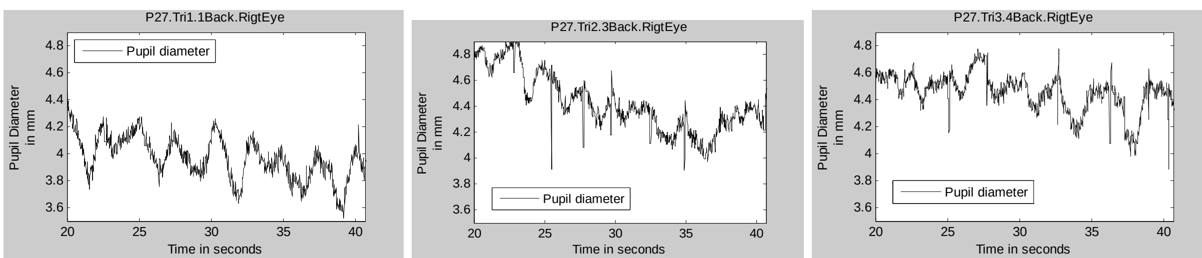

Figure 5.7: Pupil diameter increased as MWL increased. From left to right, the 1, 3, and 4-back tasks respectively served as light, medium and heavy MWL conditions.

Each trial produced an average pupil diameter within the evaluation period which was about 30 seconds. Average pupil diameters during the evaluation period were used as the pupil diameter measure. The 12 trial data were plotted (Figure 5.8) in the relation to the N-levels, which were 1, 3, and 4 back tasks for this example participant (P21). This graph shows a relatively strong, positive relationship between the N levels and the pupil diameter ($r = 0.517$, $p = 0.085$, not significant). In other words, for this participant, the pupil diameter increased with higher N levels.

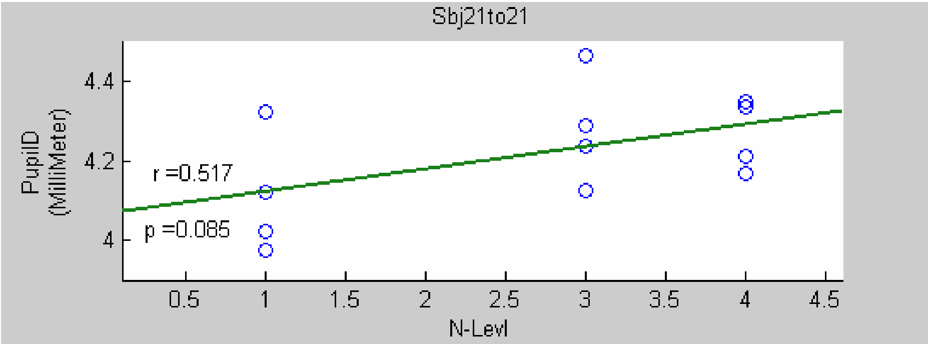

Figure 5.8. One participant's 12 trial results of pupil diameter and the 3 mental workload (MWL) levels. This shows there was a positive correlation between N-levels 1, 3, and 4 (representing MWL levels) and the pupil diameter.

Likewise, the scatter plots below were made by relating the N levels and pupil diameter for each of the 37 participants (Figure 5.9). On the participant level, most participants had strong positive correlations. The average r was 0.57 across the 37 participants. Out of 37 participants, 36 participants (97 %) had a correlation that has a positive direction. Only one (P59 in Figure 5.9) had a negative correlation. This value of 97 % satisfies the second criterion to evaluate the pupil measure as a MWL measure. This indicates that for most participants, pupil diameter increased when the task was difficult.

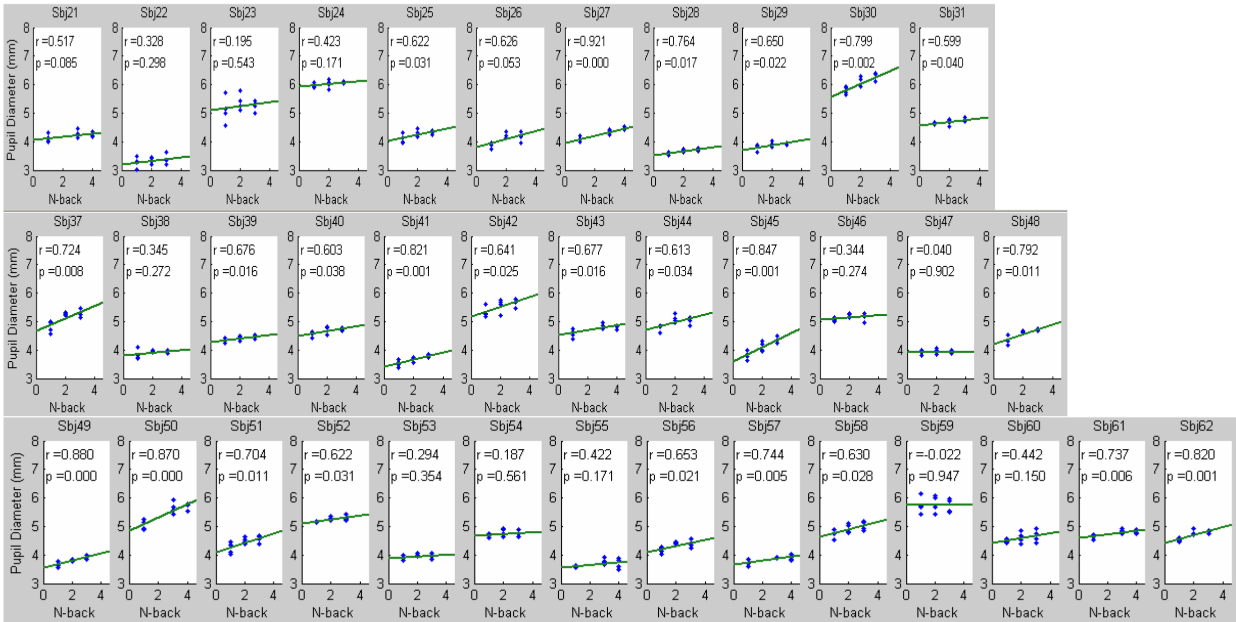

Figure 5.9. The correlation between the N levels in the N-back tasks and pupil diameter for almost all participants in these 12 trials was positive and relatively strong. The average r across 37 participants was 0.57. In these 37 participants, 97% had a positive correlation.

Figure 5.10 below shows a histogram of the correlation coefficients of the 37 participants. A normal distribution curve was fitted to the histogram. The five short, vertical lines represent standard deviations of -2, -1, 0, 1, and 1. The vast majority of the coefficients are on the positive side. This graph shows that pupil diameter satisfied the third criterion: sign consistency in the fitted normal curve distribution. The 84% line (at -1 standard deviation) was at 0.34, which is above the correlation coefficient value of zero. It is reasonable to conclude that pupil diameter produces a reliable estimation of MWL across individuals.

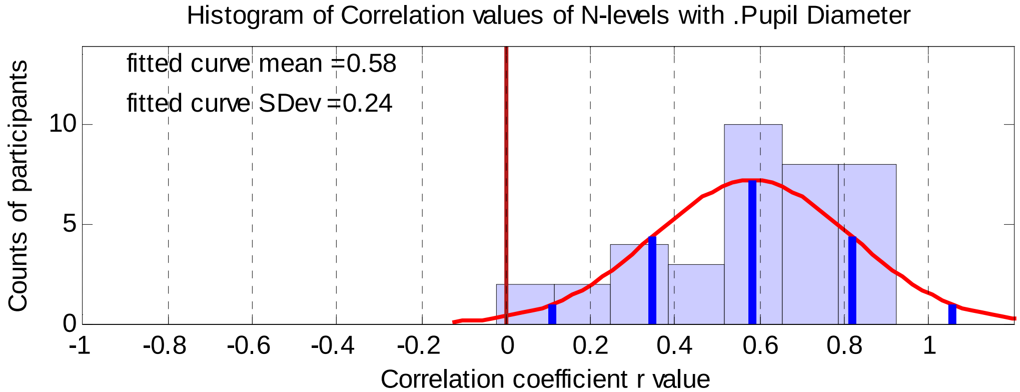

Figure 5.10. A histogram of the correlation coefficient r values between pupil diameter and N-levels. The average correlation value r was 0.58. The majority of participants had a consistent correlation. These consistent correlation values suggest that they can be generalized to the population.

One way to summarize these results for all participants in each study is to use a simple collapsed bar graph. First, data were summarized for each participant taking the average across trials. The participants' averages were then combined into a grand mean for each level of MWL in each study (Figure 5.11). These show that overall, there seemed to be consistent effect. Pupil diameter was always larger for higher MWL tasks. Especially the gap between low and medium MWL levels seemed to be large.

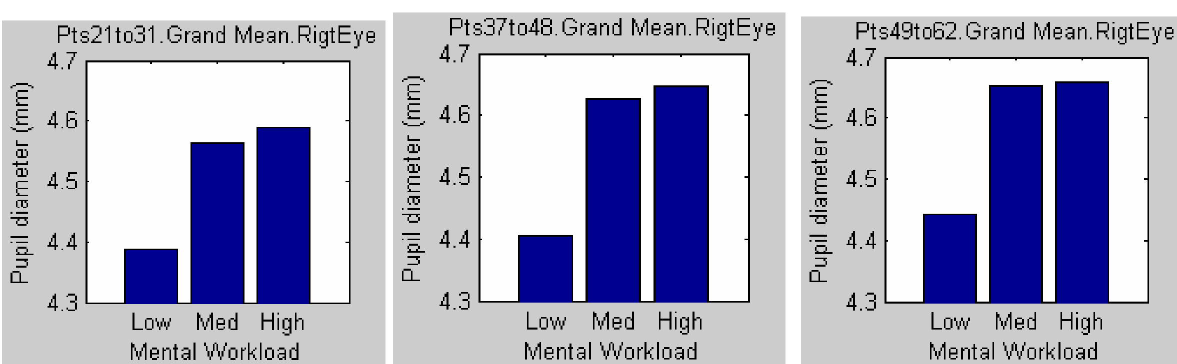

Figure 5.11. Grand mean of pupil diameter in the three MWL levels in all participants. The three graphs represent the pupil results for studies 1, 2, and 3, respectively.

However, the data in Figure 5.11 does not amenable to statistical analyses for two reasons. First, individual differences are not accounted for in Figure 5.11. Since there are large individual differences in pupil diameter baseline and pupil diameter increment, it is important to account for individual differences in a pupil analysis. Second, the three MWL levels combined both participants who did the 1, 2, and 3-back tasks and other participants who did 1, 3, and 4-back tasks. This difference on the x-axis is also not accounted for in Figure 5.11. These two challenges were ameliorated with a linear regression analysis, explained in the following paragraphs.

In order to statistically analyze the overall relationship between MWL and pupil diameter across all the participants in each study, a linear regression analysis was conducted. The predicted variable was the N level. Then, the N level was predicted using two predictors such as pupil diameter and the participant identification (ID) number. Pupil diameter was a continuous variable, while the participant ID numbers were categorical (or factor) variables. The participant IDs were treated as nominal variables in the statistic software "R."

The linear regression equation had one term on the left side, which was the estimated N levels, and 22 terms or more on the right side. For example, Study 1 had 11 participants and 22 terms on the right side. The predictors variables were the 11 participants' personalized baselines (or y-intercepts) and the 11 personalized slopes, multiplied by pupil diameter. Since the 22 terms were used for the prediction, the degrees of freedom was 105, which was calculated from the total number of trials, 127, subtracted by the number of predictor terms, 22, in Study 1.

Table 5.4 below shows the summary of the three studies on the relationship between MWL and pupil diameter. For each of the three studies, there was a statistically significant increase in the pupil diameter with increased N levels, $R^2 = 0.46$ for Study 1, $R^2 = 0.43$ for Study 2, and $R^2 = 0.48$ for Study 3.

TABLE 5.4

SUMMARY OF PUPIL DIAMETER

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.46**	**0.43**	**0.48**
Adjusted R^2	0.35	0.32	0.38
Standard error	0.77	0.73	0.80
F-statistics	4.23	3.87	4.72
p-value	0.001	0.001	0.001

Summary of the linear regression analysis predicting the N levels using the pupil diameter and the categorical participant identification numbers.

These results indicate that the pupil diameters were strongly affected by MWL – the more effort that was required to do the task, the larger the pupil diameter. The R^2 value of roughly 0.45 indicates that pupil diameter is strongly influenced by MWL, however, pupil diameter alone can explain only approximately 45% of MWL variability.

5.4.2 Saccadic Intrusions and MWL

Second, the saccadic intrusions (SI) measure was analyzed for its relation to MWL. The algorithm (explained in Chapter 3) categorized saccadic eye deviations (SEDs) into three groups using eye movement amplitudes. Also the algorithm produced three dependent variables (DVs) related to these SEDs in each trial. The three categories were microsaccades, SIs, and gaze aversion. This section shows the results on the SI measure which have specific amplitudes of eye movements ranging from 0.4 deg to 4.1 deg.

Figure 5.12 shows the change of eye angle (including SIs) during three trials on three MWL level conditions (1, 3, and 4-back tasks from left). The x-axis represents the time in seconds. The figures on top show the horizontal eye angle, theta (in green), and the calculated fixation baseline (in pink). The figures on bottom show the absolute values of eye deviation from the calculated fixation baseline (in red). The red lines indicate only the SI measure, having the amplitude changes from 0.4 and 4.1 degrees.

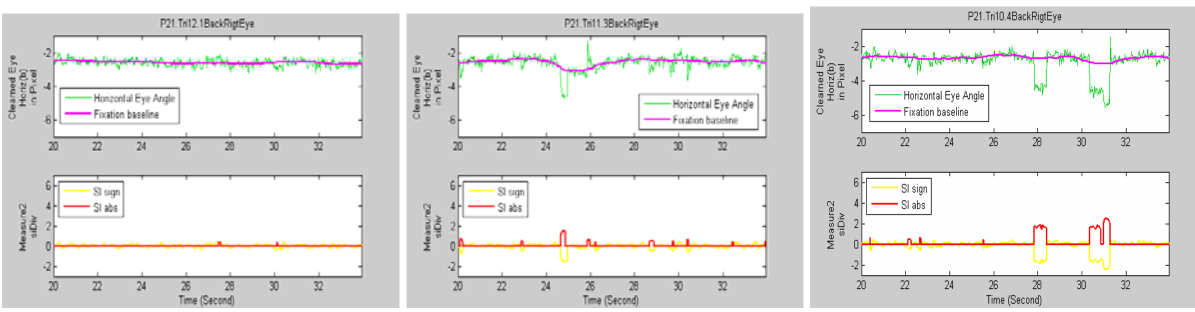

Figure 5.12: The three upper panels (light, medium, and heavy mental workload conditions) show samples of the horizontal eye movements and the calculated fixation baselines. The three lower panels show the corresponding saccadic intrusions.

