PHYSIOLOGICAL INDICATORS OF COGNITIVE WORKLOAD AS DETECTORS OF PERFORMANCE REDLINES

A Dissertation

by

CAROLINA RODRIGUEZ PARAS

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Thomas K. Ferris
Co-Chair of Committee,	Farzan Sasangohar
Committee Members,	Nancy J. Currie-Gregg
	Louis G. Tassinary
Head of Department,	Lewis Ntaimo

August 2021

Major Subject: Industrial Engineering

ABSTRACT

High cognitive workload in multitasking, safety-critical environments can lead to performance degradation, resulting in increased safety risks or errors. The problem arises when the person reaches the redline of cognitive workload, which occurs when the task demands exceed the available cognitive resources necessary to continue performing the tasks at an adequate performance level. While the redline is different for every person, an ability to understand how to detect the threshold can provide input for strategies in mental resource allocation to prevent performance degradation.

Historically, the redline of cognitive workload has been measured using subjective and performance measures. Subjective measures provide insight into the human's perceived mental workload level, and may be collected concurrently with the task, but this may require additional cognitive resources. When subjective measures are collected after the task, memory decay may occur. Performance is usually observed after the task, and when workload reaches a certain level, any additional increases in cognitive workload can lead to degraded performance. Physiological measures have also been used to measure cognitive workload, but physiological patterns have yet to be investigated as indicators of the redline of cognitive workload.

Physiological measures, such as heart rate, heart rate variability, and skin conductance reflect changes occurring in both branches of the autonomic nervous system: the sympathetic and parasympathetic. They can be collected in real-time through unobtrusive, wearable devices. The current research aims to gather a better

understanding of how high levels of cognitive workload are reflected in physiological measures, and can be used to infer when the person is approaching their redline of cognitive workload.

This research contributes to the body of knowledge by providing a better understanding of the redline of cognitive workload, which can be detected based physiological patterns. The complex relationship between both branches of the autonomic nervous system (sympathetic and parasympathetic) were explored, and used to investigate when the person may be approaching their redline of cognitive workload.

DEDICATION

To my parents and brother

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Ferris, and committee co-chair, Dr. Sasangohar, for giving me the opportunity to do research in multiple projects, and for all their teachings. I would like to extend my sincere gratitude to my committee, Dr. Ferris, Dr. Sasangohar, Dr. Currie-Gregg, and Dr. Tassinary, for their guidance, continuous support, and encouragement through the course of this research.

I would also like to thank the lab members from the Human Factors and Cognitive Systems (HF&CS) lab and the Applied Cognitive Ergonomics (ACE) Lab. Also, many thanks to all the volunteers that participated in the three studies.

I would also like to extend my sincere gratitude to my family, for all their love, encouragement, and support through the years.

CONTRIBUTORS AND FUNDING SOURCES

The studies presented in Chapters 2 and 3 were designed, conducted, analyzed, and completed by me. Harrison Wissel-Littmann helped with the data collection for the third study. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Funding for my research has been provided by a graduate teaching fellowship from the Texas A&M College of Engineering, and by research and teaching assistantship from the Department of Industrial and Systems Engineering at Texas A&M University.

NOMENCLATURE

ANS	Autonomic Nervous System
PSNS	Parasympathetic Nervous System
SNS	Sympathetic Nervous System
HR	Heart Rate
HRV	Heart Rate Variability
EEG	Electroencephalography
EDA	Electrodermal Activity
NASA-TLX	NASA Task Load Index
WPI	Workload Profile Index
SSSQ	Short Stress State Questionnaire

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Safety-critical work environments require certain performance levels to maintain safety standards. However, in volatile, uncertain, complex, and ambiguous (VUCA) environments (Bennett & Lemoine, 2014; Rouvrais et al., 2018), unexpected events can increase cognitive workload, leading to increased safety risks (Carayon & Alvarado, 2007; Endacott, 2012; Milczarek et al., 2007). Additionally, constantly high workload experienced in the workplace can lead to feeling overload, which can lead to burnout (Portoghese et al., 2014; Sweeney & Summers, 2002; Van Bogaert et al., 2013), costing an average of $125 billion to $190 billion dollars annually (Blanding, 2015; Garton, 2017).

When experiencing elevated cognitive workload, the real problem arises when the task demands overwhelm the cognitive resources of the person, which can lead to an increased safety risk and degraded overall performance. While some systems have some flexibility, most data-rich and complex environments need to maintain adequate performance levels to ensure safety. For example, physicians that work under high workload and stressful conditions also have an increased risk for occupational health hazards (van den Hombergh et al., 2005). Job stress in the oil and gas industry has been linked the personnel’s ability to avoid dangerous situations, and may increase accidents (Rundmo et al., 1998). In aviation, high workload and aircraft go-around may lead to losing state awareness, which has led to aviation accident (Schmidt et al., 2021). High cognitive workload can be experienced in most work domains, which can increase the risk of errors and lead to possible accidents.

Historically, cognitive workload has been measured through subjective and performance-based measures. However, these measures are usually collected after the task is completed. Otherwise, if collected simultaneously, there is a risk that the task will be interrupted. More recently, physiological measures have also been implemented as a way to monitor cognitive overload, which can be collected in real-time with wearable sensors and minimal task disruption. Monitoring physiological patterns associated with high levels of cognitive workload can provide information about the person's current workload to help to minimize performance decrements, thus maintaining safety.

The research described in this dissertation investigates physiological patterns indicative of autonomic activity related to when the person is approaching their cognitive redline threshold. The collective findings from this research contribute to a new definition of the redline of cognitive workload based on physiological data collected through experiments.

Additional contributions provide evidence to support the effectiveness of physiological measures to detect changes in workload when the person is approaching the limit of their cognitive resources based on analysis and comparison of the three types of data collected and investigated (i.e. physiological measures, subjective ratings, and performance). The findings providing evidence of physiological patterns could minimize performance loss associated with high levels of cognitive workload, especially when monitored through wearable devices, thus improving safety and operations.

Cognitive Workload

Several definitions exist for cognitive workload. For example, cognitive workload has been defined in terms of stress and strain, where stress is the is related to the mental load of the task demands or the peak load, and the strain is the impact of the demands on the person (Moray, 1979). More recently, cognitive workload has been defined in terms of the mental resources necessary to complete a task or a set of tasks (Wickens, 2002). More recently, workload has been described based on the task characteristics of the operator, and the task context (Young et al., 2015). These definitions imply a limit to mental capacity.

Several models have been proposed to frame the limitations of human mental resources, including bottleneck models and resource models of attention. Bottleneck models present a constriction in the way the information is processed (Pashler, 1984, 1990). For example, Broadbent's (1958) model explains humans process information with a selective filter. Treisman's (1964) model, a revised version of Broadbent's model, adds that unattended stimuli can be attenuated. More recently, other research has continued to add to bottleneck models (Driver, 2001; Lachter et al., 2004), but all models are consistent in having a bottleneck or constricting factor in how the information is processed.

Contrary to bottleneck models, resource models assume the constraint in information-processing is related to the amount of available mental resources. For example, Kahneman's (1973) unitary model consists of a single mental resources, and several tasks can be performed simultaneously as long as the total capacity is not exceeded. Wickens' (2008) multiple resource theory (MRT) consists of different resource pools that allocate resources. In this model, multitasking is possible when the task demands are allocated to different resource pools.

For the purpose of this research, workload is defined as the extent to which a limited set of mental resources are engaged to process information, strategize, and complete the tasks. A major objective of this research was to investigate the physiological patterns associated with cognitive overload.

Multitasking

Multitasking occurs when the person performs concurrent tasks, but oftentimes the tasks require sequential operations (Gutzwiller et al., 2014). Sequential operations have been described based on interruption management and task management, while concurrent task operations have been described based on multiple resources and threaded cognition.

Multitasking in sequential operations requires managing the interruptions or the tasks. Interruptions are defined as the introduction of new tasks on top of the main task (Alkahtani et al., 2020). Similarly, task management requires prioritizing one task on top of the others (Freed, 2000).

Concurrent multitasking requires that operations be performed simultaneously, usually explained based on the cognitive resources utilized. For example, the multiple resource theory model posits that people have different pools of attentional resources (Wickens, 2008). Under this model, multitasking can be effectively performed as long as the tasks requires resources from different pools. Similarly, threaded cognition explains that each task can be represented as a processing thread, which is then coordinated by a procedural resource (Salvucci & Taatgen, 2008). Thus, there is no need for executive processes for executing the tasks concurrently (Salvucci & Taatgen, 2008). The research presented in this dissertation investigates concurrent multitasking, particularly related to the multiple resource theory model (see Chapters 2, 3, and 4).

In multitasking environments, cognitive workload needs to be balanced, as cognitive underload and overload can possibly lead to performance degradation. The Yerkes-Dodson arousal curve, an inverted U-shape, describes the relationship between cognitive arousal and performance, where low and high arousal lead to low performance (Yerkes & Dodson, 1908). Previous studies have investigated underload (e.g.)(Yeh & Wickens, 1988; Young et al., 2011; Young & Stanton, 2002a, 2002b) and overload (e.g)(Fox et al., 2007; Jaeggi et al., 2007; Kirsh, 2000; Mackay, 2000). This research will focus on the overload problem, particularly when a person approaches the limit of their mental resources.

The Redline of Cognitive Workload Model

The problem arises when the person gets overloaded, either performing a single task or multitasking. With overload, the cognitive resources are near their limit, and the person has no other resources to maintain multitasking, reaching a performance redline.

The redline of cognitive workload has been described in terms of available mental resources to perform a task or set of tasks. Essentially, residual capacity is depleted as the cognitive resources approach their limit. Thus, at the redline, performance will decrease (Grier et al., 2008; Wickens, 2008; Young et al., 2015), see Figure 1. Performance loss can be minimized by ensuring proper workload levels (neither too high nor too low).

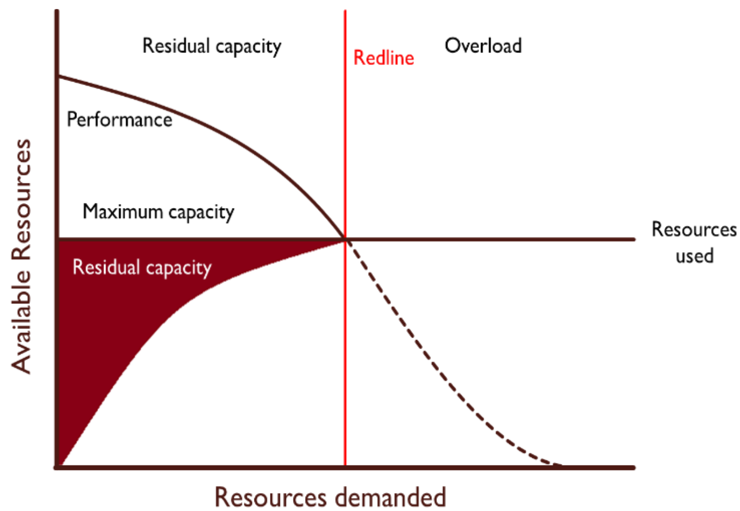


Figure 1: The redline of cognitive workload model (adapted from Wickens et al., 2015).

Previous studies have found performance redlines based on subjective, self-reported measures of cognitive workload. For example, a score in the range of 40 +/- 10 in the Subjective Workload Assessment Technique (SWAT) was correlated with performance decrements (Reid & Colle, 1988). Multitasking performance remained stable under the lower range, but performance was negatively impacted when the range score was exceeded. A similar study validated this range, the same score range, suggesting the limit may vary depending on the different tasks (Colle & Reid, 2005).