These examples mainly emphasize three facts. First, it seems the eye movements were very stable while the observer was engaged in a light MWL task (the 1-back task in Figure 5.12 left), whereas there were more eye deviations from the calculated fixation baseline in the medium and heavy MWL tasks (such as 3, and 4-back tasks).

Second, the type of eye deviations seemed to be of the saccadic (rapid) type of eye movements (especially in Figure 5.12 right), rather than the slowly drifting type of eye deviation. Therefore, third, the algorithm to detect SIs seemed to work well. The algorithm ignored the slow eye deviations, observed from the 24 to 27 second marks in the medium MWL conditions (Figure 5.12 top middle). At the same time, the algorithm detected only “saccadic” (rapid) eye deviations (SEDs) in the same period of the eye data. These SEDs within fixational eye movement (FEM) periods, specifically SIs which had a middle range amplitude, were the interest of this dissertation. For more details, see the Algorithm chapter. These three main findings provide evidence consistent with the main hypothesis; higher MWL tasks would induce more SIs, rather than to induce other types of SEDs and slow drift eye deviation.

The algorithm (See Chapter 3) quantified SIs into a single value (the SI measure), which was used as a dependent variable (DV) for each trial. The cumulative amplitude degree of the saccadic eye deviation measure during the evaluation period of about 30 seconds was divided by the seconds (e.g. 30 seconds), producing a SI measure in deg/sec. An example is shown in Figure 5.13. The SI measures were larger when the participant (e.g., P21) was engaged in higher N-back tasks such as 3- and 4-back tasks, while the SI measures were practically non-existent for a simple task like the 1-back task. This example graph shows a relatively strong, positive relationship between the N levels and the SI measure (r = 0.591, p = 0.043) within one participant.

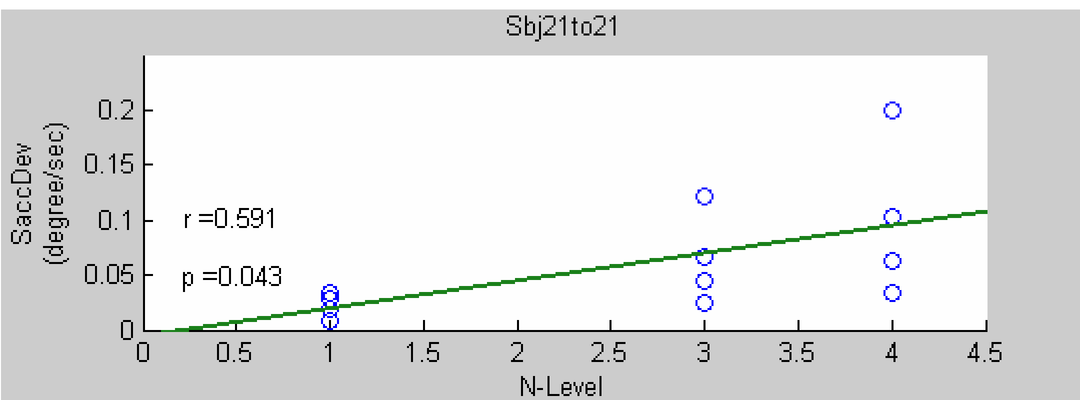

Figure 5.13. One participants 12 trial results of the correlation between the saccadic intrusion (SI) measure and the 3 mental workload (MWL) levels, such as 1-back (light MWL), 3-back (medium MWL), and 4-back tasks (heavy MWL).

Figure 5.14 shows the 37 participants' scatter plots of the correlation between the N-back levels and the SI measures. The correlation coefficient r values ranged from -0.30 to 0.84, with an average r value of 0.45. Out of 37 participants, 34 participants (92%) had a positive correlation, and three (P45, P48 and P49 in Figure 5.14) had a negative correlation. This value of 92 % satisfies the second criterion, the 84% threshold line.

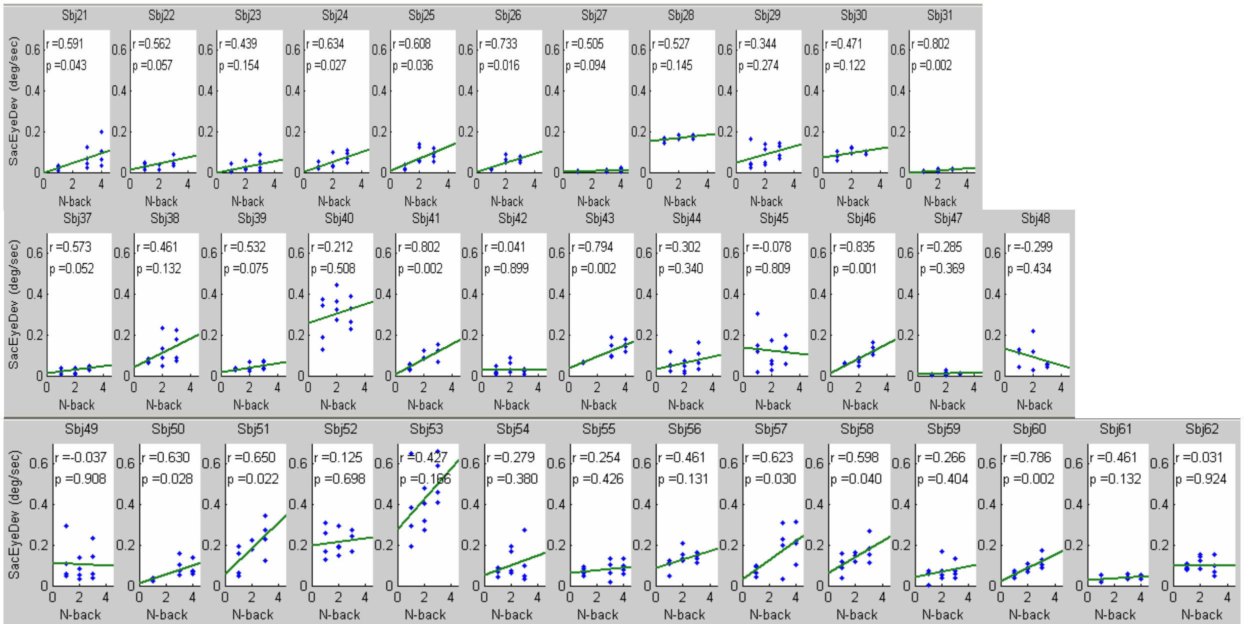

Figure 5.14. Correlation between the saccadic intrusion (SI) measure and N-back task levels. The N-back levels represent the mental workload (MWL).

Figure 5.15 shows a histogram of the correlation coefficients of the 37 participants. They consistently had a positive correlation coefficient value of 0.44. The fitted normal curve has the negative standard deviation at 0.17, which is above zero. This means, that the vast majority (theoretically 84% of people or more) had positive correlations which satisfies the third criterion. Because of the consistency in the positive values, this measure may well be generalizable to the rest of the population.

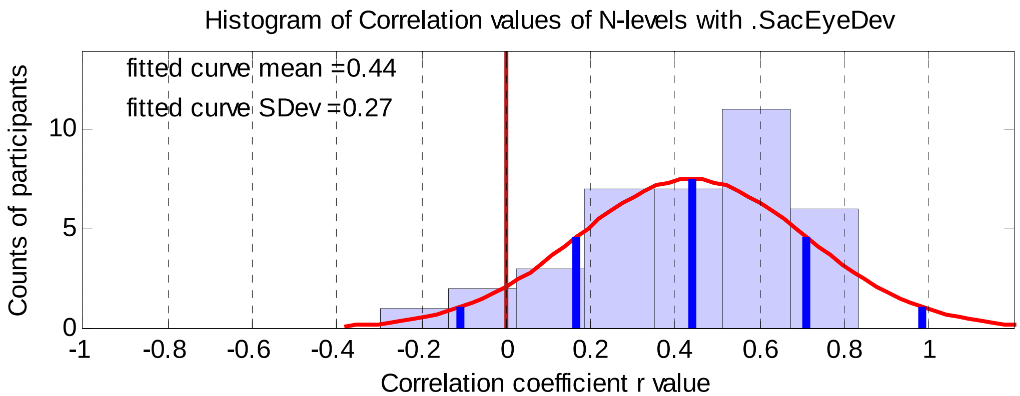

Figure 5.15. A histogram of the correlation coefficients in 37 participants between the N levels and the SI measure shows that the saccadic intrusion (SI) measure reliably estimated the MWL with the correlation coefficient r values of around 0.44.

Each participant's summary was collapsed into a grand mean score for each study (Figure 5.16). In each of the three studies, the SI measure increased as the MWL level increased. These show that overall, there seemed to be consistent effect. There were larger, longer, or a greater number of saccadic eye deviations when MWL was high. In Study 2 and 3, eye gaze moved from one location to another. However, the algorithm extracted SIs alone, and the SI measure increased with increased MWL.

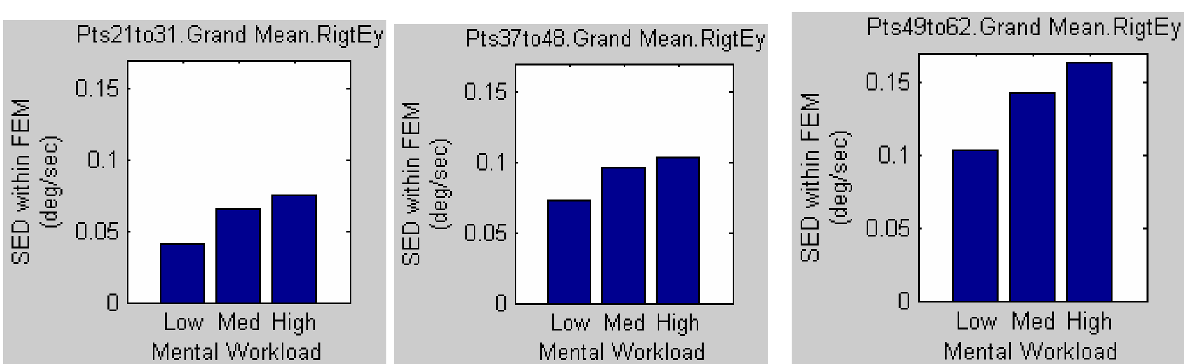

Figure 5.16. Grand mean of the saccadic intrusion (SI) measure in the three MWL levels of all participants. Overall, the increased MWL lead to the increased SI measure.

The overall relationship between MWL and the SI measure across all the participants in each study was statistically analyzed using a linear regression model. The N levels were predicted using two predictors: the SI measure and participant's IDs. The SI measure was a continuous variable, while the participant ID numbers were categorical variables.

A linear regression analysis was conducted for each of the three studies. Table 5.5 shows the summary of the three studies on the relationship between MWL and the SI measure. For each of the three studies, the result was statistically significant in increasing the SI measure with increased N levels, $R^2 = 0.38$ for Study 1, $R^2 = 0.33$ for Study 2, and $R^2 = 0.30$ for Study 3.

TABLE 5.5

SUMMARY OF THE SI MEASURE

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.38**	**0.33**	**0.30**
Adjusted R^2	0.25	0.20	0.16
Standard error	0.82	0.79	0.92
F-stats	3.03	2.52	2.21
p-value	0.001	0.001	0.002

Summary of the linear regression analysis predicting the N levels using the saccadic intrusion (SI) measure and the participant identification numbers as a categorical variable.

These results indicate that the MWL manipulation in the experiments induced the increased change in the SI measure. The R^2 values were smaller than the ones for the correlations between the pupil diameter and MWL (which was around 0.45 in Table 1). However, R^2 values larger than 0.30 are still large enough to explain MWL changes.

5.4.3 MicroSaccades and MWL

Within the three categories of SEDs, the SEDs with amplitudes of 0.4 deg or lower are called microsaccades in this dissertation (see the literature review in Chapter 2). Figure 5.17 shows the correlations between the microsaccade measure and the N-back levels on the participant level. The correlation r values of the 37 participants ranged from -0.81 to 0.67 with an average of -0.21. Out of 37 participants, 16 participants (43 %) had a positive correlation. This does not satisfy the second criterion of the consistent sign correlations. This is not consistent across participants, and it is not easy to generalize to the other population.

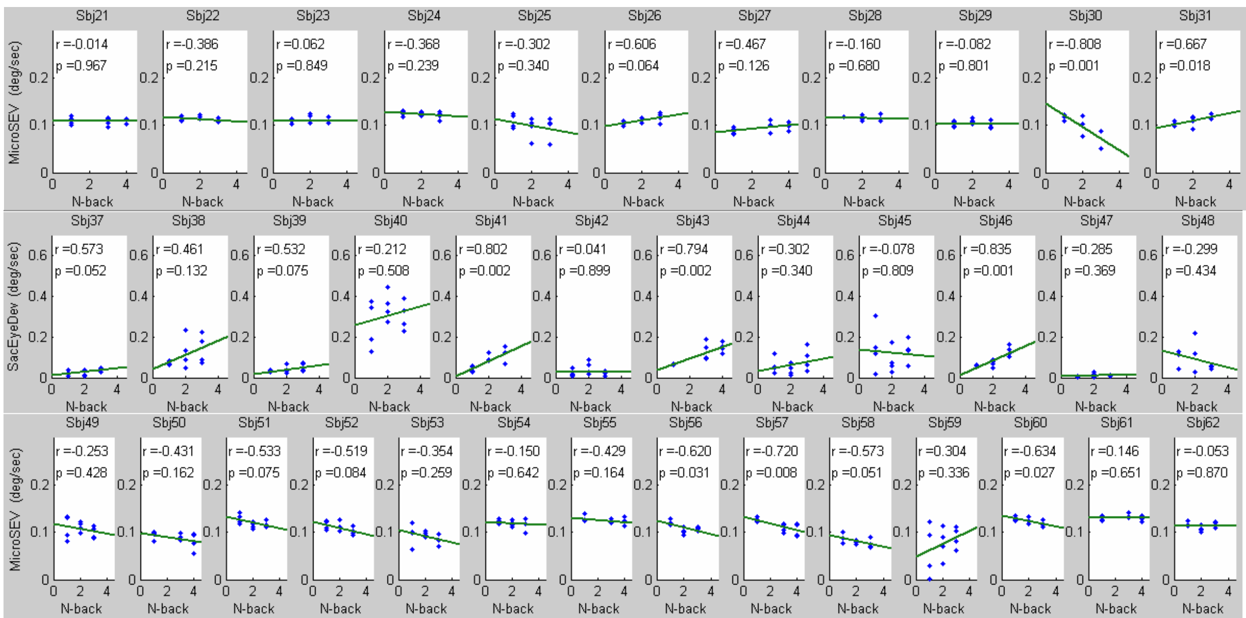

Figure 5.17. Correlation r's between the N-back levels and the microsaccade measure. The r values ranged from -0.81 to 0.67. The positive correlations were seen in 43 % of the participants.