Another example of work done to determine the upper range of acceptable workload utilized the Improved Performance Research Integration Tool (IMPRINT). The workload of combat missions was modeled in IMPRINT, and a value score of 60 or more was categorized as critically-high mental workload (Mitchell et al., 2003).

While previous research has focused on performance and subjective self-reported measures to determine the redline of cognitive workload, few studies have investigated the redline based on physiological measures. These studies, in particular, tend to suggest an asymptotic pattern, but do not explicitly relate the pattern to the redline threshold. For example, event-related potentials presented an asymptote pattern when participants were presented with the visual working memory capacity (Vogel & Machizawa, 2004). Results from this study suggest physiological measures will eventually reach a “plateau” and stop changing, creating an asymptotic pattern.

Based on the Yerkes-Dodson curve (Yerkes & Dodson, 1908), and the cognitive redline model (Wickens et al., 2015), a new model was developed to explain how the relationship between cognitive workload and performance will impact most physiological measures (for example, heart rate variability would show the opposite curve, as it decreases with increased cognitive workload), as shown in Figure 2. The model only considers medium to high workload, as underload was not studied in this research. Increased cognitive workload has been associated with decreased performance (Fox et al., 2007; Jaeggi et al., 2007; Kirsh, 2000; Mackay, 2000), and will increase some physiological measures, such as HR, and EDA. Performance will continue to degrade, and when the person reaches their redline, the physiological measures will reach an asymptote.

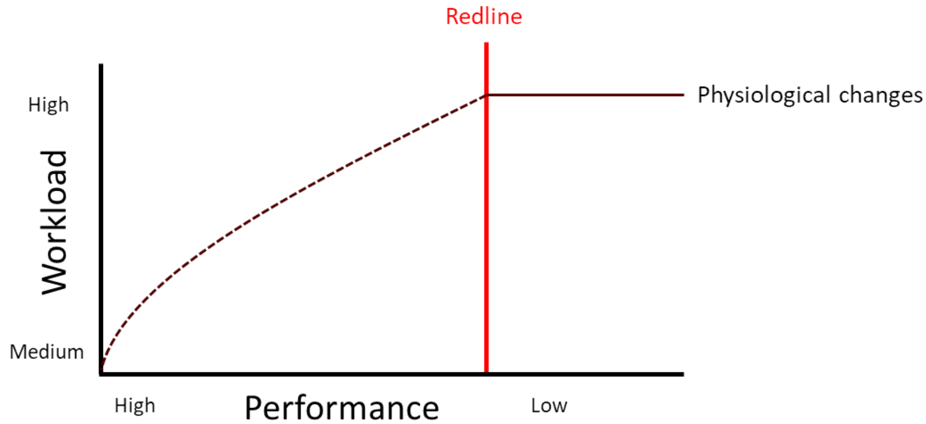


Figure 2: Cognitive workload and performance affect physiological measures.

The work presented in this dissertation will further expand the current redline definition by investigating physiological patterns and how they reflect workload when the person is approaching their redline. Thus, the proposed redline is defined as the threshold of available mental resources necessary to perform a task or set of tasks. The studies presented in this dissertation will further add to this definition by investigating the physiological patterns and autonomic arousal related to performance degradation when the person is approaching the redline threshold. Three different methods to measures workload will be used.

Methods to Detect Changes in Cognitive Workload

Cognitive workload can be measured through two main methods: subjective and objective. Subjective methods are introspective, as they depend on the person's thoughts and feelings. Objective methods are impartial, as they can be observed directly from the subject (Orlandi & Brooks, 2018). Objective methods include physiological measures, which are modulated by the autonomic nervous system, and performance-based measures, while subjective measures are based on perceived workload ratings.

While it is possible to measure workload using one of the methods for data collection, most studies use at least 2 for data comparison (Brookhuis & de Waard, 2001; Callan, 1998; Shakouri et al., 2018; Tsang & Velazquez, 1996). Using all three methods can provide a more comprehensive look at cognitive workload. To investigate physiological patterns associated with high levels of cognitive workload, data will be collected using all three methods. Table 1 contains a summary of the workload measurement technique, and examples of the measures collected.

Table 1: Workload Measurement Techniques

Workload Measurement Technique	Measures Collected
Physiological measures	• Heart rate (HR) • Heart rate variability (HRV) • Electrodermal activity (EDA) • Electroencephalography (EEG) • Pupillometry
Subjective ratings	• NASA-Task Load Index (NASA-TLX) • Short Stress State Questionnaire (SSSQ) • Workload Profile Index (WPI)
Performance measures	• Primary task • Secondary task • Reaction time

Autonomic Indices of Cognitive Workload

The autonomic nervous system (ANS) is one of the components of the peripheral nervous system (PNS), which consists of the ganglia and nerves excluding the brain and spinal cord (Hubbard, 2012; Mai & Paxinos, 2011). The PNS connects the central nervous system with the organs and limbs. Thus, the autonomic nervous system connects with glands and smooth muscle, influencing the unconscious functions of the internal organs (Mai & Paxinos, 2011). For example, autonomic functions include cardiac regulation, vasomotor activity, and some reflexes. In fact, the ANS is the most important regulator of cardiac function (Catterall, 2015). Autonomic functions are modulated by both branches of the ANS, the sympathetic (SNS) and parasympathetic nervous systems (PSNS) (Jansen et al., 1995; Waterhouse & Campbell, 2008).

The sympathetic (SNS) nervous system is responsible for the "fight or flight" response. This response is activated when the person experiences duress in the form of stress, fear, or intense exercise (Bers & Despa, 2009; Curtis & O'Keefe Jr, 2002; Fuller et al., 2010). Catecholamines (epinephrine, norepinephrine) stimulate the betta-adrenergic receptors located in the cardiac myocytes (Tsien et al., 1986), which increases the chronotropic, inotropic, and lusitropic heart states (Catterall, 2015). The chronotropic state is related to the increase in heart rate, inotropic is the strength of the cardiac muscle contraction during the systole, and the lusitropic is the relaxation rate during the diastole (Catterall, 2015). The parasympathetic nervous system (PSNS) is responsible for the "rest and digest" response, which slows heart rate in the absence of external threats (Porges, 1992). The PSNS maintains homeostasis (Jänig & Häbler, 2000; Pichon &

Chapelot, 2010; Recordati, 2003) Both the SNS and PSNS tend to act in opposite ways (i.e. one is inhibited while the other is activated), but there are exceptions where both systems may be active simultaneously.

The ANS modulates several body systems, including cardiovascular and respiratory. When the person experiences change in cognitive workload, the body reacts to the stressful situation by affecting physiological measures. Usually, the PSNS tends to withdraw so the body can enter into the "fight or flight" mode and prepare for the stressful situation (Ruscio et al., 2017). Thus, cognitive workload is reflected through changes in physiological measures (Cinaz et al., 2013; Luque-Casado et al., 2016; Mehler et al., 2011).

Physiological Measures

Data from physiological measures can be analyzed to infer changes in autonomic activity, which is related to changes in cognitive workload (Miyake et al., 2009; Ohsuga et al., 1995, 2001). The main advantage of using physiological measures is that the data can be collected in real-time, and with the advent of wearable devices, data can be collected in a wide-range of environments.

Cardiovascular measures, such as heart rate and heart rate variability are modulated by both branches of the autonomic nervous system, the sympathetic and parasympathetic nervous system. Thus, cardiovascular measures are sensitive to neurobehavioral processes (Berntson et al., 2017). Cardiovascular measures can be collected through optical or electrical sensors. Optical sensors use photoplethysmography (PPG), where a low-intensity infrared light is emitted from the sensor. The IR light is absorbed by biological tissue, but is more readily absorbed by blood. The sensor then detects changes in the light intensity, being able to detect blood volume changes in the blood vessels (Temko, 2017). While PPG sensors are not as accurate as the electrical sensors, they are wearable and inobtrusive, and they offer the opportunity of measuring heart rate while the person is moving. The electrocardiogram (ECG) collects the electrical signals of the heart, denoted with letters P, Q, R, S, and T to represent each of the segments of the heart rate cycle (Yeh & Wang, 2008).

Cardiovascular measures are perhaps the most commonly used physiological measures to assess changes in cognitive workload (Tao et al., 2019). Some studies have found heart rate to be insensitive to workload changes (e.g. Heine et al., 2017; Shakouri et al., 2018), but a few systematic reviews have found both heart rate and heart rate variability as sensitive measures to detect changes in cognitive workload (Charles & Nixon, 2019; Jorna, 1992; Tao et al., 2019).

The heart rhythm depends on the sinus node depolarization (Peltola, 2012). Decreased parasympathetic activity or increased sympathetic activity results in cardiac acceleration, and the opposite results in cardiac deceleration (Hutter & Trautwein, 1956). Heart rate has been shown to increase with high cognitive workload (Backs, 1995; Backs & Seljos, 1994; Brookhuis & de Waard, 2001; Veltman & Gaillard, 1996).

Heart rate variability (HRV) measures the natural beat to beat variations occurring within subsequent heart beats (Acharya et al., 2006; Malik, 1998) caused due to the dynamic balance that occurs between both branches of the ANS (Levy & Martin, 1984), that is, the polarization and depolarization of the sinoatrial node (Rocchetti et al., 2000). Usually, the R-R peaks in the QRS complex are analyzed, as these peaks are the most easily identifiable (Laguna et al., 1990). Heart period is the mean between the successive R-R peaks (Backs, 1995), which is different from heart rate, which is represented as the number of beats per minute. Heart period has been used to detect changes in cognitive workload (Backs, 1995; Veltman & Gaillard, 1996).

HRV can be analyzed based on different methods, which depend on the length of the time interval (short or long), and the type of analysis (time or frequency domain, geometric methods) (Task Force of the European Society of Cardiology, 1996). A short time window is considered to be 5 minutes, though some analysis have used 30-second interval (Huikuri et al., 1993).

The time domain uses statistical analysis, such as mean, standard deviation (SDNN), and the root mean square deviation (RMSSD) that occurs between the normalized (N-N) R-R intervals (Acharya et al., 2006). Other time-based analyses may look at percentages, for example, pNN50 is the percentage of NN intervals greater than 50ms (Acharya et al., 2006). The frequency domain analysis transforms the data to reflect changes in the autonomic nervous system, where the low frequency (LF), corresponding to 0.04 to 0.15 Hz reflects sympathetic activity, and the high frequency (HF), corresponding to 0.15 to 0.4 Hz reflects parasympathetic activity (Acharya et al., 2006; Bilchick & Berger, 2006). HRV has also been used to detect differences between the sleep-wake cycle (Malik, 1998). For shorter time periods, RMSSD and pNN50 are usually preferred, as they do not show differences due to the circadian rhythm (Sztajzel, 2004). However, some of the time-based and frequency-based measures are correlated. RMSSD and pNN50 show correlations with the HF analysis (Sztajzel, 2004). Studying both methods of HRV analysis is important, as time domain measures explain the amount of variability, while the frequency domain measures provide an explanation behind that variability (Stein et al., 1994). Several studies have linked decreased HRV with increasing cognitive workload (Hidalgo-Muñoz et al., 2018; Luque-Casado et al., 2016; Mehler et al., 2011).

Geometric methods present geometric patterns associated with the R-R interval, such as the triangular index, which plots the R-R interval on the X-axis, and the R-R length intervals on the y-axis (Acharya et al., 2006).