Figure 5.18 shows a histogram of correlation coefficient r values for the 37 participants between the N levels and the microsaccade measurements. The distribution of the coefficient r values appears spread in a wide area from -0.8 to 0.6. This measure seems to be difficult to generalize to the rest of the population. Because of this inconsistency and spread distribution, the microsaccade measure was not included in the multiple-regression in the later section.

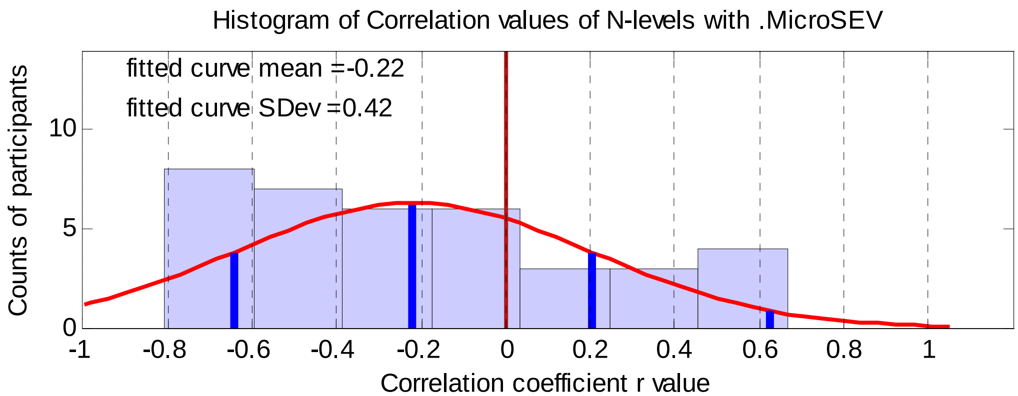

Figure 5.18. A histogram of correlation coefficient r value between the N level and the microsaccade measure. The microsaccade measure was not a reliable estimate of MWL.

In the linear regression analysis, the R^2 values for the correlation between N-levels using the microsaccade measure and the participant identification number as a categorical variable in Table 5.6 were 0.24, 0.31, and 0.28 for the three studies . The R^2 value of 0.24 is sufficiently large to predict a driver's MWL. But again, due to the inconsistency (Figure 5.18), this measure was not included in the multiple regression in the later section.

TABLE 5.6

SUMMARY OF MICROSACCADES

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.24**	**0.31**	**0.28**
Adjusted R^2	0.08	0.17	0.14
Standard error	0.91	0.80	0.93
F-stats	1.55	2.28	2.04
p-value	0.08	0.01	0.01

Summary of the linear regression analysis predicting the N levels using the microsaccade measure and the categorical participant identification numbers.

5.4.4 Gaze Aversion and MWL

The third of the three types of SEDs was gaze aversion, which was SEDs with amplitudes of 4.1 deg or larger. For each trial, gaze aversion was quantified into one single DV, the gaze aversion measure (see the Algorithm chapter). The correlation analysis was conducted to examine the relation between the gaze aversion measure and MWL. On the individual level, the r values ranged from -0.40 to 0.74, having an average of 0.22 (Figure 5.19). Out of 37 participants, 27 participants (73 %) had a positive correlation. This value of 73% does not satisfy the criterion of 84 % consistent correlations. The distribution of the correlation r values is plotted in Figure 5.20.

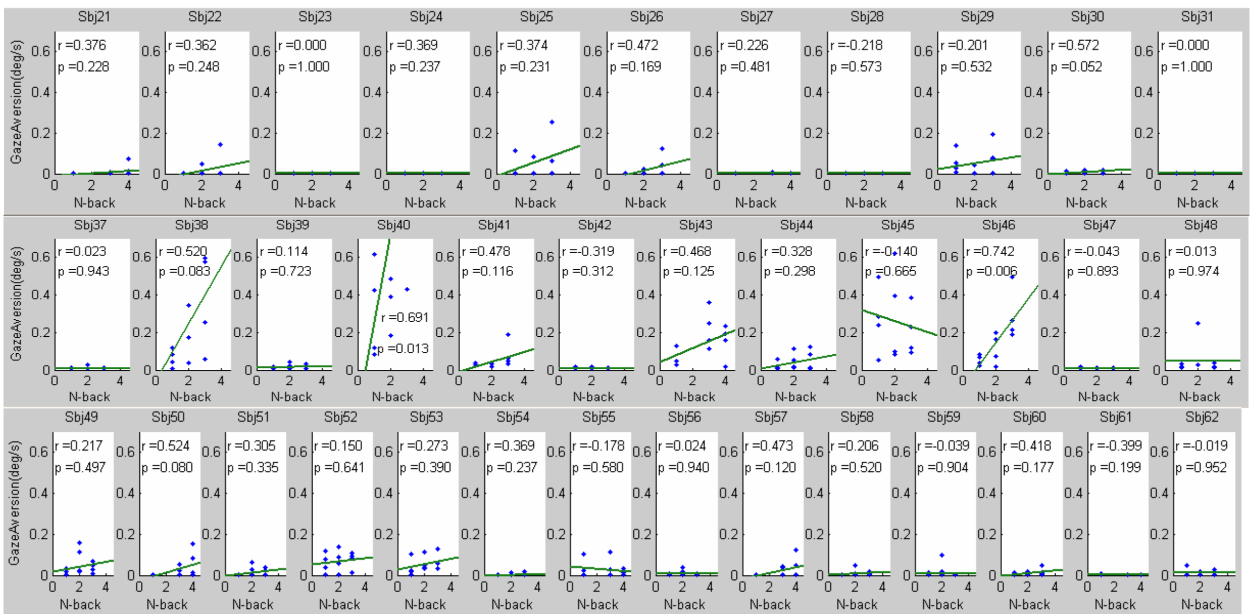

Figure 5.19. Correlation between the N-back levels and the gaze aversion measure (deg/sec) for each participant. The mean r value for the 37 participants was 0.22.

Figure 5.20 shows a histogram of correlation coefficients between the N levels and the gaze aversion measure. The fitted normal curve is fairly broad, ranging from -0.40 to 0.74. The inconsistency is shown by the location of the curve in relation to zero. The criterion value was the -1 standard deviation, which did not lie above zero. That means that the distribution spread across both negative and positive values. This measure is difficult to generalize to the rest of the population for reliably estimating MWL.

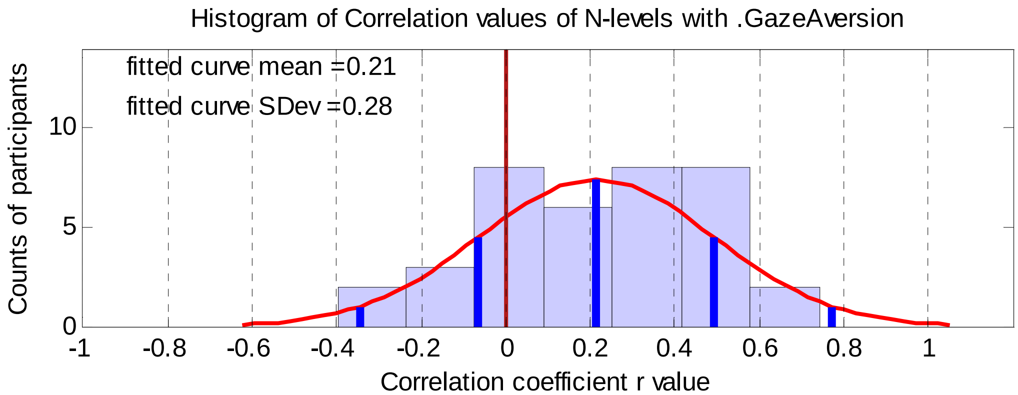

Figure 5.20. The normal curve is spread across the correlation value of zero. This gaze aversion measure is difficult to generalize to the rest of the population because the distribution of the r values disperses from the negative to the positive values. This makes it difficult to use the gaze aversion measure for a linear regression model.

A linear regression analysis was conducted for each of the three studies. Overall, the R^2 values were 0.18, 0.21, and 0.19 for the three studies to predict the N levels using the gaze aversion measure and the participant identification number as a categorical variable (Table 5.7). These R^2 values are high enough to contribute to MWL accounts. However, this gaze aversion measure was not used in the multiple regression later in this chapter due to the inconsistency of the correlation signs.

TABLE 5.7

SUMMARY OF GAZE AVERSION

	Study 1 Fixation Task	Study 2 Random Dot Task	Study 3 Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
Deg of freedom	105	117	140
R^2	**0.18**	**0.21**	**0.19**
Adjusted R^2	0.03	0.06	0.04
Standard error	0.94	0.86	0.99
F-statistics	1.12	1.36	1.24
p-value	0.28	0.15	0.21

Summary of the linear regression analysis predicting the N levels using the gaze aversion measure and the participant identification numbers as a categorical variable.

This section examined the relationship between N-back levels and each of the three categories of SEDs, such as SIs (medium amplitudes), microsaccades (low amplitudes), and gaze aversion (large amplitudes). All of them had the R^2 values 0.18 or better. That means all the three kinds of SIs showed some relation to the N-back levels. Among the three kinds of SEDs, the SI measure had the highest R^2 value (0.38) and had correlation coefficients consistent across participants in the three studies. Among these three measures, only the SI measure was used in the multiple correlation in the later section in this chapter because the SI measure was the only one that satisfied the criterion of consistent sign correlations.

5.4.5 Fourier Transformation

The fifth, sixth, and seventh single predictors to predict MWL were Fourier transformation (FT) power density. Figure 5.21 shows a series of eye movement example data in three consecutive trials on the three MWL levels (1, 3, and 4-back tasks) from left to right. The fixation eye movements seemed to be more stable in lower N-back tasks and instable in the higher N-back tasks. This instability in the eye movements were quantified into three single values using the FT power density.

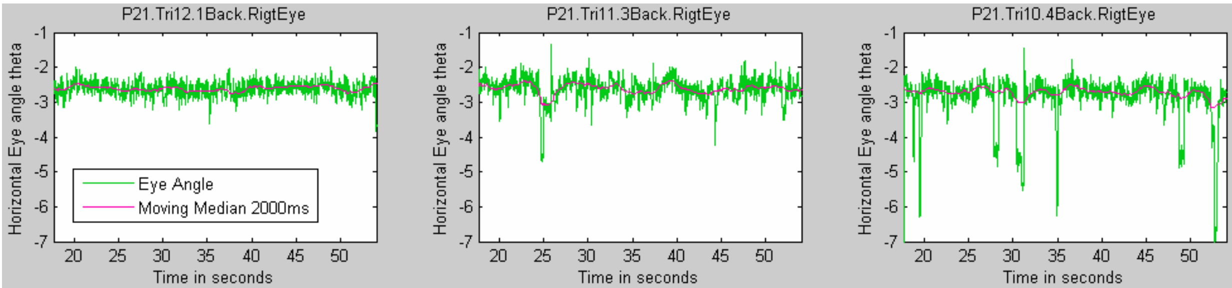

Figure 5.21. Eye movement data in three trials; 1-back (Figure 5.21 left), 3-back (Figure 5.21 middle), and 4-back tasks (Figure 5.21 right). These tasks served to implement light, medium, and heavy mental workload (MWL) levels, respectively. Gaze fixation is looser in the higher N-back tasks (heavy MWL).

The eye movement data in a time-domain representation (Figure 5.21) were transformed into a frequency-domain representation (Figure 5.22) using FT. The focus of this dissertation is the power density related to saccadic intrusions (SIs). As the Algorithm chapter (Chapter 3) explained, SIs are most likely to show their influence on the eye movements in specific frequency components, which range from 1.14 to 16.67 Hz. Using this range, the FT data were quantified into three values; power density in the lower frequency band from 0 to 1.14 Hz that is not related to SIs, power density in the middle frequency band from 1.14 to 16.67 Hz, which is related to SIs, and power density in the higher frequency band from 16.67 to 25 Hz that is not related to SIs. See more in the algorithm chapter.

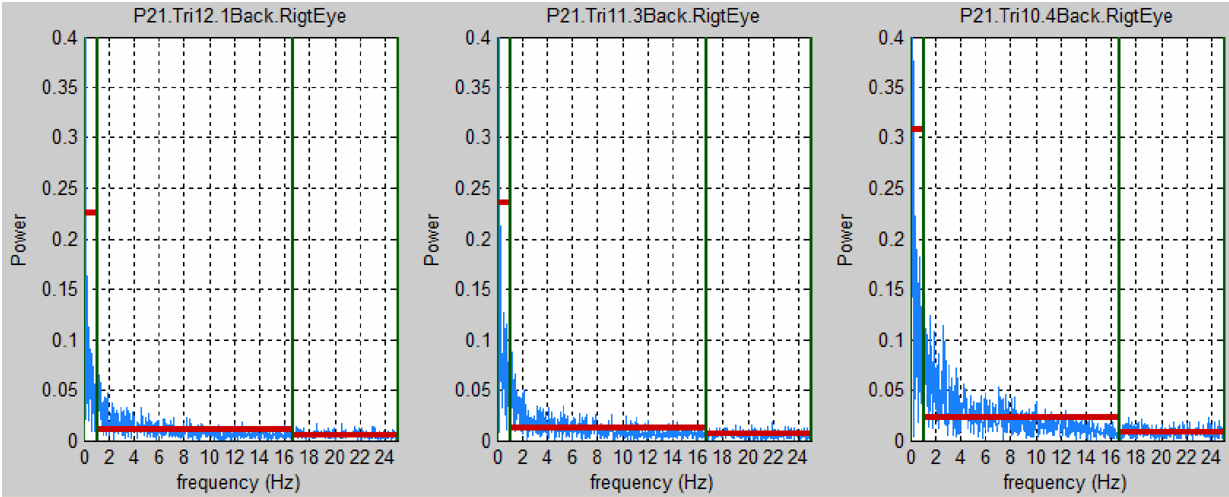

Figure 5.22. Examples of frequency-domain representations transformed from Figure 5.21 by using Fourier transformation (FT). These three graphs were for three mental workload (MWL) levels: light, medium, and heavy MWL from left to right. The power density (the y-axis) seems to be larger at the majority of the frequencies (the x-axis) with heavy MWL (Figure 5.22 right).

5.4.6 Low Band FT

For the 37 participants, the correlation values r ranged from -0.74 (in P41) to 0.68 (in P46). Thirty participants (81%) had a positive correlation. The other seven participants who had a negative correlation were P23, P24, P41, P45, P47, P49, and P59 (Figure 5.23). The author has defined the three criterions to evaluate the seven physiological measures. One of the criterions was to have 84 % of participants showing the consistent correlations. The low band FT with 81 % of participants having a positive correlation did not satisfy this criterion.