Electrodermal activity (EDA), also referred to as galvanic skin response (GSR), measures the electrical conductance and resistance of the skin. Skin conductance is only modulated by the SNS, which controls the eccrine sweat glands, thus affecting the electrical resistance and conductance of the skin (Solovey et al., 2014). The general skin potential can be broken down into two main components: tonic and phasic. Tonic changes correspond to skin conductance level (SCL), which represents slow changes, while the phasic corresponds to the skin conductance response (SCR), associated with the rapid peaks that usually occur after a stimulus (Boucsein, 2012). If skin conductance is collected using gelled electrodes, the solution needs to be isotonic as it has a similar composition to sweat (Schmidt & Walach, 2000). Several locations have been discussed as possible electrode placement, including the distal phalanxes of several fingers, the thenar and hypothenar eminences of the palm, the wrist, and the medial inner side of the foot (Ranogajec & Geršak, 2014). Previous studies have investigated how skin conductance can be used to measure cognitive workload (Ghaderyan & Abbasi, 2016; Kosch et al., 2019; Reimer et al., 2009). Of particular interest is the study that investigated event-related potentials reach an asymptotic pattern when visual working memory resources are depleted (Vogel & Machizawa, 2004), which may be related to the redline of cognitive workload.

Electroencephalography (EEG) records the electrical activity that originates in the brain cortex by using electrodes that go around the scalp and forehead. A reference electrode to filter data noise is usually placed on the ear. EEG analysis has identified several frequency bands: 0.5-3 Hz (delta), 4-8 Hz (theta), 8-13 Hz (alpha), 13-30 (Beta) and 40 to 50 Hz (gamma) (Mehta & Parasuraman, 2013). Some of these bands have been studied as indicators of cognitive workload. For example, the alpha frequency band shows attenuated activity as a response to increased cognitive workload (Berka et al., 2007; Galy et al., 2012), while the theta and beta frequency bands increase their activity (Fairclough et al., 2005). Additionally, the alpha frequency band has been found to be negatively correlated with task engagement (Lubar et al., 1995; Offenloch & Zahner, 1990), for which it has been suggested as an engagement index (Kamzanova et al., 2011; Pope et al., 1995).

Pupillometry studies pupil dilation. The pupil dilates as a response to changes in arousal and mental effort, and constricts as a response to brightness and to near fixation (Mathôt, 2018). The pupil tends to dilate quickly as a response to mental effort, and more complicated tasks are associated with larger pupil dilation mean values (Iqbal et al., 2004). The pupil dilation magnitude is usually on the order of 0.5 mm, and the pupil quickly returns to normal after completing the mental task that triggered the dilation (Klingner et al., 2008). Pupil dilation has been studied as an effective way to measure workload (Beatty, 1982). Pupillometry can also be collected with minimally-intrusive or non-intrusive devices, which makes it easier to implement (Palinko et al., 2010; Recarte & Nunes, 2003). However, several factors affect the results, such as changes in brightness (Pomplun & Sunkara, 2003). Additionally, the angle of the gaze can also impact the results (Pomplun & Sunkara, 2003). When collecting pupillometry, it is important to properly calibrate the device before each study to ensure proper data collection.

Baselining Procedure for Physiological Analysis

It is important to consider collecting a physiological baseline, which can be used in the analysis. For this research, the physiological baseline was collected while participants followed a paced-breathing app, using a ratio of 1:2 inhalation to exhalation. This breathing ratio has been found to decrease heart rate and blood pressure (Modesti et al., 2010). Using this ratio decreased the normal breathing rate, which is around 12 breaths per minute (Roscoe, 1992) to 6 breaths per minute, which have been recorded as producing higher amplitudes of heart rate variability (Song & Lehrer, 2003). Additionally, participants listened to nature sounds, as these have been found to decrease heart rate (Alvarsson et al., 2010; Ghezeljeh et al., 2017). Thus, the data is collected when the person is relaxed and experiencing minimal cognitive workload.

Previous studies that have investigated changes in physiological measures have also collected data a physiological baseline of varying lengths. Five minutes have been used (Wilhelm et al., 2004), as well as 10 to 15 minutes to allow the different body pressures to stabilize (Kergoat & Faucher, 1999). Pilot testing was done in the laboratory to determine the best time length to collect the baseline. Physiological measures (EDA, HR, HRV) were collected, and participants provided subjective feedback. Most pilot subjects agreed that they felt restless by the end of the 15 minutes, but reported that 10 minutes was a good, relaxing period. The physiological data was compared for the 10 and 15-minute periods, with very similar results. The shorter, 10-minute baseline was selected to ensure participants were relaxed before the study.

Subjective Measures

Subjective methods depend on personal workload perception, so direct input from the person is required (Rubio et al., 2004). Thus, the data collection can occur simultaneously while performing the task, or after the task is complete. If the data collection occurs while the task is ongoing, then the task needs to be interrupted. Otherwise, waiting to collect the data after the task is over may lead to memory decay, as the person may forget some of the details. Retrospective data collection may be used to avoid memory decay. For example in guided reconstruction, the person watches the video of themselves performing the task (Cacioppo et al., 1988). Subjective measures are cost-effective to administer, and provide insight into the person's workload perception, which may differ from data collected from objective methods. However, there are several shortcomings, which is why they are sometimes correlated with objective methods (McKendrick et al., 2019).

The Short Stress State Questionnaire (SSSQ) is a shorter version of the Dundee Stress State Questionnaire (DSSQ), and measures stress through 3 main factors: 1) task engagement, 2) distress, and 3) worry (Helton, 2004; Helton et al., 2005). Questions used to assess task engagement are based on participant’s motivation to complete the task, as well as the concentration and energy levels required to complete the task. Distress encompasses any negative effects that arise from the task, or from lacking control over the task. Worry refers to participant’s negative thinking styles (Matthews et al., 2013). The test is composed of a pre-, and post-test portion to compare how stress changed depending on the task. The SSSQ has been used in previous studies along with NASA-Task Load Index (Matthews & Campbell, 2010).

NASA-Task Load Index (TLX) classifies workload according to 6 different categories: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Participants rate each of the different 6 categories on a scale of 1 to 100. Then, participants rate the pairwise comparisons, consisting of a pair of categories, comparing all categories against each other for a total of 15 pairs. Participants then rate the highest category they thought provided the most workload during the task. Then, the score for each category is multiplied by the weight assigned to that category, and all are added together and divided by 15 to obtain the aggregate score (Hart & Staveland, 1988).

The Workload Profile Index (WPI) is based on Wickens' Multiple Resource Theory (MRT) model (Wickens, 2008), as users rate the tasks based on each individual resource used. For example, the task loads are rated according to the MRT dimensions in the processing state (perception/cognition, and response), sensory modality (vision, audition, and haptic), and the processing code (verbal/symbolic, and spatial/analog) (Rizzo et al., 2016; Rizzo & Longo, 2017; Rubio et al., 2004). To complete the questionnaire, users rate each of the MRT categories on a scale of 0 to 1, 0 if the resource was not used, and 1 if it was used. At the end, all the scores for each category are added together to obtain the aggregate score (Rubio et al., 2004). WPI has been investigated as a subjective tool to determine if there is any interference between different tasks (Phillips & Boles, 2004).

Performance-Based Measures

Performance data is usually collected while the person is performing the task, but the analysis is performed after the data collection is complete. In studies with high natural validity, or in real-life situations, collecting performance metrics may pose a challenge, especially as certain performance metrics may not be as easily observed unless an accident occurs. There are several ways to measure performance, including primary task performance, and secondary task performance.

Primary task performance depends mainly in observing how the person performs the task. Several metrics can be used, for example, reaction time, completion time, and number of errors.

When performing the secondary tasks, participants are usually instructed to perform the primary task as well as possible. Measuring performance on the secondary task can provide insight into how many attentional resources are required to perform the primary task. A common secondary task includes the n-back task, which consists of a presentation of words or shapes the subject has to keep in their working memory while performing the primary task (Brouwer et al., 2012; Herff et al., 2014; Pergher et al., 2019; Wang et al., 2015).

Autonomic Indices of the Redline of Cognitive Workload

Previous attempts to define the redline of cognitive workload have used subjective measures and performance. For example, the SWAT score of 40+/-10 (Colle & Reid, 2005; Reid & Colle, 1988) and the IMPRINT score of 60 (Mitchell et al., 2003). However, some studies have detected an asymptotic pattern in physiology. For example, event-related potential data followed an asymptotic pattern when the visual working memory was overwhelmed by a visual working memory task (Vogel & Machizawa, 2004). This recognizable pattern may be helpful in finding the redline. Similarly, heart rate, respiration rate, and galvanic skin response were found to follow a “plateau” pattern as the individual tasks in the study became increasingly more challenging (Mehler et al., 2009). A previous study done by author in a controlled multitasking environment found a similar pattern for heart rate variability (HRV) (Rodriguez Paras et al., 2015). The HRV pattern was consistent with the NASA-TLX subjective assessments, further explaining that physiological measures may stop fluctuating and stabilize into an asymptote pattern.

Physiological measures are modulated by the ANS, but not all measures are modulated by both branches. Cardiovascular measures (e.g. heart rate, heart rate variability), and breathing rate are modulated by the SNS and PSNS (Berntson et al., 2017; Stein et al., 1994). However, EDA is only modulated by the SNS (Boucsein, 2012; Dawson et al., 2017). The following research questions are based on the interaction between the SNS and PSNS activation and withdrawal. It is theorized that with increasing cognitive workload, the SNS activity increases, and PSNS withdraws. However, as the threshold of mental capacity approaches, the SNS activity will stabilize, thus reaching the asymptotic pattern.

Research Questions

The research addresses three research questions based on improving the detection of the redline of cognitive workload based on physiological measures.

The first research question investigates the sensitivity of physiological measures when the person is close to reaching their cognitive redline limit. Previous studies have analyzed physiological measures as workload indicators, but few to date have investigated the sensitivity of these measures at increasing and high workload levels. Heart rate tends to be the measure most commonly analyzed to detect cognitive workload changes (Charles & Nixon, 2019). Other measures include heart rate variability, electrodermal activity, and electroencephalography. All of these measures, except for EDA, are modulated by both branches of the autonomic nervous system. There is a debate based on the sensitivity of these measures, for example, the interpretation of heart rate data in relation to cognitive workload (Heine et al., 2017; Hidalgo-Muñoz et al., 2018; Shakouri et al., 2018). **Which physiological measures are more sensitive to changes in workload at very high levels that are near or exceeding the redline?** Chapter 2 provides information on the study that was conducted to answer this question.

The second research question expands on the results of the first questions, as now the physiological data is correlated with other workload measures, including performance and subjective measures. **To what extent do changes in these sensitive physiological variables compare with other standard measures of task performance and subjective measures?** The study that collected three types of data is explained in Chapter 3.

The last question combines information from the results of the previous two questions. Understanding the association among all three workload measures, as well as which physiological measures are more sensitive, this last question will investigate sympathetic and parasympathetic patterns associated with workload changes. **To what extent do physiological measures indicate the redline of cognitive workload based on patterns of the different branches of the autonomic nervous system?** The last study is presented in Chapter 4.

Proposed Cognitive Redline Risk Matrix and Analysis

Based on the current research results, a diagram can be helpful to understand when someone may be approaching their redline. The cognitive redline risk matrix (Figure 3) is based on all three types of workload measurement techniques, where each is assigned a value from 1 to 3. When performing the task, physiology is analyzed and compared to the baseline. If the measures are fluctuating, but still significantly close to baseline, then a 1 is assigned (some natural variation is expected when measuring physiological measures). A rating of 2 is provided if the measures increase, but are not significantly different from baseline, and if the measures have stopped changing and are significantly different from the baseline, then a 3 is assigned. Subjective ratings depend on how the person subjectively feels about the task, and depending on the data collection, these could be collected based on existing questionnaires, such as NASA-TLX, or another survey or questionnaire specific to the task could be developed. The rating of 1 is assigned if the person feels the workload is manageable. A rating of 2 is assigned if the workload is perceived as high, but statistically, is not significantly different from the baseline (level 1), and the rating of 3 is provided when the task demands overwhelm the person, and the subjective score is significantly different from level 1. Performance measures can be observed from the task. If performance is acceptable, a rating of 1 is provided. A rating of 2 observes some performance degradation, but not significantly different from the baseline observed at level 1. The third rating is awarded to severe performance degradation, which is significantly different from the baseline observed at level 1.

The risk levels would be observed for each of the classifications and plotted in the cognitive redline risk matrix, which uses a color-coded gradient to indicate how close the person is to their redline. The green color indicates the person may be far from reaching their redline, while the red color indicates the person may be closer to their redline.

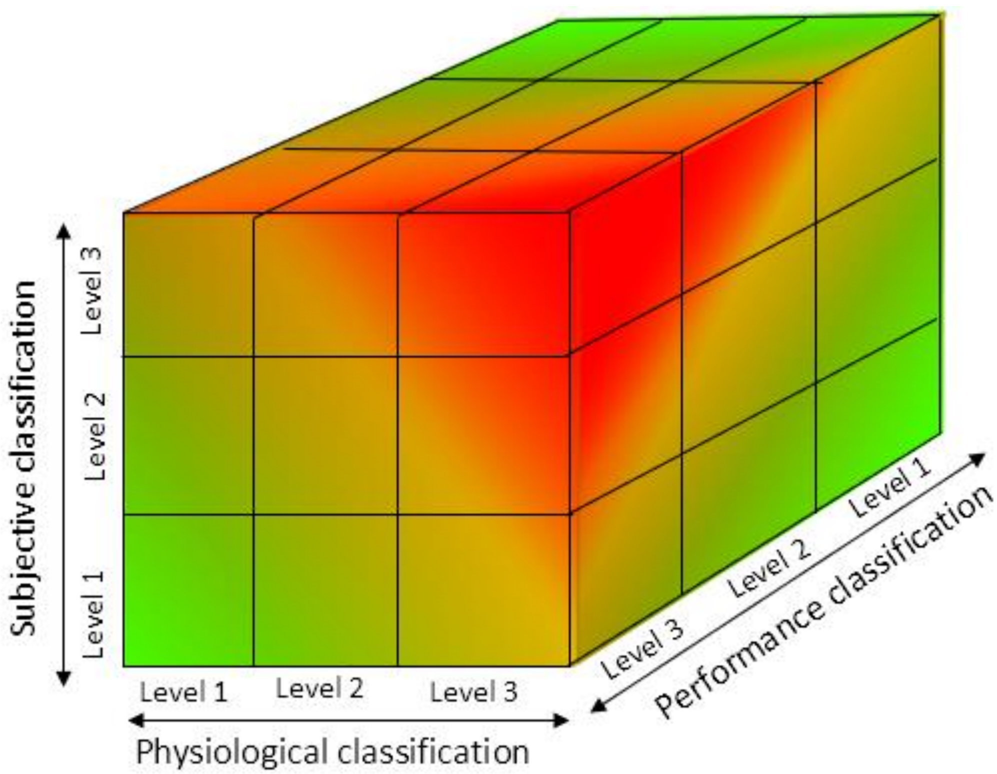


Figure 3: Redline of cognitive workload risk matrix proposed diagram.

Organization

This dissertation is organized in 5 chapters, the first of which is a literature review discussing cognitive workload and its measuring methods. An overview is provided on previous work related to the redline of cognitive workload, and the novel approach this research is taking by documenting the redline limit using autonomic indicators. Thus, the literature review also includes information on how autonomic indicators can be used as workload indices.

The research questions, based on the literature review, are also detailed in the introductory chapter. The subsequent chapters (chapters 2, 3, and 4) focus on covering a specific study to address each of the research questions. Finally, chapter 5 provides a discussion based on the findings and provides an overview for future work.

CHAPTER II

PHYSIOLOGICAL INDICES OF WORKLOAD

Introduction

Humans must maintain adequate performance levels, particularly in safety-critical domains (e.g. healthcare, aviation, oil and gas), where even a small mistake can be costly and have adverse consequences. Such environments usually involve multitasking, which can influence the level of cognitive workload experienced by the workers. Identifying when someone is approaching the threshold of their available mental resources could prevent performance degradation, thus having a positive outcome in safety. This creates a need to study and understand how cognitive resources (e.g. attention, memory) provide indications of when people are approaching their threshold and their performance degrades. Additionally, it is important to understand the relationship between utilization of cognitive resources and performance.

When a person approaches the threshold of their cognitive resources, their performance tends to degrade, as described by the redline of cognitive workload model (Grier et al., 2008; Wickens, 2008). As defined in Chapter 1, previous studies have found a correlation between elevated workload levels and performance degradation, suggesting at an ideal cognitive workload level to sustain adequate performance. For example, a Subjective Workload Assessment Technique (SWAT) score of 40 +/- 10 predicts possible performance difficulties, providing a redline range (Colle & Reid, 2005; Reid & Colle, 1988), suggesting the person has exceeded their workload capacity.

Similarly, several studies have provided workload scores for the Improved Performance Research Integration Tool (IMPRINT). In the visual, auditory, cognitive, and psychomotor (VACP) option of IMPRINT, workload is calculated on a 7-point scale according to each of these mental resources. Receiving a score of 7 is indication that the person has exceeded the available resource capacity (McCracken & Aldrich, 1984). The IMPRINT advanced model has the additional workload option of speech, and the five resources are rated based on a 7-point rating scale. The individual scores are then added to obtain the overall workload value. According to subject matter experts (SMEs) familiar with the mounted combat system (MSC), a score of 60 or more was considered as high workload for the MSC task, and thus associated with performance degradation (Mitchell et al., 2003).

While these studies have clearly established the relationship between subjective workload ratings and performance degradation, the studies only explain when the person experiences high levels of workload, and don't detect when the risk for reaching the threshold is growing. Future investigations in this dissertation take a more practical approach by using physiological measures to identify the redline threshold, particularly investigating the detection of workload dynamics that may impact performance, as workload is only a problem if it is related to a risk of performance degradation.

While physiological changes have been linked with changes in cognitive workload (see Chapter 1), to date, to the author's knowledge, this hasn't been done. As described in Chapter 1, physiological measures reflect changes in the autonomic nervous system. Under extreme workload conditions, the person experiences the fight or flight response, associated with sympathetic activation. Thus, physiological data can be used to infer changes in cognitive workload. The theory presented in this dissertation explains that, near the redline threshold, physiological measures can show less responsiveness to cognitive workload changes, similar to an asymptotic pattern. The first study presented in this dissertation is designed to explore the possible relationship between common physiological measures and how they reflect changes in workload when the person is approaching their potential redline threshold. Subjective measures and performance will also be analyzed.

This first study aims to answer the first research question postulated in the larger body of research: **Which physiological measures are more sensitive to changes in workload at very high levels that are near or exceeding the redline?** Several physiological measures (i.e. breathing rate, heart rate, heart rate variability, encephalography, and electrodermal activity) were selected based on findings from previous studies. For example, heart rate and heart rate variability (Luque-Casado et al., 2016; Mehler et al., 2011; Nickel & Nachreiner, 2003; Yang et al., 2013), breathing rate (Roscoe, 1992), and electrodermal activity (Boucsein, 2012) reliably indicate changes in sympathetic arousal which can be associated with changes in workload. The alpha wave from electroencephalogram has also been demonstrated as relating to cognitive workload changes (Knoll et al., 2011).

Two other measures (subjective methods and performance) will be used in addition to physiological measures to infer the redline of cognitive workload. Several subjective questionnaire-style methods were used, as each questionnaire provides a different perspective on participants' cognitive workload. Participants completed tasks set in the MATB-II environment with 5 different levels of imposed workload, while their physiological data were collected. MATB-II has been used in previous studies to impose cognitive workload, while physiological measures are collected to measure such changes. For example, electrophysiological measures (heart rate and respiration rate) showed significant differences in response to different taskloads associated with higher levels of mental workload (Nixon & Charles, 2017). Another study found decreasing NN intervals with increased cognitive workload (Dell'Agnola et al., 2018). Similarly, RMSSD and pNN50 measures of HRV were lower for the higher stress level induced by MATB-II (Kennedy & Parker, 2017). Overall, MATB-II has been used to induce varying levels of workload while physiological measures were collected to distinguish across different workload levels.

Findings from this research can expand the definition of the redline threshold to include physiological indices. Additionally, the study explored which physiological measures are more sensitive to changes in cognitive workload when the person is approaching their redline threshold.

Methods

Thirty participants completed the IRB-approved study (IRB approval: IRB2014-0499D), see Appendix A for IRB materials. Participants were at least 18 years old, had normal, or corrected-to-normal vision, and no impairments of the hands that prevented them from using a mouse, keyboard, and joystick.

Participants completed a series of tasks in a multitasking context with controlled levels of difficulty. Physiological data were collected to infer patterns in workload assessment data.

Procedure

Participants were given a brief overview of the study when they arrived at the laboratory. After consenting to participate in the study, participants completed a background questionnaire, which sought information about demographics, videogame and flight simulator experiences, and caffeine consumption. Then, participants proceeded to complete the questions in the pre-experiment section of the Short Stress State Questionnaire (SSSQ) (Helton, 2004).

Participants then put on the physiological devices (see Figure 4) , which included the Zephyr BioHarness3 (Zephyr Corporation, 2012), Shimmer GSR (Realtime Technologies Ltd, 2017), and NeuroSky MindWave (NeuroSky, Inc., 2015),with help from the experimenter. Each of the devices was calibrated to ensure proper streaming and storage of data before proceeding with the study.

Figure 4: Physiological devices used in the study. From left to right: Shimmer GSR, NeuroSky MindWave, and Zephyr BioHarness.

The physiological baseline data (explained in Chapter 1) were then collected, which were used in the statistical analysis as a covariate. The baselining process consisted of following the instructions to perform paced breathing while listening to nature sounds (johnnielawson, 2013) through noise-cancelling earphones for 10 minutes (Kergoat & Faucher, 1999). The paced breathing instructions were presented using the Paced Breathing app (TrexLLC, 2014), as shown in Figure 5, and consisted of a ratio of 1:2 inhalation to exhalation, as previous studies have found this ratio to decrease heart rate and blood pressure (Modesti et al., 2010). Thus, the data collected during this time period are related to when the person is in a relaxed state, the most basic level of ANS activity, and experiencing minimal cognitive workload. All physiological variables (HR, HRV, EDA, EEG, BR) investigated in this study were collected during the baselining procedure.

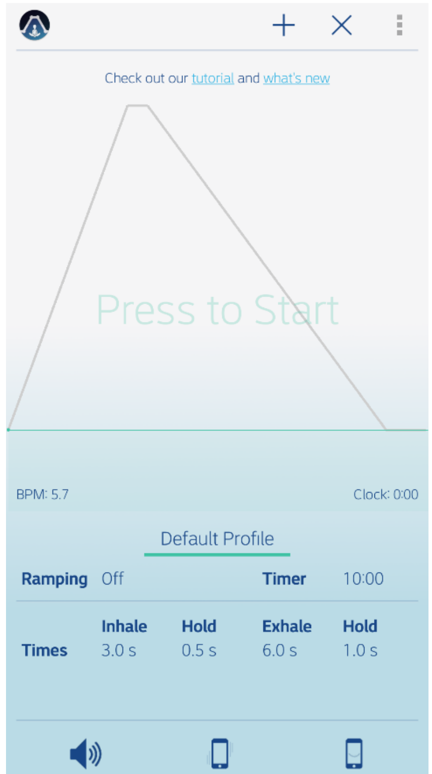


Figure 5: Paced Breathing app (TrexLLC., 2014).