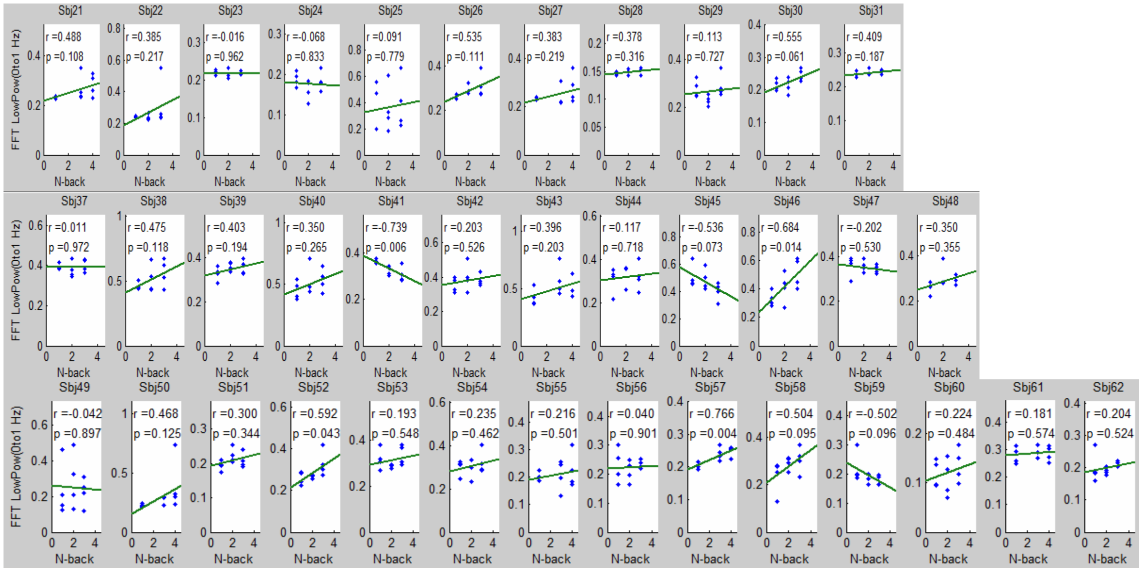

Figure 5.23. Individual level correlation between the N levels and the average power density in the lower band Fourier Transformation (FT). The correlations r ranged from -0.74 to 0.68. Out of 37, 30 participants (81%) had a positive correlation.

Plotted in Figure 5.24 is a histogram of these correlation values between the N-back levels and the low band FT measure. The average correlation r was 0.22. The -1 standard deviation was at r = -0.11, which was below zero. This does not satisfy the criterion. The distribution was too broad from negative to positive values to consistently estimate the N back levels.

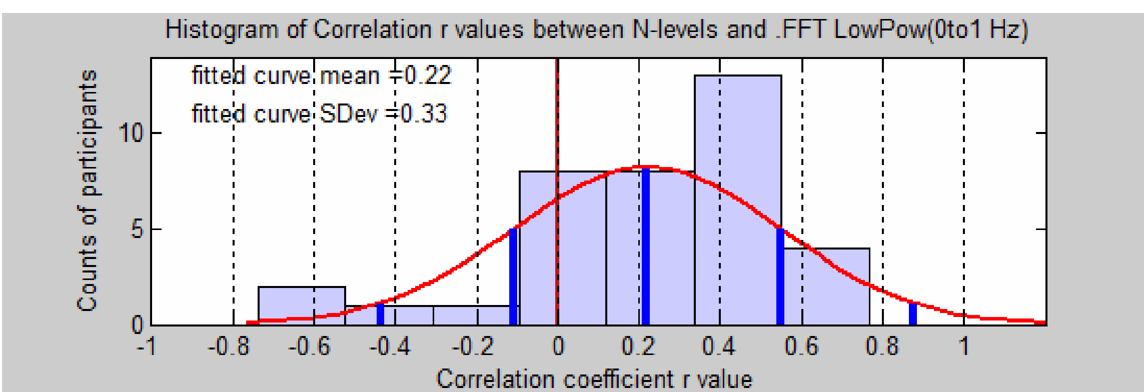

Figure 5.24. The fitted normal curve on a histogram of the correlation r values between the N-back levels and the low band FT measurements. This graph shows the r values overly spread across the correlation value of zero.

The average values representing the lower band FT frequencies were used for a linear regression to predict the N levels in each of the three studies. Three linear regressions produced three R^2 values 0.21, 0.22, and 0.24 (Table 5.8).

TABLE 5.8

SUMMARY OF LOW FREQUENCY BAND FOURIER TRANSFORMATION

Low Band FT)

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	22	24	28
R^2	**0.21**	**0.22**	**0.24**
Adjusted R^2	0.05	0.06	0.09
Standard error	0.93	0.86	0.96
p-value	0.19	0.12	0.03

Table 5.8. Summary of three linear regression analyses to predict the N levels (i.e. mental workload) using two predictors, such as the average power density for lower band frequencies (0 to 1.14 Hz) in Fourier transformation (FT) frequency-domain representation and the participants' identification numbers as a categorical variable.

5.4.7 Middle Band FT

Figure 5.25 participants' correlation r values which ranged from -0.23 (in P29) to 0.80 (in P58). Twenty-eight participants (76%) had a positive correlation. The other nine participants who did not have a positive correlation were P29, P37, P45, P48, P53, P55, P59, P61, and P62. This percentage of 76 % does not satisfy the criterion of 84 %, so this middle band FT is not consistent enough to reliably estimate the N levels or mental workload (MWL).

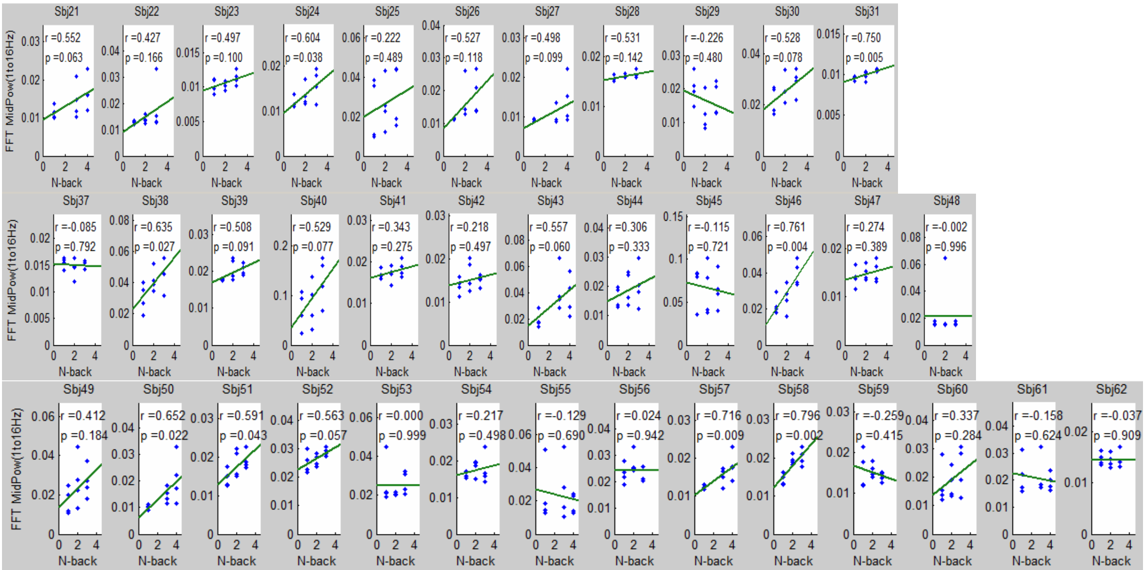

Figure 5.25. Correlation r values between the N levels and the medium band FT power density for each of the 37 participants. The r values ranged from -0.23 to 0.80. Twenty-eight participants (76%) had a positive correlation.

A histogram of the r values is plotted in Figure 5.26. The fitted normal curve has the mean of r = 0.34, which was higher than the two other FT bands (r = 0.22 for the low band in Figure 5.24 and r = 0.27 in Figure 5.28). Also the -1 standard deviation line was at 0.03, which was above zero. This value satisfies one criterion; a good MWL measure should have the normal fit curve’s 84 percentile rank (which is the -1 standard deviation line) above zero. This usually indicates that the vast majority of the participants had a positive correlation, but this was not the case; the previous paragraph shows that 76% of participants achieved a positive correlation. This makes the medium band FT measure difficult to generalize to the rest of the population.

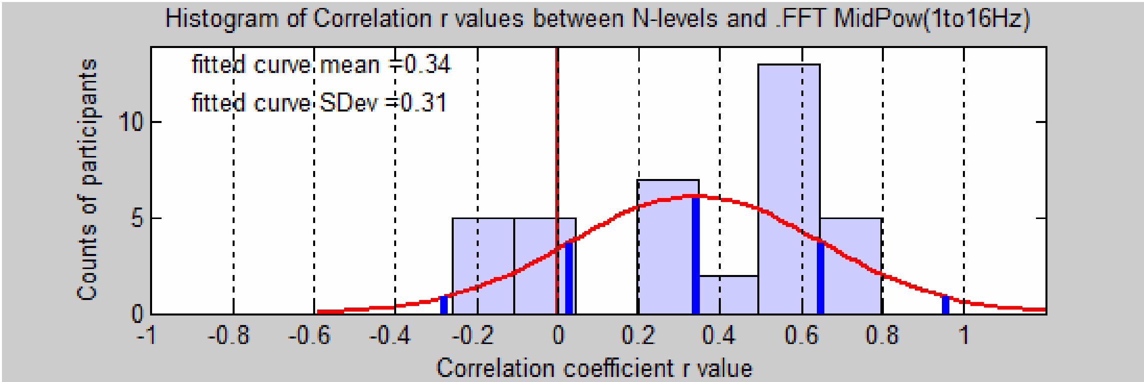

Figure 5.26. A histogram of the correlation r values between the N levels and the medium band FT measurements. The average r among the 37 participants was the highest (r = 0.34) among the three bands of FT. The fitted normal curve shows that the -1 standard deviation line at 0.03, which is above the criterion line of zero. However, the actual counts of participants who had a positive correlation were only 76%, which was the below the criterion line of 84% (the theoretical percentage of the distribution above the -1 standard deviation). This is difficult to generalize to the rest of the population.

The results of the linear regression analyses were summarized in Table 5.9. The R^2 values were 0.32, 0.23, and 0.28 (Table 5.9). Since the medium frequency band was supposed to relate to SIs, higher R^2 values were hypothesized. The results show that Study 1 (Table 5.9) had the higher R^2 value than Study 2 and Study 3.

TABLE 5.9

SUMMARY OF MIDDLE FREQUENCY BAND FOURIER TRANSFORMATION

		Study 1	Study 2	Study 3
Mid Band FT FT (1.14 to 16.67 Hz)		Fixation Task	Random Dot Task	Free-viewing Task
	# of Participants	11	12	14
	Total trials	127	141	168
	Predictor terms	22	24	28
	R^2	**0.32**	**0.23**	**0.28**
	Adjusted R^2	0.18	0.08	0.14
	Standard error	0.86	0.85	0.94
	p-value	0.01	0.07	0.01

Summary of three linear regression analyses to predict the N levels using two predictors, such as the average power density for medium band frequencies (1.14 to 16.67 Hz) in FT and the participants' identification numbers as categorical variables.

5.4.8 High Band FT

Figure 5.27 shows participants' correlation values r ranging from -0.38 (P54) to 0.86 (P46). Thirty-one participants (83.8%) had a positive correlation, and the other six participants (P29, P45, P53, P54, P55, and P62) had a negative correlation. This does not satisfy the criterion; a good MWL measure should have a positive correlation with MWL (i.e., the N levels) in 84.1 % or more participants.

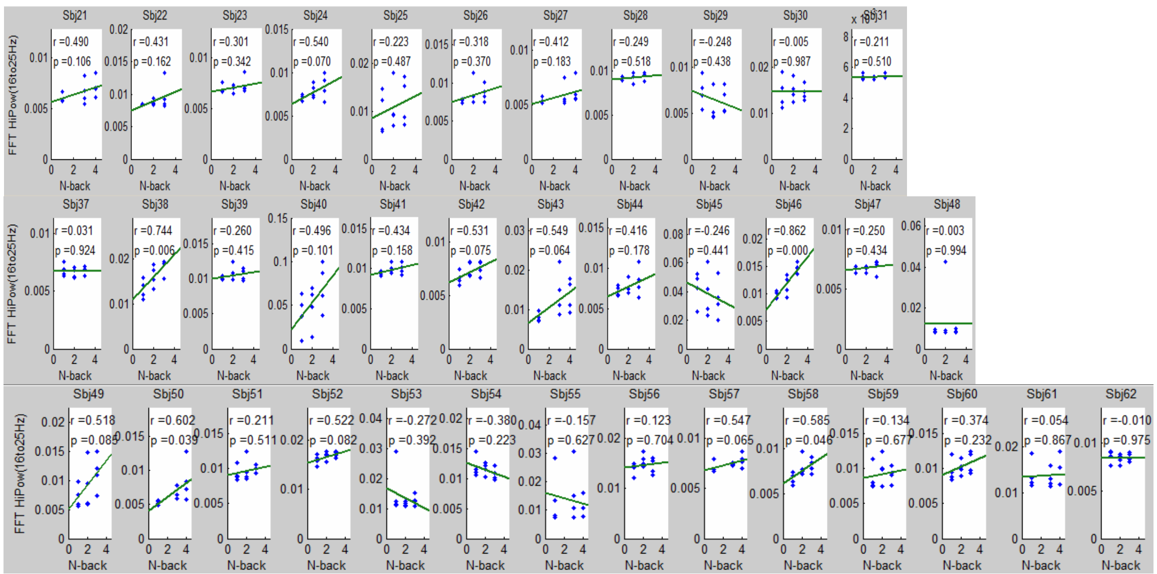

Figure 5.27. Individual level correlation r values between the N level and the high band FT power density in 37 participants. Eighty-four percent of participants had a positive correlation.

These correlation r values are plotted in a histogram in Figure 5.28. The average r was 0.27 for the 37 participants. The -1 standard deviation line was -0.03, which was below zero. This does not satisfy the criterion; a good MWL measure should have a fitted normal curve with the -1 standard deviation value of zero or higher. This high band FT measure is difficult to generalize to the rest of the population.