Participants then proceeded to complete a training on how to operate the Multi-Attribute Task Battery-II (MATB-II, see section “Software to Manipulate Mental Workload” in this chapter), in order to reduce the likelihood of learning effects (Prinzel et al., 2000). The training consisted of a PowerPoint presentation with instructions for each of the tasks, explained by the experimenter. Participants were encouraged to ask questions during the presentation. Then, participants completed 2 training scenarios in the MATB-II software, each with a 3-minute duration. The training difficulty was similar to the medium and hard scenarios (see Table 2 in section “Software to Manipulate Mental Workload”).

The experimenter coached the participant during the training scenarios to ensure the participants understood the correct functioning of the MATB-II software. At the end of each MATB-II training scenario, participants had the chance to ask any questions before proceeding with the study.

After completing the training, participants took a 3-minute resting break, where, similar to the baselining process, they were instructed to follow the paced breathing app (TrexLLC, 2014) while listening to nature sounds (johnnielawson, 2013). This break was used to minimize the risk of a carryover between different scenarios, allowing physiological measures to return to a relaxed state with minimal cognitive workload before participants completed the next MATB-II scenario.

Then, participants completed a 2-minute MATB-II scenario. The difficulty of the scenario ranged from very easy to very difficult (see Table 2 for scenario difficulty). At the end of each MATB-II scenario, participants completed the NASA-TLX questionnaire, and the Workload Profile Index (WPI). Two different subjective workload assessments were collected, as the NASA-TLX provides diagnostic information about the sources of workload (Hart & Staveland, 1988), and the WPI provides information based on the cognitive resources used for the task (Rubio et al., 2004)see Chapter1). The 3-minute break and MATB-II scenario cycle of trials was iterated a total of 5 times.

After participants completed the fifth and final MATB-II scenario, they proceeded to complete the NASA-TLX pairwise comparison survey(Hart & Staveland, 1988). The NASA-TLX pairwise comparison is a weighting procedure, which compares each of the 6 different NASA-TLX categories, presented in pairs, and the assigned weight is then used to obtain a global score (Hart & Staveland, 1988). The SSSQ post-test questionnaire (Helton, 2004) was the final questionnaire completed by the participants. Then, participants removed the physiological devices to conclude the study. A summary of the study procedure is given in Figure 6.

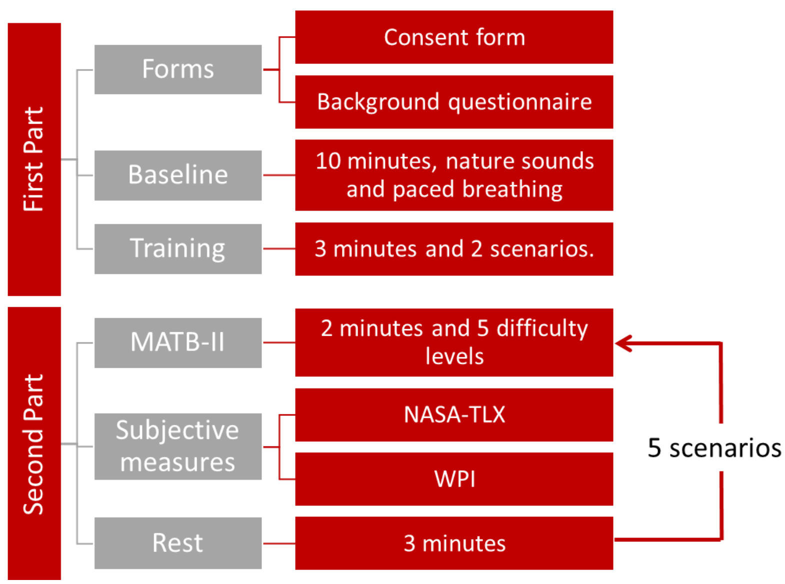


Figure 6: Experimental procedure.

Apparatus

This section discusses the different physiological devices, subjective questionnaire-style methods, and the software used to manipulate mental workload utilized during data collection. A summary of the dependent measures collected throughout the study is provided in Table 4.

Physiological Devices

Three types of devices were used to collect physiological data in real-time while the participants conducted the study activities, including the pre-experiment questionnaires, pre-experiment physiological baselining procedure (see “Physiological Baseline” in Chapter 1), experimental trials, and post-study questionnaires. The devices were calibrated using a standard calibration procedure for each device after participants put them on, and then the data collection started.

Cardiovascular measures (HR and the R-R interval) and breathing rate were collected with the Zephyr BioHarness3 (Zephyr Corporation, 2012), as shown in Figure 7. BioHarness3 consists of an adjustable chest strap, which provided skin contact at key electrode locations, and a "biomodule" with embedded ECG, breathing, and other sensing capabilities to collect ECG, breathing rate, and other data (Zephyr Corporation, 2012). The chest strap was moistened by the experimenter before use to ensure proper skin contact, and thus minimize data loss. Several studies have investigated the use of the BioHarness3 as a tool to collect HR, HRV, and BR (Hailstone & Kilding, 2011; Johnstone, Ford, Hughes, Watson, & Garrett, 2012a, 2012b; Johnstone, Ford, Hughes, Watson, Mitchell, et al., 2012).

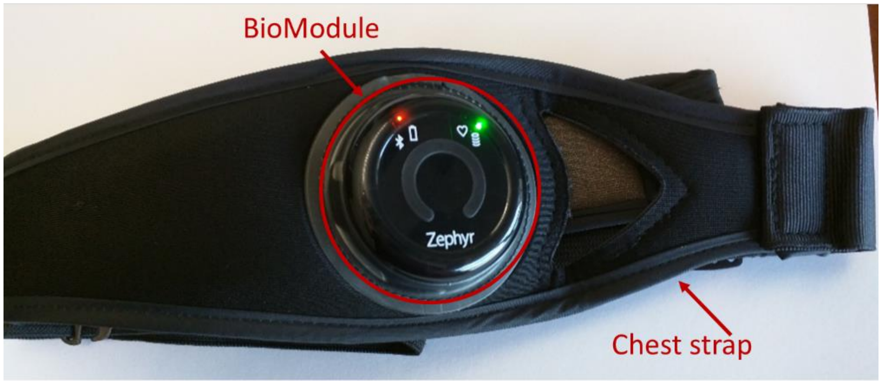


Figure 7: Zephyr BioHarness3.

The collected cardiovascular data were recorded as heart rate, in beats per minute, and heart rate variability, recorded as the R-R interval in seconds (Zephyr Technology, 2016). The R-R intervals were extracted, and then analyzed based on time-domain features, such as pNN50, before the statistical analysis. Breathing rate is provided in breaths per minute (bpm), which includes the whole breathing cycle (one inhalation and one exhalation), measured by a pressure sensor in the strap (Zephyr Technology, 2016). A summary of all physiological measures collected in the study and their units is provided in Table 4.

The NeuroSky MindWave (NeuroSky, Inc., 2015) recorded EEG data via a single dry electrode (Figure 8) placed in the FP1 position, according to the international 10-20 system (Herwig et al., 2003; Homan et al., 1987, p. 10). The reference electrode is clipped on the ear to reduce data noise.

Figure 8: NeuroSky Mindwave. The figure on the left presents the front of the device; the single electrode is placed in the FP1 position is visible in the figure on the right. The clip attaches to the ear as the reference electrode.

The EEG data are reported by the software in the form of frequency bands, including Alpha and Beta (de Munck et al., 2009; Zheng & Lu, 2015). The power units represent the relative amplitude of each individual EEG band (NeuroSky, Inc., 2014). The alpha wave data were analyzed, as this power band has been previously studied to measure changes in cognitive workload (Brouwer et al., 2012; Matthews et al., 2017).

The ShimmerGSR system (Realtime Technologies Ltd, 2017) collected EDA data. The device consisted of an amplifier worn around the wrist, and 2leads attached to isotonic gel electrodes were placed on the thenar and hypothenar eminences (see Figure 9 (Dawson et al., 2017) in the palm of the non-dominant hand.

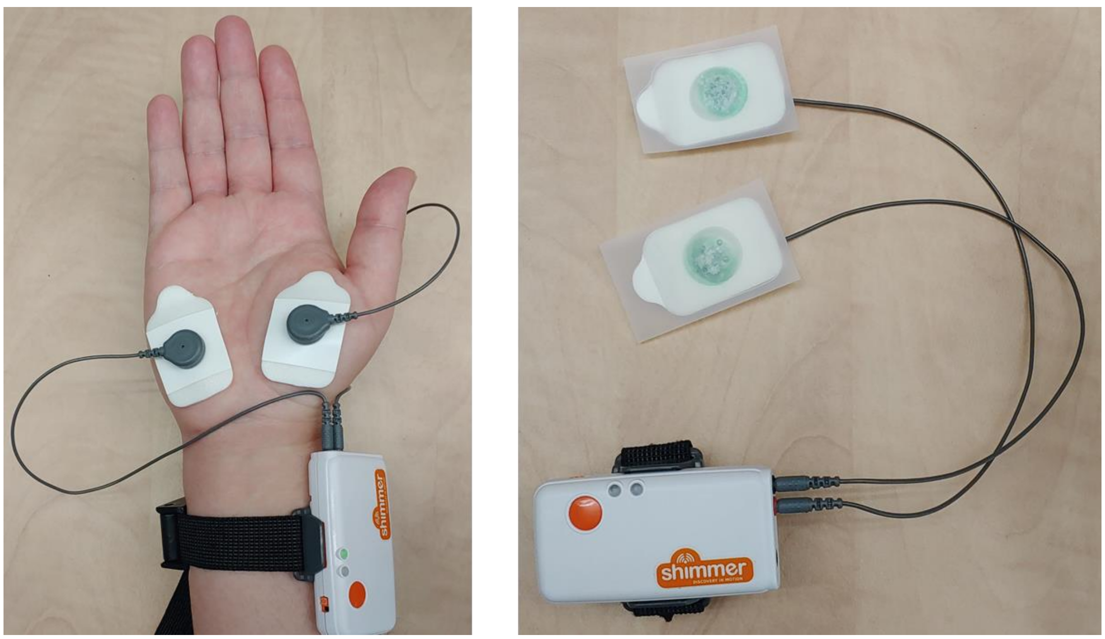


Figure 9: Shimmer GSR. Figure on the left shows the correct lead placement on the thenar and hypothenar eminences. The right figure shows the amplifier with the leads.

Even though EDA is traditionally collected from the non-dominant hand to avoid movement and motion artifacts, in this study the data were collected from the dominant hand, as the nondominant hand was used to control the keyboard and mouse and would have created more motion artifacts in the data. Extensive pilot testing was done by the experimenters, which consisted of collecting EDA data from both wrists while interacting with the MATB-II software. Results from the pilot test determined that using the nondominant hand on the joystick minimized motion and thus the motion artifact.

The raw ShimmerGSR data were analyzed using Ledalab, a Matlab plug-in, to extract skin conductance response (Bach, 2014; Karenbach, 2005). The extracted data were then statistically analyzed.

Software to Manipulate Mental Workload

The Multi-Attribute Task Battery-II (MATB-II) (Santiago-Espada et al., 2011), a software developed by NASA, was used to manipulate the workload conditions (see Figure 10). The software consisted of four different tasks, modeled after those found in an aircraft cockpit, including system monitoring (SYSMON), tracking (TRACK), communications (COMM), and resource management (RESMAN) (Comstock Jr & Arnegard, 1992), as shown in Figure 10. A summary of the tasks is provided in Table 3.