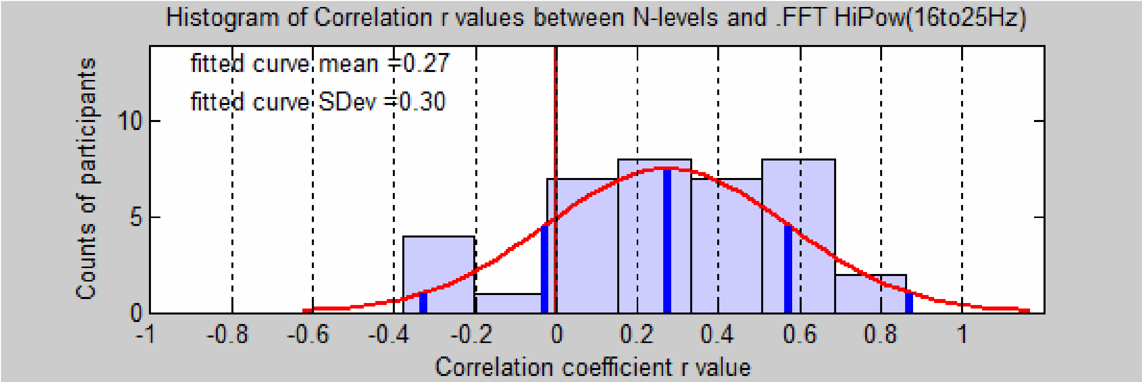

Figure 5.28. Correlation between the N levels and the high band FT power density.

Table 5.10 shows the summary of the linear regression analysis for each of the three studies. The N back level was predicted using the high band FT power density and participants' identification numbers. The linear regression analyses produced three R^2 values 0.20, 0.27, and 0.23 (Table 5.10).

TABLE 5.10

SUMMARY OF HIGH FREQUENCY BAND FOURIER TRANSFORMATION

High Band FT FT (16.67 to 25 Hz)		Study 1 Fixation Task	Study 2 Random Dot Task	Study 3 Free-viewing Task
	# of Participants	11	12	14
	Total trials	127	141	168
	Predictor terms	22	24	28
	R^2	**0.20**	**0.27**	**0.23**
	Adjusted R^2	0.04	0.13	0.08
	Standard error	0.93	0.83	0.97
	p-value	0.22	0.02	0.06

Summary of three linear regression analyses to predict the N levels using two predictors, such as the average power density for medium band frequencies (16.67 to 25 Hz) in FT and the participants' identification numbers as categorical variables.

The results in Tables 5.8, 5.9, and 5.10 indicate that the three bands of FT power density predicted the N levels relatively accurately, having nine R^2 values ranging from 0.20 to 0.32 in for the three studies. In other words, the MWL manipulation in the experiments induced more eye movement fluctuation. As the first study to use FT analyses for SIs, the results offer some evidence in favor of a relationship between the FT power density in the certain frequency bands and MWL manipulation.

However, the FT power density measures for SI analyses were still too crude to consistently demonstrate similar results across participants. Each of the three frequency bands of FT power density measures did not match the criteria (Figure 5.23, 24, 25, 26, 27, and 28) for becoming a future measure to estimate MWL. The two necessary criterions were (1) to have the actual mass body (84% of the participants) consistently on the positive correlation side (or consistently on the negative correlation side), and (2) to have the fitted normal curve mass body (the area between the negative and positive standard deviations of one) consistently on the positive (or consistently on the negative) correlation side. It seems difficult to use any of the three bands of the FT measures to reliably estimate MWL at this point. A specific method to improve the FT measures is discussed in the Discussion chapter (Chapter 6).

5.4.9 Group Trends Analysis

Trends tests were conducted to examine the seven physiological variables' changes when MWL varied. Most participants had four trials for each of the three MWL levels and the average of those trials was used in the analyses. The three average scores for each participant were examined in their rank-order trends. For example, pupil diameter is expected to have an order of small, medium, and large for the three MWL levels of low, medium, and heavy. Figure 5.29 shows out of 37 participants, 20 participants had pupil diameter changes in the expected order

[Low < Med < High], 16 had [Low < High < Med], and 1 had [High < Low < Med]. If the trends of the rank order were random, the six possible combinations would have equally distributed counts in each of the six trend combinations. A chi-square test revealed that the probability to have 12 or more participants out of 37 in one of the six categories is lower than 0.1% (See Figure 5.29). As expected, both the SI measure and pupil diameter had the most participants in the L<M<H category. Eighteen participants had the increasing trends (L<M<H) for the SI measure, and 20 participants for pupil diameter. The chances are lower than 0.001 that this would be the case were the results truly random. This shows that both SI measure and pupil diameter significantly and consistently increased when MWL was high.

Similarly, the mid band FT had 21 and the high band FT had 19 participants with increasing trends (L<M<H), out of 37 participants. The probability of this occurring randomly is lower than 0.001. This shows that eye movements with 1.14 Hz frequency or higher significantly and consistently increased when MWL was high.

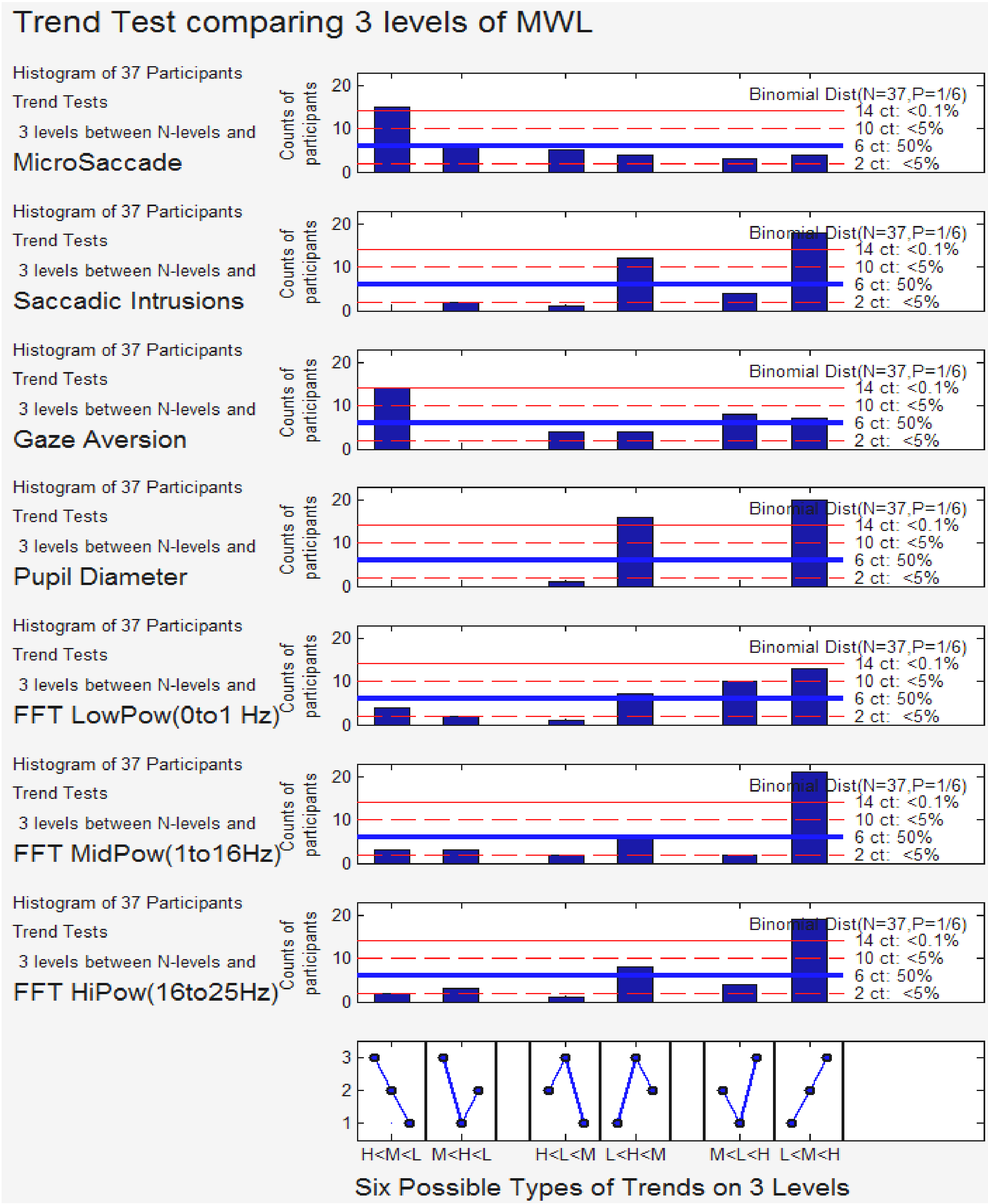

Figure 5.29. Results of the trends tests. There were six possible types of trends on the three mental workload (MWL) levels. The number of participants who had the increasing trend (Low < Mid < High) was significantly higher than the chance level (1/6) on the saccadic intrusion measure, pupil diameter, low FT, mid FT, and high FT.

5.4.10 Section Summary

The previous sections examined the relationship between the N-back levels (serving as MWL) and each of the seven single predictors with participant's identification number as a categorical variable in the linear regressions. Table 5.11 below shows the summary of these seven single predictors as well as the two performance measures and one subjective rating measure. Each of these nine is a measurement to estimate MWL, however, only the physiological measures (the seven predictors) can be used in estimating a driver's MWL in real-time.

TABLE 5.11

SUMMARY OF TEN MENTAL WORKLOAD MEASURES

	Predictors	R^2			# of participants with positive correlation	Usage in Multiple regression
		Study 1	Study 2	Study 3		
		Fixation Task	Random Dot Task	Free-viewing Task		
Performance Measures	N-back Performance (%)	0.47	0.50	0.40		No
	Reaction Time	0.44	0.36	0.41		No
Subjective Rating	Subjective Rating (Uni-scale)	0.88	0.81	0.78		No
Physiological measures	Pupil Diameter	0.46	0.43	0.48	36 / 37 (97%)	Yes
	SI measure	0.38	0.33	0.30	34 / 37 (92%)	Yes
	microsaccade measure	0.24	0.31	0.28	16 / 37 (43%)	No
	gaze aversion measure	0.18	0.21	0.19	27 / 37 (73%)	No
	FT (Low band)	**0.21**	**0.22**	**0.24**	30 / 37 (81%)	No
	FT (Mid band)	**0.32**	**0.23**	**0.28**	28 / 37 (76%)	No
	FT (High band)	**0.20**	**0.27**	**0.23**	31 / 37 (83.8%)	No

R^2 values to predict the N levels using each of the single predictors with help of the participants identification number as a categorical variable.

Several conclusions can be drawn from the results summarized in this table. First, Table 5.11 shows that all ten MWL measures were related to the N-back levels with an R^2 value of 0.18 or better. They are all large enough to be meaningful. Second, each of the ten MWL measures had consistent R^2 values across the three studies. The largest gap was 0.10 in the N-back performance (ranging from R^2 = 0.40 to 0.50) and another 0.10 in the subjective rating (ranging from R^2 = 0.78 to 0.88). This shows all of the nine MWL measures are reliable across studies.

Third, Table 5.11 shows, that out of the seven physiological measures, only pupil diameter had high R^2 values (0.43 or better). This has been shown in the past by many researchers (Just and Carpenter, 1993; Beatty & Wagoner, 1978). Fourth, out of the three SED measures (i.e., SIs, microsaccades, and gaze aversion), the SI measure has the highest R^2 values (0.30 or better). This is consistent with the dissertation hypothesis that SIs increases when MWL increases. Fifth, out of the three FT measures, the middle band has highest R^2 values (0.23 or better). The middle band FT is related to SIs. This is also consistent with the dissertation hypothesis.

Sixth, the two right columns show the percentage of participants who had a positive correlation between the single predictor and the N-back levels. In an automatic MWL estimation system, consistent directions of physiological changes are more preferable. The two measures that had more than 84.1 % consistency among the 37 participants were pupil diameter and the SI measure. These two were used for the multiple regression analysis in the next section.

5.5 Predicting MWL Using Multiple Predictors

5.5.1 Multiple Predictors

In this section, the N-levels were predicted using several of the single predictors together. These predictors were pupil diameter, the SI measure, and the individual identification number. Since there are three predictors, the number of predictor terms on the right hand side ofthe linear regression equation is three-fold the participant number. For example, Study 1 had 11 participants, therefore, Study 1 had 33 predictor terms. Subtracting 33 from the total trials, 127, the degrees of freedom for Study 1 was 94 (See Table 5.12 below).

TABLE 5.12

SUMMARY OF MULTIPLE REGRESSION USING PUPIL DIAMETER AND THE SI MEASURE

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	33	36	42
Deg of freedom	94	105	126
R^2	**0.58**	**0.58**	**0.54**
Adjusted R^2	0.44	0.44	0.40
Standard error	0.71	0.66	0.78
F-stats	4.09	4.17	3.67
p-value	0.001	0.001	0.001

Summary of three linear regression analyses for the three studies. Each analysis predicted the N levels using three kinds of predictors: the pupil diameter, the saccadic intrusion (SI) measure, and participants' identification number.

Table 5.12 above shows the results of the multiple regression models for each of the three studies. The R^2 values were 0.58, 0.58, and 0.54. Each of these was closer to 1.00 than the R^2 values using either one of the predictors (such as pupil diameter in Table 5.4 or the SI measure in Table 5.5). Since the pupil measure and the SI measure rely on different information resources (one derived from the pupil, and the other derived from eye movements), there seems not much overlap between the predictors; the R^2 value is much higher than the prediction by a single predictor in the previous section (such as Sections 4.1 and 4.2 in this chapter).

Table 5.12 also shows the most optimal estimation of MWL using two of the most consistent, reliable, and predictable physiological measures: pupil diameter and saccadic intrusions. (See Section 4: Predicting MWL using single predictor in this chapter for the statistical support of consistency, reliability, and predictability). However, if a model could focus on just the predictability aspect ignoring consistency and reliability, the R^2 values would increase, which is shown in Table 5.13 below.

Table 5.13 below shows the results of multiple regression models using the seven physiological measures. As the previous section explained, the five measurements (microsaccades, gaze aversion, low band FT, mid band FT and high band FT) are not usable because of the inconsistency of the physiological change directions. However, a future algorithm may pin-point specific physiological measures and may have consistent correlation signs. The R^2 values were as high as 0.83, 0.87, and 0.77. However, one of the adjusted R^2 values (Study 3) decreased from 0.40 in Table 11 to 0.31 in Table 5.12. This shows the multiple regression model in Table 5.13 has unwanted predictor(s).

TABLE 5.13

SUMMARY OF MULTIPLE REGRESSION ANALYSES USING ALL THE SEVEN PHYSIOLOGICAL MEASURES

	Study 1	Study 2	Study 3
	Fixation Task	Random Dot Task	Free-viewing Task
# of Participants	11	12	14
Total trials	127	141	168
Predictor terms	86	96	112
Deg of freedom	41	45	56
R^2	**0.83**	**0.87**	**0.77**
Adjusted R^2	0.46	0.58	**0.31**
Standard error	0.70	0.57	0.84
F-stats	2.28	3.08	1.68
p-value	0.002	0.001	0.017

Results of three multiple regression analyses using all the seven physiological measures and the participants' identification numbers as a categorical variable. Although the correlation values are very high, some predictors were not consistent, therefore, it is difficult to generalize to the rest of the population.

5.6 Comparison Between Predictors

5.6.1 Introduction

This dissertation used seven kinds of physiological responses as a measure to predict MWL: pupil diameter, three types of SED measures, and three types of Fourier transformation (FT) measures. This section compares the predictability among these seven predictors. Usually the beta weights in a linear regression can be a good measure for determining the value of predictors in a regression model. However, the units and the distribution of each predictor are different from each other, which makes direct comparison difficult. Therefore, another set of variables, the individual's correlation coefficients, were used for the comparison of the predictors. Correlation coefficient *r* values were calculated between MWL and a single predictor within each participant. The *F*-test or the *t*-test compares the *r* values between physiological measures.

5.6.2 Correlation Coefficient Values for Each Individual

Within each participant, the correlation between MWL and a single predictor was calculated using the same method as shown in Section 4.1 (the correlation between MWL and the pupil diameter in Participant 21).

Figure 5.27 shows each participant's correlation coefficient r value between the N levels (which is assumed to be equivalent to the MWL levels) and pupil diameter. The average r of all the participants in each of the three studies was 0.57 or better.

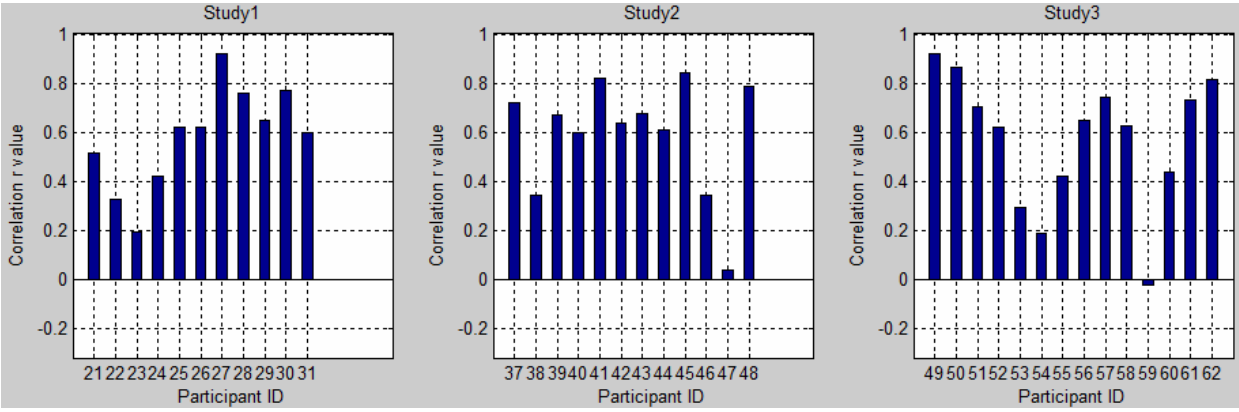

Figure 5.30. Correlation coefficient r values between pupil diameter and MWL in each of 37 participants. Study 1 had an average r value of 0.58, ranging from 0.2 to 0.92. Study 2 had an average r value of 0.59, ranging from 0.04 to 0.84. Study 3 had an average r value of 0.57, ranging from -0.02 to 0.93.

Likewise, Figure 5.28 shows the correlation coefficient r values between the N-levels and the SI measure for each of 37 participants. The overall average coefficient r value was 0.45.

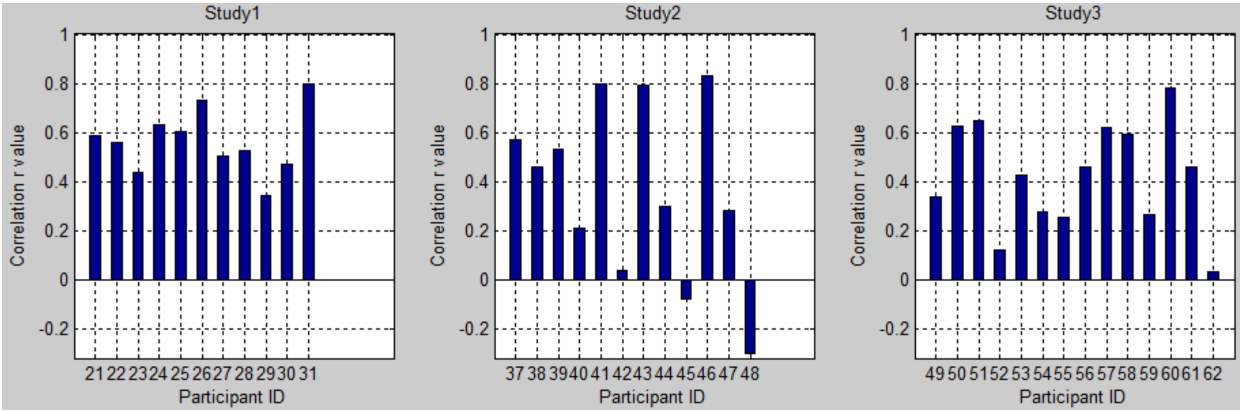

Figure 5.31. Correlation coefficient r values between the saccadic intrusion (SI) measure and MWL in 37 participants. Study 1 had an average r value of 0.57, ranging from 0.34 to 0.80. Study 2 had an average r value of 0.37, ranging from -0.30 to 0.84. Study 3 had an average r value of 0.42 ranging from 0.03 to 0.79.

5.6.3 Comparison Between Predictors

Three paired-sample *t*-tests were conducted in three studies to evaluate the various predictors on their predictability, specifically between pupil size and MWL and the SI measure and MWL. Table 5.14 below shows that the summary of the three *t*-tests. The average of all participants' correlation coefficients to predict the N-levels were always higher with the pupil diameter than with the SI measure. However, with the sample sizes 14 or smaller, the differences were not statistically significant.

TABLE 5.14

SUMMARY OF COMPARISON BETWEEN PUPIL DIAMETER AND THE SI MEASURE

IN TERMS OF THEIR CORRELATIONS WITH MENTAL WORKLOAD

	Study 1	Study 2	Study 3
Observations	11	12	14
Mean (Pupil)	0.58	0.59	0.57
Mean (SI)	0.57	0.37	0.42
Variance (Pupil)	0.04	0.06	0.08
Variance (SI)	0.02	0.13	0.05
t Stat	0.25	1.77	1.60
P(T<=t) two-tail	0.81	0.09	0.12
Result	Not Significant	Not Significant	Not Significant

Results of *t*-tests to examine the two types of correlation coefficients. One type was for the correlation between MWL and pupil diameter. Another was for the correlation between MWL and the SI measure. Although pupil diameter was correlated with MWL more strongly than the SI measure, the differences were not significant.

5.6.4 Comparison Among the Three Types of Saccadic Eye Deviation

The SI quantification algorithm (Chapter 3) was developed based on empirical data characterizing the properties of SIs (e.g., amplitude range and frequency range). If MWL invoked SIs, the eye movement in the range that is related to SIs should have stronger correlation with the N level.

First, this dissertation categorized eye deviation amplitude into three ranges – low, middle, and high. Abadi and Gowen (2004) and McGivern and Gibson (2006) showed that the most SIs had an amplitude ranging from 0.4 to 4.1 degrees. These parameters were used as the lower and upper amplitude thresholds to distinguish the three amplitude levels of saccadic eye deviations (SEDs), such as microsaccades, SIs, and gaze aversion. If the N level is linked to the behavior of SIs, it must lead to a stronger correlation between the N level and the quantified SI measure (the middle range amplitude from 0.4 to 4.1 deg), rather than the lower range amplitude (the microsaccade measure) or the higher range amplitude (the gaze aversion measure). This seems to be true; Table 5.11 (in this Chapter) shows the R^2 values to predict the N-level were always higher with the SI measure (0.30 or better) than the other two measures (0.24 or better and 0.18 or better, respectively). In addition to that, each participant's correlations were compared using the *F* test. Table 5.15 shows the average r values in three categories of SED in three studies. In each of the three studies, the average r value was the highest for correlation between the N level and the SI measure than for the two other kinds. The results of the *F* tests show that the differences were significant (Table 5.15). The follow-up *t* tests (Table 5.15) show that the differences between the SI-MWL correlation and the Gaze aversion-MWL correlation were significant in Study 1 (p = 0.002) and Study 3 (p = 0.008), but not in Study 2 (p = 0.168).

TABLE 5.15

SUMMARY OF COMPARISON AMONG THE THREE SED MEASURES IN TERMS OF THEIR PREDICTABILITIES OF THE N-BACK LEVEL

		Study 1	Study 2	Study 3
		Fixation task	Random dot task	Free-viewing task
	# of participants	11	12	14
Average of r in all participants	r of MWL with microsaccade measure	-0.03	-0.25	-0.34
	r of MWL with the SI measure	0.57	0.37	0.40
	r of MWL with gaze aversion measure	0.25	0.24	0.17

F-statistics		10.16	5.85	21.31
p-value		0.001	0.010	0.001

Follow-up test	SI measure vs. gaze aversion measure			
T-statistics		4.04	1.47	3.10
p-value		0.002	0.168	0.008

A summary of the comparison among the three types of saccadic eye deviations (SEDs): the microsaccade measure, the saccadic intrusion (SI) measure, and the gaze aversion measure. The correlation coefficients r between the N-back level and the eye movement were shown in this table. Among these three similar eye movements, the SI measure had the strongest link to the N level. This supports the hypothesis that the N-back task induced SIs, but not other eye movements.

These results show that generally the manipulation of the N levels in the experiments were strongly linked to the SI measure, rather than to the microsaccade measure or to the gaze aversion measure. These results suggest that the N level manipulation was reflected not only in the SI measure but also in the number of Sis, rather than in the similar eye movements such as microsaccades and gaze aversion.

5.6.5 Comparison Among the Three Frequency Bands of Fourier Transformation

Another way to identify SIs is to compare them according to their frequency composition. The Fourier transformation (FT) frequencies were categorized into three groups; low, middle, and high frequency bands. SIs are believed to be reflected in the middle band FT (see the Algorithm chapter). If the middle band was more associated with SIs than the other two bands, then the middle band should show a stronger correlation with MWL (i.e., the N-back levels) than the other two bands. F-tests and t-tests were conducted to compare all participants' correlation (r) values in each study (Table 5.16). The three correlations tested were between the N levels and the low, medium, and high band FTs.

TABLE 5.16

SUMMARY OF COMPARISON OF THREE FOURIER TRANSFORMATION MEASURES

IN TERMS OF THEIR PREDICTABILITIES OF THE N-BACK LEVEL

		Study 1 Fixation task	Study 2 Random dot task	Study 3 Free-viewing task
	# of participants	11	12	14
Average r across all participants	Low band FT	0.296	0.126	0.24
	Mid band FT	0.446	0.327	0.27
	High band FT	0.267	0.361	0.20
F-statistics		2.85	4.12	0.33
p-value		0.03	0.03	0.72
Follow-up test				
	T-statistics (Low vs Mid)	1.82	1.99	0.37
	p-value	0.10	0.07	0.72
	T-statistics (Mid vs High)	3.04	0.82	0.93
	p-value	0.01	0.43	0.37
	T-statistics (Low vs High)	0.30	2.21	0.40
	p-value	0.77	0.05	0.69

A comparison of three frequency bands of Fourier transformation power density hypothesized to correlate with mental workload. The hypothesis that the middle band FT would correlate with mental workload (i.e., the N-back levels) seems to be true only in Study 1.

Since the middle band FT was associated with SIs, the author hypothesized that the middle band FT would have stronger correlations with N-back levels than the other two bands. This seems to be true only in Study 1. The middle band correlation average was 0.45, which was higher than either the low band correlation of 0.30 or the high band correlation of 0.27. The F-test was significant in Study 1 (p=0.03). Follow-up t-tests were significant (p = 0.01) between the middle band and the high band, but not significant (p=0.10) between the low band and the middle band.

While not significant in the F-tests in Study 2 and Study 3, the results in Study 1 are promising and future research should focus on refining the FT based analysis. The author suspects that the results in Study 1 were consistent with the author's hypothesis because the eye movements were controlled more in Study 1, yielding more fixational eye movements. One possible explanation for the non-significant results in Studies 2 and 3 could be the crudeness of the algorithm. The FT quantification algorithm was not as refined as the SI quantification algorithm in this dissertation. The SI quantification algorithm was specifically designed to single out only saccadic intrusions after removing regular saccades, microsaccades, gaze aversions, slow drifts, and tremors. The FT quantification algorithm did not remove any of these, therefore, the eye movements had more noise. The results of the FT analyses included the frequency components of regular saccades invoked by the random dot tasks in Study 2 and free eye movements in Study 3. As the other algorithm (Algorithm 1) had been refined over many revisions, the FT results might similarly be refined by reducing these kinds of noise from other eye movements that are not SIs.

5.7 Chapter Summary

This chapter examined the relationship between the N-back levels (serving as MWL levels) and seven physiological measures statistically. The MWL levels seemed to be strongly related with pupil diameter and the saccadic intrusion (SI) measure. Both pupil diameter and the SI measure were not just strongly related to the MWL levels, but also consistently related to the MWL levels across the 37 participants. The next chapter (Discussion Chapter) discusses the possible reasons that these two measures were successful and the other four measures were not successful and also some future ideas to improve the studies.

CHAPTER 6

DISCUSSION

6.1 Summary of Findings

As hypothesized, the results show that two physiological measures – the saccadic intrusion (SI) measure and pupil size – increased as the N-back task demanded higher working memory load. Out of the seven linear regression analyses conducted on the seven physiological measures, these two factors were best at estimating the N-back level. Both measures had the highest correlation with mental workload (MWL; see the shaded area in Table 6.1 below) and the highest generalizability; the percentages of participants who had a positive correlation coefficient between the N-back levels and the physiological measures were the highest (92 % or higher) for these two measures (Table 6.1). This indicates that each of the SI measure and the pupil measure could accurately and reliably estimates an operator's MWL.

TABLE 6

SUMMARY OF THE SEVEN PHYSIOLOGICAL MEASURES TO ESTIMATE MWL

		Predictability			Generalizability	
	Predictors	R^2			# of participants with positive correlation	Usage in Multiple regression
		Study 1	Study 2	Study 3	Three studies	
		Fixation Task	Random Dot Task	Free-viewing Task		
Physiological responses	Pupil Diameter	0.46	0.43	0.48	36 / 37 (97%)	Yes
	SI measure (medium amplitude SED)	0.38	0.33	0.30	34 / 37 (92%)	Yes
	Microsaccade measure (low amplitude SED)	0.24	0.31	0.28	16 / 37 (43%)	No
	Gaze aversion measure (high amplitude SED)	0.18	0.21	0.19	27 / 37 (73%)	No
	FT (Low band)	0.21	0.22	0.24	30 / 37 (81%)	No
	FT (Mid band), SI related	0.32	0.23	0.28	28 / 37 (76%)	No
	FT (High band)	0.20	0.27	0.23	31 / 37 (84%)	No

A summary of the predictability (middle three columns) and the generalizability (the second column from right). The R^2 values show how accurately the N-back levels were predicted using each of the seven single predictors and the participants' identification number as a categorical variable. The seven physiological responses were: pupil diameter, three categories of saccadic eye deviation (SED) (such as the saccadic intrusion (SI) measure, the microsaccade measure, and the gaze aversion measure), and three different frequency bands derived from Fourier transformation (FT) power density (i.e., low band FT and the high band FT). The pupil and SI measures had the highest correlation with N-back levels.