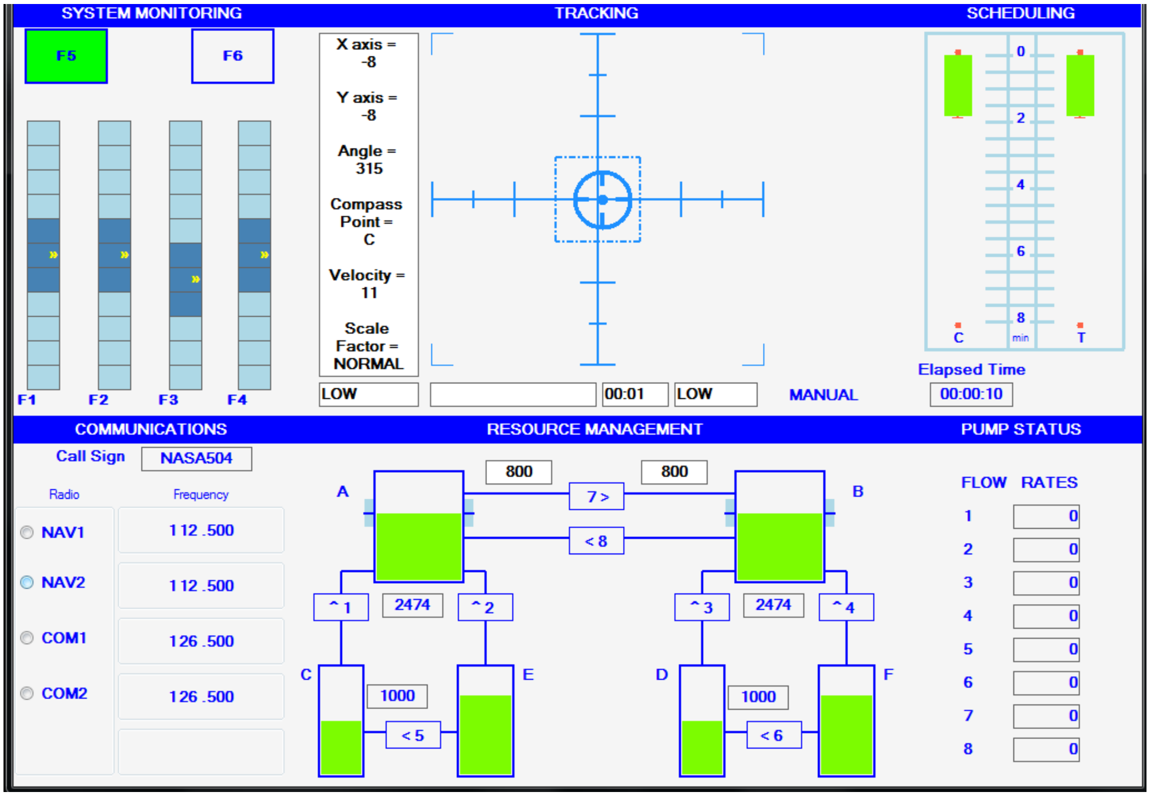


Figure 10: The Multi-Attribute Task Battery-II (MATB-II) (Santiago-Espada et al., 2011).

The software was operated through a mouse, keyboard, and a joystick. The joystick was used with the dominant hand to control the pointer in the tracking task. The mouse was used with the non-dominant hand to provide the reply in the communications task. Both the mouse and keyboard could be used to interact with the resource management and system monitoring tasks. The study setup (Figure 11) was reconfigured based on the handedness so that participants' dominant hand controlled the joystick.

Figure 11: MATB-II study configuration, with the mouse on the left and the joystick on the right.

The compensatory tracking task (TRACK) (Lee et al., 2017) consisted of controlling a pointer and centering it as close as possible within the center square (see Figure 12). The software-defined "response" variable describes the amount of randomized error that is combined with operator input for each update cycle, while the "update" variable is similar to control gain, referring to the magnitude of effect that joystick inputs have on the moving target for every update cycle (Santiago-Espada et al., 2011), similar to control gain. Both the "response" and "update" can be specified in 3 levels: "low," "medium," and "high."

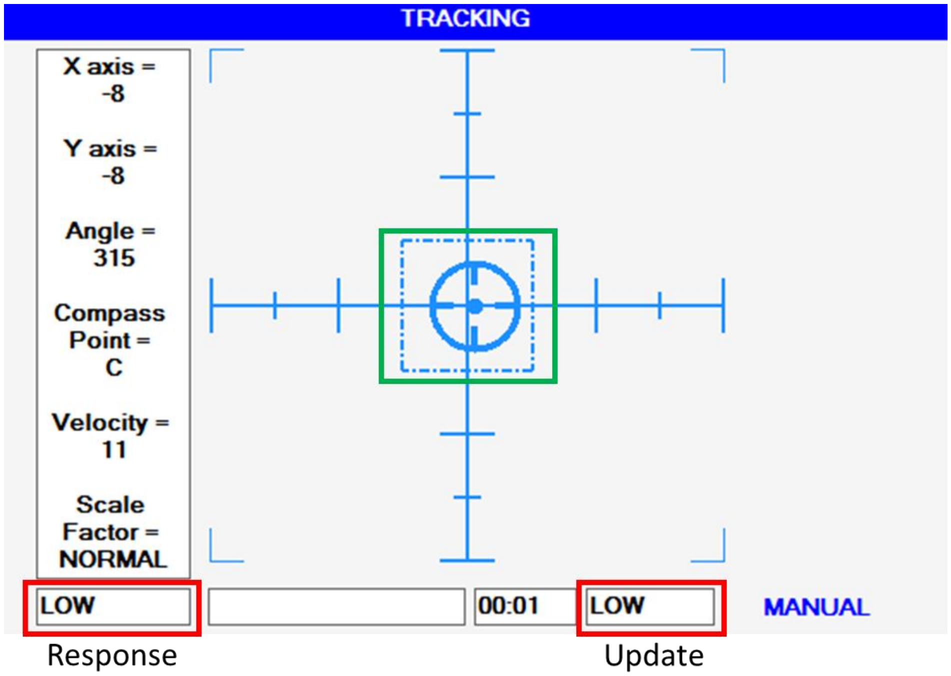


Figure 12: MATB-II tracking task (TRACK). The response and update settings (red rectangles) can be set to low, medium or high. The green square highlights the pointer, which must be centered as close as possible within the center square.

The other three MATB-II tasks had to be completed at the same time as the tracking task. The system monitoring (SYSMON) and resource management (RESMAN) (see Table 3) tasks were conducted using a combination of mouse and keyboard inputs, and the communications (COMM) task was conducted with the mouse.

The system monitoring (SYSMON) task consisted of a system of dynamic lights and scales (see Figure 13). The light system consisted of 2 lights, labeled “F5” and “F6.” Each light displayed 2 states, a “light off” and “light on.” For F5, the “light on” state was green, and required pressing F5 when the light was turned off. For F6, the “light off” state required pressing F6 when the red light turned on.

The scale system consisted of 4 fluctuating weights in the scales, labeled F1 through F4 (see Figure 13). When the weights in the scale touched either end, the system required pressing either on the scale using the mouse, or on the keyboard key (F1 through F4) corresponding to the malfunctioning scale. In the SYSMON task, the number of malfunction incidences can be modified according to the task.

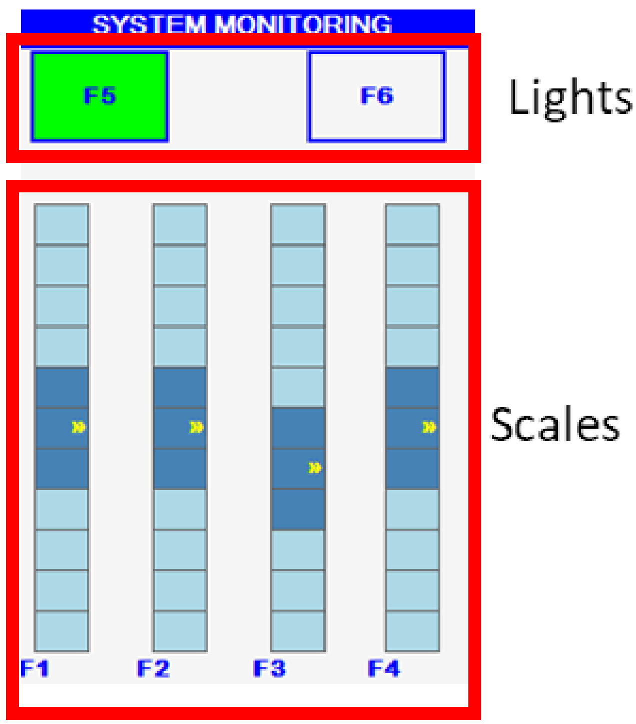


Figure 13: System monitoring task. The upper rectangle shows the light system. The lower rectangle highlights the scales.

The communications task (COMM) provided different auditory call signs that were modeled after identifiers used in air traffic control radio conversations. Only responses to the aircraft call sign, “NASA 504,” were considered correct. All other all signs had to be ignored. To reply, the mouse was used to change the radio and frequency (see Figure 14).

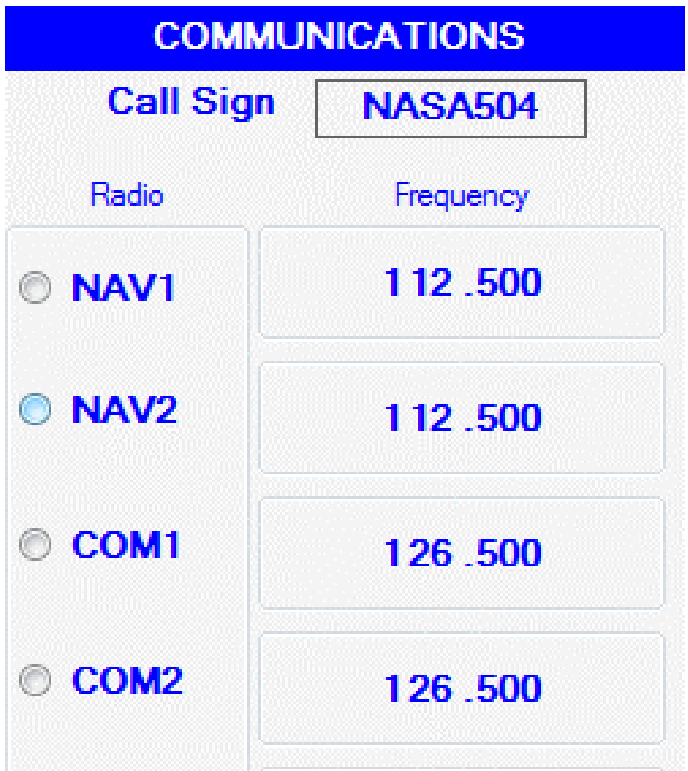


Figure 14: Communications task (COMM). Participants were instructed to listen to their callsign, and change the radio and frequency.

The resource management (RESMAN) task consisted of two fuel tanks, “A” and “B,” as shown in Figure 15. The fuel levels in tanks A and B would slowly deplete, and the fuel had to be kept as close as possible to 2,500. Additional fuel could be passed to these tanks via the other fuel tanks (“C”, “D”, “E” and “F”) by activating the different fuel pumps (labeled 1 through 8); the arrows next to the number indicated the direction of the fuel flow.

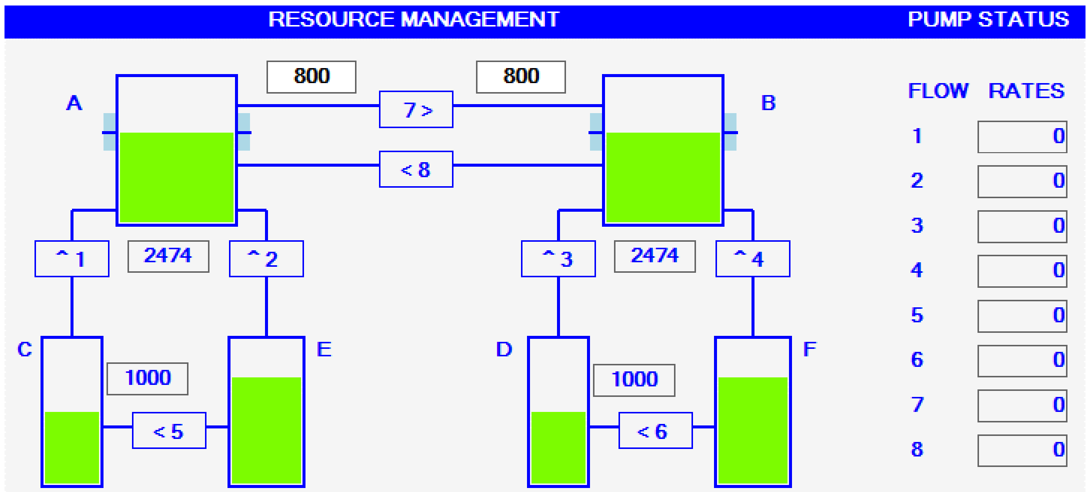


Figure 15: Resource management (RESMAN) task. Participants were instructed to keel tanks A and B as close as possible to 2,500 by operating the fuel pumps.

Cognitive workload was manipulated based on the number of "events" for each of the four MATB-II tasks. A summary of the number of events for each task and scenario is provided in Table 2. The TRACK task depended on the settings for the "update" and the "response" (see Figure 12). The cognitive workload induced by the SYSMON task was modulated by increasing the number of incidences (i.e., malfunction of the dynamic light and scale systems) that required interaction. The cognitive workload imposed by the COMM task was determined based on the number of "OWN" and "OTHER" call signs (see Table 2). Lastly, the cognitive workload from the RESMAN task was manipulated by increasing the number of failing pumps, as well as the total amount of time the fuel pumps failed.

Table 2: MATB-II Difficulty Levels

Difficulty Level	**1**	**2**	**3**	**4**	**5**
TRACK					
Response	Low	Medium	High	High	High
Update	Low	Low	High	High	High
SYSMON					
Green light	1	4	6	10	11
Red light	1	4	6	10	11
Scales	1	6	15	20	22
COMM					
Own	1	2	5	5	5
Other	1	2	3	3	5
RESMAN					
Failing pumps	2	5	5	5	10
Total time (s)	10	45	43	112	108

The MATB-II scenarios were developed based on iterative pilot testing performed by experienced MATB-II users, including the experimenter and 3 test subjects. The experimenter developed the first level with the easiest difficulty and number of events for each of the 4 tasks. Then, the other 4 scenarios were coded by increasing the difficulty and number of events for each task. The pilot subjects would then complete each of the scenarios presented in a counterbalanced order and rate them using NASA-TLX. Additionally, the pilot subjects also provided verbal feedback to understand which tasks required additional modifications. The experimenter would then take the NASA-TLX scores and verbal feedback into consideration to adjust the difficulty levels. The scenarios were iteratively refined based on users' feedback and NASA-TLX scores to have an equivalent magnitude of change between difficulty levels.

The training duration for the MATB-II scenarios was based on a previous study that used MATB-II to measure workload. The training consisted of a 5-minute long scenario with difficulty similar to the experimental trials (Karpinsky et al., 2016). Pilot testing was performed with the pilot subject, but even though the subjects were explained the use of MATB-II, they still had questions while performing the task. Thus, training was split in two 3-minute scenarios, where participants were asked after the first training scenario if they had any questions. At this point, the experimenter could go back to the MATB-II diagram and answer any questions participants may have before proceeding with the second training scenario.

Several MATB-II studies have used fewer scenarios with a longer duration. For example, 2 scenarios each with a 10-minute duration (Roy et al., 2016), 20-minute duration (Karpinsky et al., 2016), 3 scenarios with a 15-minute duration (Heard et al., 2019), or 4 scenarios with a 5-minute duration (Gevins & Smith, 2003). The pilot testing started with the last model of 4 scenarios with a 5-minute duration. However, 4 scenarios were not enough to discriminate between the high workload levels, so another scenario was added. However, pilot subjects reported feeling cognitively tired at the end of the data collection involving the sensor calibration, baseline data collection, training, and 5 scenarios with a duration of 5 minutes each, with the 3-minute break in-between. The scenarios were reduced to 2 minutes, which also allowed for ultra-short HRV data to be analyzed (Melo et al., 2018).

The break in-between tasks was modeled after a similar study that incorporated a 5-minute break between MATB-II tasks. However, this break was also used to complete the subjective assessment (Albuquerque et al., 2018). For the current study, the break was 3 minutes, but did not include the subjective assessment. However, when taking into consideration the subjective assessments, the total time between scenarios is around 5 minutes.

Performance measures (see Table 3) were calculated from the MATB-II data collected from each of the four tasks throughout each scenario. The tracking task performance consisted of the root mean square deviation in pixels from the center target. The SYSMON task collected the response time for each malfunctioning system (i.e. lights and scales) each time the participant interacted with the system. The reaction time, the number of hits and misses, and the time to reset were used to calculate the percentage of time the system was in a failed state (Splawn, 2013). The COMM performance measure was analyzed based on the responses to the “OWN” call sign, including the reaction time, number of hits and misses, and the time to reset to calculate the percent of time the system was in a failed state (Splawn, 2013). RESMAN performance consisted of the percent error between the target of 2,500 fuel units and the actual fuel level (Splawn, 2013).

Table 3: MATB-II Tasks and Performance Measures

Task	**Control**	**Workload Modification**	**Performance Measures**	**Units**
TRACK	Joystick	Response and update	Root mean square deviation of the number of pixels target was away from the center	Pixels
SYSMON	Mouse or keyboard	Number of light and/or scale malfunctions	Average response time for lights and scales	% time the system is in a failure state
COMM	Mouse	Number of calls ("OWN" and "OTHER")	Number of correctly answered "OWN" calls, and correctly ignored "OTHER" calls	% time the system is in a failure state
RESMAN	Mouse or keyboard	Number of malfunctioning pumps, and total time pumps were failing	Average deviation from the 2,500-fuel target level in tanks "A" and "B"	Fuel units

Subjective Questionnaires

Several subjective questionnaires were used to collect data, as each questionnaire provided a different perspective on participant's cognitive workload. More details about these questionnaires are provided in Chapter 1.

A background questionnaire was developed to capture additional information on demographics, such as age, sex, and student classification (e.g. undergraduate, master's, Ph.D., or professional). Other questions inquired about participants' experience with videogames, including if participants played videogames, the number of hours played per week, system used (e.g. PC, Xbox, Playstation, WiiU, Nintendo3DS, etc.), and if it required the use of a joystick. Because MATB-II's tasks are based on a flight cockpit, the questionnaire also had questions related to participant's experience with flight simulators. Lastly, the last few questions were related to caffeine consumption, as caffeine can affect physiological measures.

The Short Stress State Questionnaire (SSSQ) (Helton, 2004; Helton et al., 2005) was used to determine if the changes in physiology were due to stress, by considering a pre- and post-assessment to measure incidental stress. The first part of the SSSQ was completed before the study, and the post-study was completed at the end of the last MATB-II scenario. The total score for the SSSQ was calculated according to the formula: (pre-task score – post-task score)/(standard deviation of the pre-task score) (Helton, 2004).

The Workload Profile Index (WPI) (Rubio et al., 2004) inquired about the different workload dimensions based on Wicken's Multiple Resource Theory (Wickens, 2008). All relevant dimensions from the Multiple Resource Theory Model were included for the purpose of the study. The WPI is usually rated using a scale of 0 for no engagement, and 1 for full engagement (Rubio et al., 2004). The scales were adapted from 0 to 5 to obtain a greater resolution to the workload changes between MATB-II scenarios, based on pilot testing. Figure 14 shows the modified WPI version used in this study.

Please fill an integer number from 0 to 5 into the box.

0 = No engagement at all **5 = full engagement.**

Yellow	Workload Dimensions						
	Stage of Processing		Code of Processing		Input		Output
	Perceptual/ Cognition	Response	Spatial	Verbal	Visual	Auditory	Manual
MATB-II Task							

Figure 16: The modified Workload Profile Index (WPI) version used in this study.

NASA-Task Load Index (NASA-TLX) (Hart & Staveland, 1988) was used to determine overall cognitive workload based on 6 different dimensions. The NASA-TLX version used was included in the MATB-II software, and consisted of sliding scales. This version of NASA-TLX appeared at the end of each scenario (see Figure 17). The pairwise comparison necessary to obtain the weights was provided at the end of the study. The total NASA-TLX score was calculated by multiplying the scores from each scenario by the weights, and then dividing by 15 to keep the scale from 0 to 100 (Hart & Staveland, 1988).

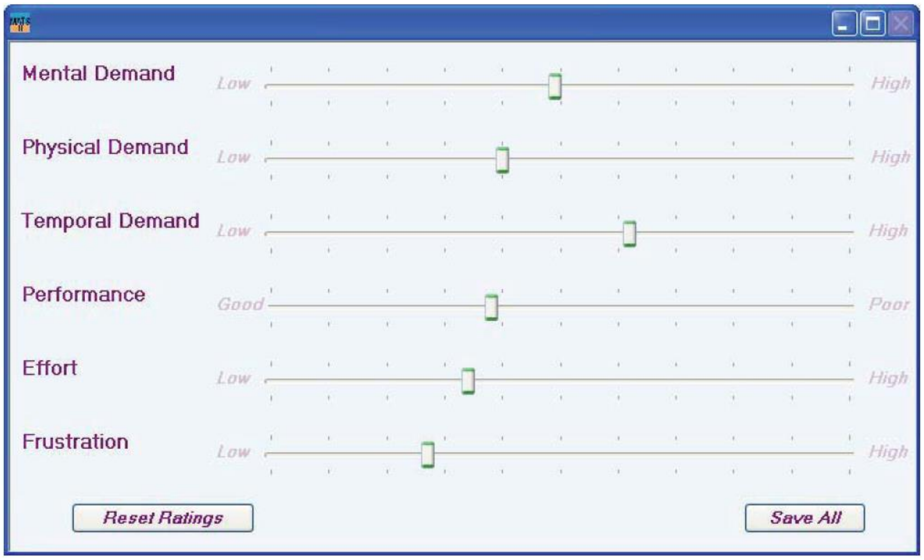


Figure 17: NASA-TLX.

Experimental Variables

The dependent variables are based on the three methods for detecting changes in cognitive workload (i.e. subjective methods, performance metrics, and physiological measures). A summary of the variables collected is listed in Table 4.

Table 4: Dependent Variables

Workload Measure Type	**Measure Collected**	**Device or Method**	**Unit**
Physiological	Heart Rate	BioHarness3	Beats per minute (BMP)
Physiological	Heart Rate Variability (HRV), pNN50	BioHarness3	% (percentile of normal to normal (NN) intervals greater than 50 ms
Physiological	Breath Rate (BR)	BioHarness3	Breaths per minute (BPM)
Physiological	Electrodermal Activity (EDA) (skin conductance response)	ShimmerGSR	Microsiemens
Physiological	Electroencephalography (EEG)	NeuroSky MindWave	Device units
Subjective	Short Stress State Questionnaire (SSSQ)	Survey	Percent change in score
Subjective	Workload Profile Index (WPI)	Survey	Average score for all categories (0 to 5 scale)
Subjective	NASA-Task Load Index (NASA-TLX)	MATB-II, pairwise comparison weights	TLX Score
Performance	TRACK	MATB-II	Pixels
Performance	SYSMON	MATB-II	% time the system is in a failure state
Performance	COMM	MATB-II	% time the system is in a failure state
Performance	RESMAN	MATB-II	Fuel units

Study Design

The study trials were presented in a counterbalanced order based on the MATB-II difficulty levels. The counterbalanced design was chosen as a way to minimize the effects from a possible learning curve and carryover effects. Because the total number of scenarios is 5, a complete counterbalanced order is for n=10, where the first 5 are presented in a balanced order, while the other 5 are presented as a mirrored order of the first 5. The full counterbalanced study was repeated 3 times, for a total of n=30 participants. The counterbalanced order is presented in Table 5.