This research makes a number of valuable contributions to science. First, a novel algorithm was developed to quantify saccadic intrusions into a SI measure. The SI algorithm was preprogrammed to examine nine characteristics of SIs in the two-dimensional eye movement data (Section 2.3.8). Since the development of the algorithm was based on the past eye movement research data, the algorithm is theoretically grounded. It works with a low resolution eye tracker (like Tobii), and can detect SIs even when there are eye movements larger than fixation (Studies 2 and 3).

Second, the research findings showed the relationship between SIs and MWL. The results showed that 18 out of 37 participants had a monotonically increasing trend of the SI measure as

MWL increased. This number was statistically much higher than the predicted chance rate (i.e., 1/6). At a basic science level, this finding may help to reveal the relationship between eye movements (the control of saccades and fixation) and brain activity related to attention and cognitive capacities. At an applied science level, this finding may lead to a future applications in estimating MWL using eye movements. In the future, an automatic system may be able to estimate MWL in a driving environment, and may prevent possible car accidents.

Third, the result offered evidence that Fourier Transformation (FT) analyses may be a useful alternative method for quantifying the effects of task demands on MWL. The results of the trends tests showed that the number of participants who had the increasing trend (L<M<H) was 21 for the mid band FT, which was the highest among the seven physiological measures. Although the FT analyses were crude and exploratory, the high consistency (the fact that 21 out of 37 had the increasing trend in Figure 5.4.9) shows its potential to contribute to estimating MWL in the future.

6.2 Strengths and Weaknesses of the Research

The strengths and weaknesses of the research in terms of its operationalization, the MWL task, algorithms, equipments, and the relationship between SIs and MWL are discussed below.

6.2.1 Operationalization 1. The N-back task is not exactly the same as MWL

The N-back task was used to manipulate MWL in this dissertation. The results showed that participants seemed to experience three distinct levels of MWL as reflected in correct responses, reaction time, subjective rating, and the NASA-TLX (see the section 3 “Confirmation of MWL manipulation” in the Results chapter). It seems clear that the N-back task was effective in manipulating the MWL levels in the experiments

However, the N-back task may not be perfectly equivalent to MWL. There may be some specific aspects of the N-back task that is not found in other MWL tasks, and that aspect might also affected SIs. For example, Owen and colleagues (2005) did a meta-analysis of 24 N-back studies. They identified six areas of the brain that were related to the N-back task, as well as some other areas that were related only to the verbal N-back task and other areas that were related to non-verbal (shape N-back and location N-back) tasks. This suggests that the verbal N-back task used in this dissertation may not simulate all kinds of MWL. Maybe the brain areas responsible for verbal N-back tasks invoked the SIs. This dissertation focused on the eye movements, and did not examine a variety of cognitive tasks that might relate to SIs. It is possible that another task that engages or demands different cognitive processes might be more effective in inducing SI. SIs were induced during the N-back task engagement perhaps because of the requirement to retrieve or maintain information in short term memory. Other MWL tasks will need to be tested in the future to provide further support for the hypothesized link between MWL in general (i.e., all kinds of cognitive task) and SIs.

This dissertation examined the relationship between the SI measure and just one kind of MWL task. The other well-accepted MWL measure is the pupil measure which has been shown to be related to the cognitive demands associated with performance of a variety of cognitive tasks including digit span (Kahneman, 1967), tone comparison (Kahneman & Beatty, 1967), word memorization (Elshtain & Schaefer, 1968), and arithmetic (Bradshaw, 1968). More experiments will be needed to confirm the link between MWL and SIs.

6.2.2 Strengths of the SI Measure of MWL

The SI measure of MWL offers several advantages. Since SIs are eye movement, they can be monitored externally thus not requiring that the operator wear anything that might interfere with their normal behavior. SIs are also relatively easy to observe because their amplitudes are much bigger than the amplitudes of the other types of fixational eye movements. This means that a lower resolution eye tracker (such as Tobii) can detect SIs.

6.2.3 Instrumentation Limitations.

The Tobii eye trackers' temporal resolution is (50 Hz, polling data every 20 ms), and spatial resolution 0.25 deg) are relatively low which can limit the experimenters ability to detect SIs. While the results show that SI can be detected using a inexpensive eye trackers the reliability this measure may be improved by using an eye tracker with superior spatial and temporal resolution.

6.2.4 Dwell Time in Quantifying Saccadic Intrusions

One of the weaknesses of the algorithm was the inability to detect the dwell time of SIs. The current algorithm could continuously accumulate eye deviations from a fixation baseline. However, since microsaccades and SIs typically occur once in a few hundred milliseconds, the ideal algorithm probably should have the capability to detect SI episodes discretely one by one, not continuously and not cumulatively.

6.2.5 Free Eye Movements in Quantifying Saccadic Intrusions

Another strength of this dissertation is the robustness of the SI algorithm. This dissertation detected SIs in a fixation task (Study 1), a random dot task (Study 2), and a free-viewing task (Study 3), and the algorithm was robust detecting SIs under different experimental paradigms. The algorithm detected each of the fixational periods and calculated eye deviations within each of the fixational periods. This is a big step from the past research by Abadi and Gowen (2004), Engbert (2006), and McGivern and Gibson (2006), where they used only a fixation task. However, the current algorithm can be improved. It makes errors occasionally, especially when detecting regular saccades. The algorithm also occasionally “misses” regular saccades occasionally. Further enhancements in the algorithm should improve the ability to estimate MWL using SI measure proposed in this dissertation.

Some types of eye movements were intentionally constrained so simplify the analyses These restrained types of eye movements (such as VOR, OKR, smooth pursuits) will need to be included in future studies.

6.2.6 Better Prediction Using Multiple Predictors

Since the SI measure and pupil diameter rely on different sources (i.e., eye movements vs. pupil size), the two measures can separately contribute to estimating MWL. Moreover, estimation of MWL was more accurate by using a combination of the two measures than each alone, as shown in the Section 5.5 (“Predicting MWL Using Multiple Predictors”).

6.2.7 Low Noise Using Multiple Predictors

Another potential strength of the multiple predictor method is that it may produce the overall noise in the measurements. Physiological responses tend to be noisy. For example, when MWL is estimated using pupil diameter, illumination changes can be a source of measurement noise since they will affect the size of the pupil. Likewise, large eye movements are a source of noise in a SI measure when MWL is estimated through a SI measure alone. The noise in one measure is expected not to affect the other measure and may even reduce overall noise if data from the two measures are relatively uncorrelated as pupil size and eye movement patterns should be. An algorithm that uses a combination of measures may prove to be more reliable than one used in experiments performed in this dissertation.

6.2.8 Scope of Each Predictor

Another possible strength of the multiple predictor method is that the two predictors may have different ranges for reflecting MWL. Figure 5.11 shows the the pupil diameter measure saturated after the medium MWL level; pupil diameter increased from the low to medium MWL levels, but did not increase much from the medium to high MWL levels. Therefore, pupil diameter may be better at reflecting MWL at lighter levels. Since this ceiling effect was not observed in the SI measure (Figure 5.16), it might be better at reflecting MWL differences in higher ranges.

6.3 Neurological Explanation

The third item on the strengths and weaknesses is the neurological explanation. This is part of discussion connecting MWL and SIs. The offer evidence of a relationship between MWL as induced by the N-back and the frequency of SIs. The specific neurological mechanism responsible for the increase in SEDs (including SIs and microsaccades) as cognitive demand increases is unknown. However, several research teams that have that extensively examined the relationship between attention and SEDs (such as microsaccades and SIs). have suggested that SEDs originate in the Superior Colliculus The Engbert lab did a series of studies on microsaccades (one kind of SED) and attention (e.g., Engbert, 2006; Engbert and Kliegl, 2003; Engbert and Kliegl, 2004; Laubrock, Engbert and Kliegl, 2005; Rolfs, Engbert and Kliegl, 2005). They found microsaccades increased their occurrence and amplitude about 400 ms after attentional demand. The Abadi lab also did a series of studies on SIs (one kind of SED) and attention and found similar results (e.g., Gowen, Abadi and Poliakoff, 2005; Abadi and Gowen, 2004; Gowen, Abadi, Poliakoff, et al., 2007; Gowen and Abadi, 2005). Both teams suggest that the neural mechanism accounting for the link between SEDs and attention is fixation neurons in the superior colliculus. Although there is a difference between attention and MWL in their studies and in this dissertation, this dissertation endorses their explanations based on fixation neurons in the superior colliculus. Additionally, this dissertation borrows an adjunct idea on attention from another review paper by Hutton (2008).

Hutton (2008) cites a study by Pratt, Lajonchere and Abrams (2006) which offers evidence that maintaining target fixation requires attention. Pratt et al. used gap trials in which participants were instructed first to fixate on a center target, and then fixate on a peripheral target when it appeared. There were several conditions. In the No-Gap condition, there was no temporal gap between the offset of the center target and the onset of the peripheral target. In the Gap condition, there was a 200 ms temporal gap between them. Pratt et al. compared latency time of the saccade to the peripheral target between the conditions, and found that the Gap condition had faster latency time than the No-Gap condition. Hutton (2008) reviews that the time gap disengaged the attention to the center target, so that the recharged attention could engage in the peripheral target faster. Pratt et al. (2006) also used three other conditions that would demand different levels of attention, and found saccade latency times increased as a function of the imposed attention demands.

Pratt et al. (2006) attributed their results to fixation neurons in the superior colliculus. When fixation neurons in the superior colliculus are more active, fixation is more stable, but this seems to consume attention resources. In other words, when attention resources are used for an attentional task or a MWL task, fixation neurons in the superior colliculus may not have adequate access to the resources for self regulation. Therefore, fixation can become unstable. Hutton’s (2008) view has not been proven neurologically, however, it seems quite plausible. This view could explain the instability of fixation, and the weak fixation allowed SEDs to intrude in the fixation when MWL was high.

6.4 Chapter Conclusions

The three main contributions of this dissertation to the sciences are the development of an algorithm to quantify saccadic intrusions (SIs), the demonstration of the relationship between SIs and mental workload (MWL), and the exploratory demonstration of the usage of Fourier Transformation (FT) to analyze SIs. The SI quantification algorithm was designed to identify SIs not only when the eye gaze fixates at one location, but also when the eye freely moves. Future developments of the algorithm should focus on improve the algorithm's ability to accurately distinguish regular saccades from other types of eye movements.

CHAPTER 7

CONCLUSIONS

The purpose of this dissertation was to examine if mental workload (MWL) caused changes in physiological eye behavior – such as saccadic intrusions (SIs) and pupil diameter – in a laboratory setting. In order to operationalize this research goal, SI behavior was quantified into a numerical value by developing a novel algorithm. The author first listed the characteristics of SI behavior (see Section 3.8, “The term Saccadic Intrusions” in Chapter 2 “Literature Review”) based on the past eye movement research (Abadi and Gowen, 2004; Engbert, 2006; Carpenter, 1988), then developed a new SI algorithm to detect all of these characteristics in time-series eye movement data and quantify them into a single value during a time period (see Chapter 3, “Algorithm”).

The experiments tested both the cause-effect of the MWL-and-SI relationship and the effectiveness of the new SI algorithm for quantifying SI behavior. While MWL was manipulated by using the auditory N-back task, eye activities were recorded. The analyses show that as MWL increased, so did pupil size and the SI measure. Both of the relations (the MWL-pupil relation and the MWL-SI relation) were strong enough to have the average individual correlations 0.57 and 0.45, respectively, out of 37 participants in the three studies (see Sections 4.1, “Pupil diameter and MWL” and 4.2, “Saccadic Intrusions and MWL” in Chapter 5, “Results”). These results indicate that both pupil diameter and the SI measure were accurate indicators of MWL in laboratory conditions. Also, both relations were consistent enough to have 97% and 92% of participants with positive correlations. These results indicate that both measures were almost always producing the same directional changes when MWL was imposed; the pupil diameter almost always increased and SIs almost always increased. The high consistency in both measures suggest that these measures may be generalizable to the rest of the population.

When comparing the ability of the pupil diameter and SI measures to predict MWL, pupil diameter had an average correlation 0.57 while the SI measure had an average correlation of 0.45. However, the predictabilities were essentially the same in Study 1 (r = 0.58 for pupil diameter and r = 0.57 for the SI measure) where the illumination was stable and there were no large eye movements (see Section 6.2, “Correlation Coefficient Values” in the Results chapter). This indicates that the SI measure has the potential to estimate MWL as accurately as pupil diameter does in controlled conditions. Additionally, eye movements seem to provide much more information than pupil diameter does (see Section 4.3, “Pros and Cons of Measuring Pupil Diameter” in the Literature Review). The Tobii eye tracker used in these experiments has a fairly low resolution, which meant that not all of the information from the eye movements were available for analysis. Nonetheless, the novel SI detection algorithm yielded data that was robust enough to have a significant correlation with MWL. The use of SI measures may also be beneficial because unlike pupil diameter, SI measures have the leeway to be refined by using a higher resolution eye tracker and by improving the SI detection algorithm. That refinement may increase the SI measure’s ability to estimate MWL.

The contribution of this dissertation was twofold. First, it proposes a SI detection algorithm to estimate MWL. Second, it demonstratesthe relationship between MWL and SIs . Since SIs consistently increased with increased MWL, SIs can be used as a reliable indicator of MWL. As of today, none of the pre-existing physiological measures that are observable from outside of the body (such as pupil diameter and blink rate) can accurately estimate MWL. Since the SI measure is shown to be about as accurate as pupil diameter, and has leeway to improve its MWL predictability with today's technology, it is a potential candidate to become the first physiological measure that can accurately estimate MWL in   environments.

REFERENCES

LIST OF REFERENCES

Abadi, R. V., & Gowen, E. (2004). Characteristics of saccadic intrusions. *Vision research, 44*(23), 2675-2690.

Andreassi, J. L. (2000). *Psychophysiology: Human Behavior & Physiological Response* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*, 623-636.

Baddeley, A. D., & Hitch, G. J. L. (1974). Working Memory, In G.A. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 8, pp. 47-89), New York, NY: Academic Press.

Baker, L., Goldstein, R., & Stern, J. A. (1992). Saccadic eye movements in deception (DoDPI92-R-0003). Fort McClellan, AL: Department of Defense Polygraph Institute. DTIC# ADA304658.

Beatty, J., & Wagoner, B. L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science*, *199*, 1216–1218.

Boff, K. R., & Lincoln, J. E. (1988). *Engineering Data Compendium: Human Perception and Performance*. Sec. 1.235, AAMRL, Wright-Patterson AFB, OH.

Bradshaw, J. L. (1968). Pupillary changes and reaction time with varied stimulus uncertainty. *Psychonomic Science, 13*, 69-70.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Camilli, M., Terenzi, M., & Nocera, F. D. (2007). Concurrent validity of an ocular measure of mental workload. In D. de Waard, G.R.J. Hockey, P. Nickel, and K.A. Brookhuis (Eds.), *Human Factors Issues in Complex System Performance* (pp. 117 - 129). Maastricht, the Netherlands: Shaker Publishing.

Carpenter, R. H. S. (1988). *Movements of the Eyes* (2nd ed.), London, UK: Pion Press.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3):6, 1-15.

Chapman, L. J., Chapman, J. P., & Brelje, T. (1969). Influence of the experimenter on papillary dilation to sexually provocative pictures. *Journal of Abnormal Psychology, 74,* 396-400.

Chapman, C. R., Oka, S., Bradshaw D. H., Jacobson, R. C., & Donaldson, G. W. (1999). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain evoked potentials and pain report. *Psychophysiology*, *36*, 44-52.

Clementz, B. A., Sweeney, J. A., Hirt, M., & Haas, G. (1990). Pursuit gain and saccadic intrusions in first-degree relatives of probands with schizophrenia. *Journal of Abnormal Psychology, 99*(4), 327-35.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2003). Development of gaze aversion as disengagement from visual information. *Developmental Psychology, 38*(3), 438-445.

Elshtain, E. L., & Schaefer, T. (1968). Effects of storage load and word frequency on papillary responses during short term memory. *Psychonomic Science, 12*, 143-144.

Engbert, R. (2006). Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. *Progress in Brain Research, 154*. 177-192.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*, 1035-1045.

Findlay, J. M. (1971). Frequency analysis of human involuntary eye movement. *Kybernetik, 8,* 207-214.

Geacintov, T., & Peavler, W. (1974). Pupillography in industrial fatigue assessment. *Journal of Applied Psychology, 59,* 213-216.

Gopher, D., and Donchin, E. (1986). "Workload: An examination of the concept". In K. Boff, and L. Kaufman, (Eds.): *Handbook of Human Perception and Performance,* NewYork, NY: John Wiley, Vol II, 41, 1-49.

Gowen, E., Abadi, R. V., & Poliakoff, E. (2005). Paying attention to saccadic intrusions. *Cognitive Brain Research, 25*(3), 810-825.

Gowen, E., Abadi, R. V., Poliakoff, E., Hansen, P., & Miall, R. C. (2007). Modulation of saccadic intrusions by exogenous and endogenous attention. *Cognitive Brain Research, 1141*: 154-67.

Guyton, A. C. (1977). *Basic human physiology: Normal function and mechanisms of disease.* Philadelphia, PA: Saunders Co.

Hancock, P. A., Lesch, M., & Simmons, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis and Prevention, 35*. 501-514.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors, 35,* 579-601.

Hutton, S. B. (2008). Cognitive control of saccadic eye movements. *Brain and Cognition, 68*, 327-340.

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive Affective, & Behavioral Neuroscience*, *7*(2), 75-89.

Johnson. A. & Proctor, R. W. (2004). *Attention: Theory and Practice*. Thousand Oaks <<state>>: Sage Publications.

Just , M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology, 47*, 310-339.

Just, M. A., Carpenter, P.A., & Miyake, A. (2003). Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics, 4*(1-2), 56-88.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D & Beatty, J. (1967). Pupillary responses in a pitch discrimination task. *Perception and Psychophysics, 2,* 101-105.

Kahneman, D. & Peavler, W. S. (1969). Incentive effects and papillary changes in association learning. *Journal of Experimental Psychology, 79*, 312-318.

Krauskopf, J., Graf, V., & Gaarder, K. (1965). Lack of inhibition during involuntary saccades. *American Journal of Psychology, 79*(1), 73-81.

Krauzlis, R. J. (2005). The control of voluntary eye movements: New perspectives. *Neuroscientist*, *11,* 124–137.

Lasker, A. G., Mazzocco, M. M. M., & Zee, D. S. (2007). Ocular motor indicators of executive dysfunction in fragile X and Turner syndromes. *Brain and Cognition, 63*, 203–220.

Laubrock, J., Engbert, R. & Kliegl, R. (2005). Microsaccade dynamics during covert attention. *Vision Research, 45,* 721-730.

Laubrock, J., Engbert, R., Rolfs, M. & Kliegl, R. (2007). Microsaccades are an index of covert attention. *Psychological Science. 18*(4). 364-366.

Leigh, R. J., & Kennard, C. (2004). Using saccades as a research tool in the clinical neurosciences. *Brain*, *127*, 460-477.

Leigh, R.J., & Zee, D.S. (2006). *The Neurology of Eye Movements*. New York, NY: Oxford University Press.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., et al. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (Tech. Rep. 851). Fort Bliss, TX: U.S. Army Research Institute, Field Unit.

Martinez-Conde, S. (2006). Fixational eye movements in normal and pathological vision. *Progress in Brain Research, 154*. 151-176.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory & Cognition, 27,* 817-835.

McGivern, R. C., & Gibson, J. M. (2006). Characterization of ocular fixation in humans by analysis of saccadic intrusions and fixation periods: A pragmatic approach. *Vision Research, 46,* 3741-3747.

Murray, S .J., McFarland, K. & Geffen, G. (2005). The effects of task complexity and practice on dual-task interference in visuospatial working memory. *Brain Impairment, 6*(1), 13-20.

Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers, 47,* 617-644.

Obinata, G., Tokuda, S., Fukuda, K. (2009). Quantitative evaluation of mental workload by using model of involuntary eye movement. *Engineering Psychology and Cognitive Ergonomics, 5639,* 223-232.

Obinata, G., Tokuda, S., & Shibata, N. (2008). Mental workloads can be objectively quantified in real-time using VOR (vestibulo-ocular reflex). *Proceedings of the 17th International Federation of Automatic Control World Congress,* 15094-15099.

O'Donnell, R. D., & Eggemeier, F. T. (1986). "Workload Assessment Methodology," (42/1-42/9). In K. Boff, L. Kaufman and J. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol. II: Cognitive Processes and Performance*. New York, NY: Wiley Interscience.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-Back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25,* 46-59.

Papousek, I., & Schulter, G. (1999). Quantitative assessment of five behavioral laterality measures: Distributions of scores and intercorrelations among right-handers. *Laterality, 4* (4), 345-362.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision 10*: 437-442.

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78,* 391-408.

Pratt, J., Lajonchere, C. M., & Abrams, R. A. (2006). Attentional modulation of the gap effect. *Vision Research, 46,* 2602–2607.

Proctor, R. W., & Van Zandt, T. (1994). *Human Factors in Simple and Complex Systems*. Boston, MA: Allyn and Bacon.

Ratliff, F., & Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology, 40,* 687-701.

Recarte, M. A., & Nunes, L. M. (2002). Mental load and loss of control over speed in real driving. Towards a theory of attentional speed control. *Transportation Research, 5,* 111–122.

Recarte, M. A., & Nunes, L. M. (2003). Mental Workload While Driving: Effects on Visual Search, Discrimination and Decision Making. *Journal of Experimental Psychology: Applied, 9*(2), 119–137.

Rayner, K. (1984). Visual selection in reading, picture perception, and visual search: A tutorial review. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance (Vol.10)*. Hillsdale, NJ: Erlbaum.

Rice, R. E. & Stohl, C. (2006). Communication and human factors. *Handbook of Human Factors and Ergonomics.* (G. Salvendy, ed.). Wiley, Hoboken, New Jersery. Chapter 6. 150-176.

Robinson, D. A. (1963). A method of measuring eye movements using a search coil in a magnetic field. *IEEE Transaction of Biomedical Electronics, 10*, 137-145.

Robinson, D. A. (1981). The use of control system analysis in the neurophysiology of eye movements. *Annual review of neuroscience, 4,* 463-503.

Robinson, D. A., Gordon, J.L., & Gordon, S.E. (1986). A model of the smooth pursuit eye movement system. *Biological Cybernetics, 55* 43-57.

Rolfs, M., Engbert, R., & Kliegl, R. (2005). Crossmodal coupling of oculomotor control and spatial attention in vision and audition. *Experimental Brain Research, 166,* 427–439.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W.T. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision, 77*(1-3), 157-173.

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology, 45,* 679–687.

Snowden, Thompson, & Troscianko, 2006). *Basic Vision: An Introduction to Visual Perception.* New York: Oxford.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods: A Practical Guide for Engineering and Design.* Burlington, VT: Ashgate.

Stone, L. S., Beutter, B. R., & Lorenceau, J. (2000). Visual motion integration for perception and pursuit. *Perception, 29*, 771-787.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23-32.

Taguchi, G., Chowdhury, S., & Taguchi, S. (2000). *Robust Engineering.* New Jersey: McGraw-Hill.

Tokuda, S. (2008). Using saccadic eye movements to measure mental workload. *Journal of Vision. 8* (17). 86a.

Tokuda, S., Morris, N., & Chaparro, A. (2007). The effects of visual-spatial and verbal working memory load on driving performance. *Journal of Vision 7*(9): 683, 683a.

Tokuda, S., Obinata, G., Usui, T., Inuzuka, Y. & Hamada, H. (2009). Measuring the driver's mental workload in the peripheral vision area using vestibulo-ocular reflex. *Transactions of Society of Automotive Engineers of Japan, 40*(1). 205-211.

Tsang, S. P., & Vidulich, M. A.(2006). Mental workload and situation awareness. In G. Salvendy, (Ed.) *Handbook of Human Factors and Ergonomics*. Wiley, Hoboken, New Jersery. Chapter 9. 243-268.

Van Orden, K. F., Limbert, W., Makeig, S., Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*, 111-21.

Wickens, C. D. (1984). Processing Resources in Attention, In Raja Parasuraman and D.R. Davies, (Eds.) *Varieties of Attention*, New York: Academic Press, 63.102.

Wickens, C. D. (1991). Processing resources and attention. In D. Damos (Ed.), *Multiple-task performance* (pp. 3-31). London: Taylor & Francis.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). New Jersey: Prentice Hall.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Wickens, C.D. (2004). Multiple resource time sharing model. In N.A. Stanton, E. Salas, H.W., Hendrick, A. Hedge, & K. Brookhuis (Eds.) *Handbook of Human Factors and Ergonomics Methods.* (40-1-40-7). London: Taylor & Francis.

Wickens, C. D. (2007). Attention to Attention and Its Applications: A Concluding View. In A.E. Kramer, D.A. Wiegmann, and A. Kirlik (Eds.) *Attention: From Theory to Practice*. London: Oxford.

Zee, D. S., Yee, R. D., Cogan, D. G., Robinson, D. A., & Engel, W. K. (1976). Ocular motor abnormalities in hereditary cerebellar ataxia. *Brain*, *99*(2), 207–234.

APPENDICIES

Appendix A
TABLE A.1.

DEMOGRAPHIC DATA

Time Order	Study#	Data used	Reason to discard	pt ID	Gender	Age	Handed-ness	Correction	Heavy MWL
17	2	Used		37	Male	24	Right	Contact	3
18	2	Used		38	Female	18	Right	None	3
19	2	Used		39	Female	22	Right	None	3
20	2	Used		40	Female	21	Right	Contact	3
21	2	Used		41	Female	18	Right	Contact	3
22	2	Used		42	Male	27	Right	Contact	3
23	2	Used		43	Male	18	Right	None	4
24	2	Used		44	Female	18	Right	None	3
25	1	Used		21	Female	18	Left	None	4
26	1	Used		22	Female	20	Right	None	3
27	Discarded	Not used	Tobii	N/A	Female	19	Right	None	3
28	1	Used		23	Female	18	Right	None	3
29	1	Used		24	Male	20	Right	None	3
30	1	Used		25	Female	21	Right	None	3
31	1	Used		26	Female	18	Right	None	3
32	1	Used		27	Female	18	Right	Contact	4
33	1	Used		28	Female	55	Right	None	3
34	Discarded	Not used	N-back	N/A	Male	30	Right	None	3
35	1	Used		29	Female	19	Left	None	3
36	1	Used		30	Female	18	Right	Contact	3
37	Discarded	Used	Tobii	32	Female	25	Left	None	3
38	2	Used		45	Female	18	Right	Contact	3
39	2	Used		46	Female	20	Right	None	3
40	2	Used		47	Female	19	Left	None	3
41	2	Used		48	Female	18	Right	Contact	3
42	1	Used		31	Female	19	Right	None	3
43	3	Used		49	Female	18	Right	Contact	3
44	3	Used		50	Female	18	Right	None	4
45	3	Used		51	Male	19	Right	None	3
46	3	Used		52	Female	19	Left	None	3
47	3	Used		53	Male	19	Right	Contact	3
48	3	Used		54	Female	18	Right	None	3
49	3	Used		55	Female	18	Left	Contact	4
50	3	Used		56	Female	18	Right	Contact	3
51	3	Used		57	Female	18	Right	None	4
52	Discarded	Not used	N-back	N/A	Male	32	Right	None	3
53	3	Used		58	Male	20	Right	None	3
54	Discarded	Not used	Premise	N/A	Female	18	Right	None	3
55	3	Used		59	Female	21	Right	None	3
56	3	Used		60	Male	19	Left	Contact	3
57	3	Used		61	Female	18	Left	Glasses	4
58	3	Used		62	Female	19	Right	Contact	3

Demographic data of 37 participants. The 12 columns consist of the chronological order in experiments, the study number, usage in the analyses, the reason to discard the participant's data, the participant ID number in this dissertation, the ethnicity of the participant (Af-Am = African American, So-Am = South-American), gender, age, the dominance of the hand usage, vision correction methods in the experiments, and the highest N level in the N-back task. A total of five out of 37 participants were removed from the analyses (indicated in shades) for the reasons written in the fourth column. Read the Section 1 "Participants" in Chapter 4 "Methods."

Appendix B
TABLE A.2

DEMOGRAPHIC SURVEY

Name	Participant #
Age	
Eye Correction None Glasses Contact Lens	
Handedness Left Right	
How many hours did you sleep last night?	
What time did you wake up today?	
Do you have any vision problem now?	

Experiment Date
Experiment Time
Experiment End Time
Experimenter Name

Table A.2. A questionnaire used for learning the participants' eye condition.