Because participants were looking at the screen when each of the MATB-II scenarios was opened, a color word-based code was assigned to each scenario to prevent the participants from guessing the difficulty levels, and thus influencing the subjective responses.

Table 5: Counterbalanced Order

n	Order of MATB-II Scenarios				
1	1	2	5	3	4
2	2	3	1	4	5
3	3	4	2	5	1
4	4	5	3	1	2
5	5	1	4	2	3
6	4	3	5	2	1
7	5	4	1	3	2
8	1	5	2	4	3
9	2	1	3	5	4
10	3	2	4	1	5

Results

Thirty participants (male = 18, female = 12) completed the IRB-approved study (IRB approval: IRB2014-0499D). The mean age for participants was 24.3 years old, (standard deviation = 3.7 years).

All data were analyzed using SAS Studio version 9.4 (SAS Institute Inc., 2021) statistical software. Physiological measures, performance measures, and NASA-TLX were analyzed using the PROC MIXED code in order to reduce maximum likelihood estimates. The significance level was set at $\alpha = 0.05$. Due to the large amount of inherent noise in the physiological data, marginal significance as effects less than alpha = 0.10 is also reported. A t-test was used to compare the SSSQ score (PROC TTEST), and the WPI was analyzed using Friedman's test, a non-parametric test using PROC FREQ.

Physiological measures were analyzed using a within-subjects ANCOVA analysis, where the physiological baseline data were used as the covariate. All other measures (subjective and performance) were analyzed using within-subjects ANOVA. Post-hoc tests involved standard t-tests to determine where the differences between groups occurred.

Physiological Measures

The results from HRV (see Figure 18) were calculated using the pNN50 metric, which is the percentage of the N-N intervals greater than 50 milliseconds (more information about the pNN50 is in Chapter 1). pNN50 showed a significant difference among the MATB-II workload levels ($F(4, 34) = 5.47$, $p<0.005$). Tukey-Kramer post hoc revealed workload level 1 to be significantly different from levels 3 (p = 0.008), 4 (p = 0.003), and 5 (p = 0.004). Level 2 did not significantly differ from levels 3 (p = 0.291), 4 (p = 0.149), or 5 (p = 0.199). The asymptotic pattern associated with the redline of cognitive workload started showing at workload level 3, as there is no significant difference between levels 3 and 4 (p = 0.997), and levels 3 and 5 (p = 0.999). Levels 4 and 5 were not significantly different from each other (p = 0.999), further establishing the asymptote pattern.

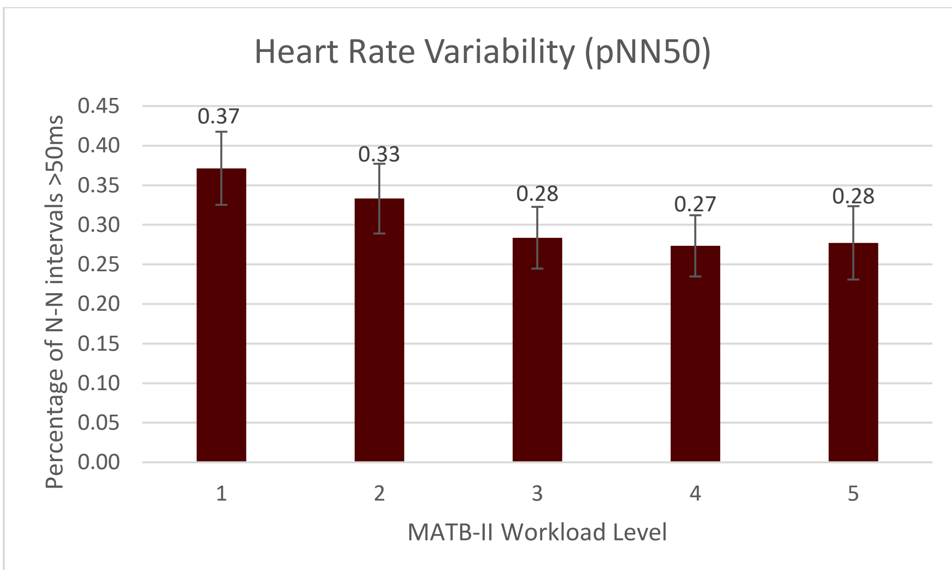


Figure 18: Heart rate variability, showing the pNN50 results. Error bars show standard error.

Skin conductance response (Figure 19) was not significant, but the results are marginally significant (F(4, 21) = 2.23, p = 0.079). Tukey-Kramer post-hoc results indicated level 1 is not statistically different from level 2 (p = 0.999), but significantly different from levels 3, 4, and 5. There is no significant difference among the other levels.

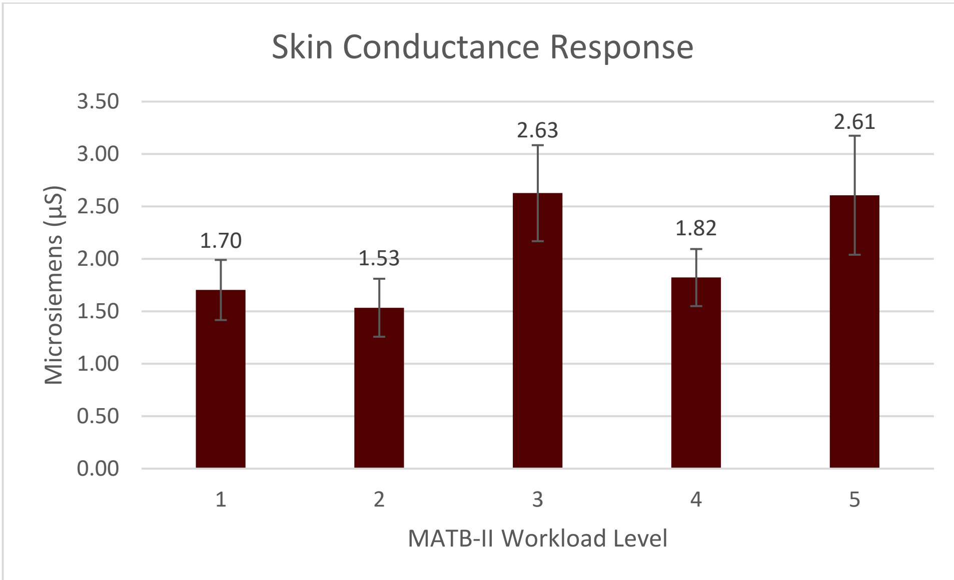


Figure 19: Skin Conductance Response. Error bars show standard error.

Heart rate (Figure 20) did not show any significant effects among the MATB-II workload levels (F(4, 38), p = 0.1068).

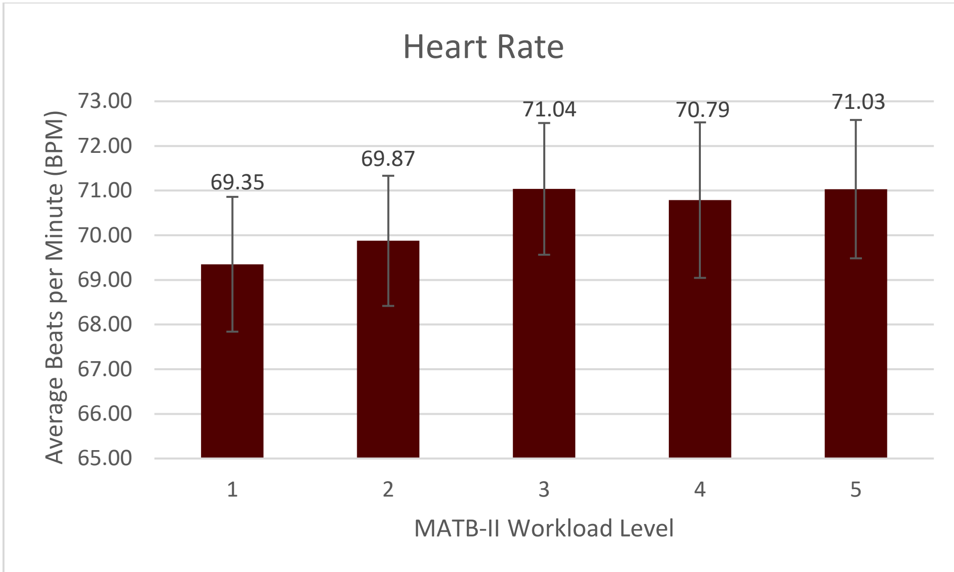


Figure 20: Average heart rate. Error bars show standard error.

Breathing rate was not significant (F(4, 37) = 1.97, p = 0.103).

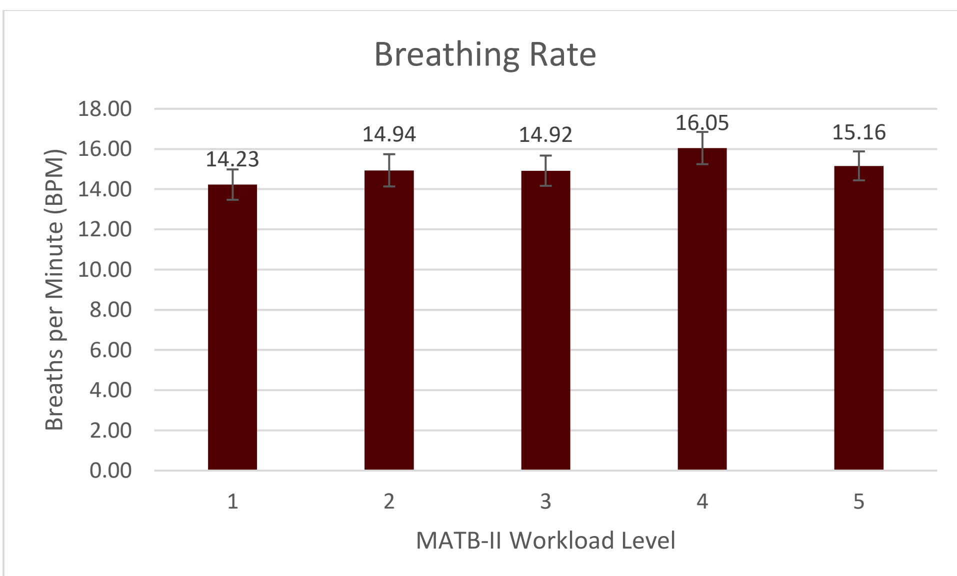


Figure 21: Average breathing rate. Error bars show standard error.

The alpha wave from EEG (Figure 22) result was significantly affected by higher workload levels ($F(4, 37) = 2.69$, $p = 0.035$). Tukey-Kramer post hoc results indicated that level 1 is not significantly different from levels 4 ($p = 0.1605$) and 5 ($p = 0.293$). Level 2 is not significantly different from workload level 3 ($p = 0.997$), and workload levels 4 and 5 were also not statistically different ($p = 0.998$). Levels 2 and 3 were not significantly different, which was the same case with workload levels 4 and 5.

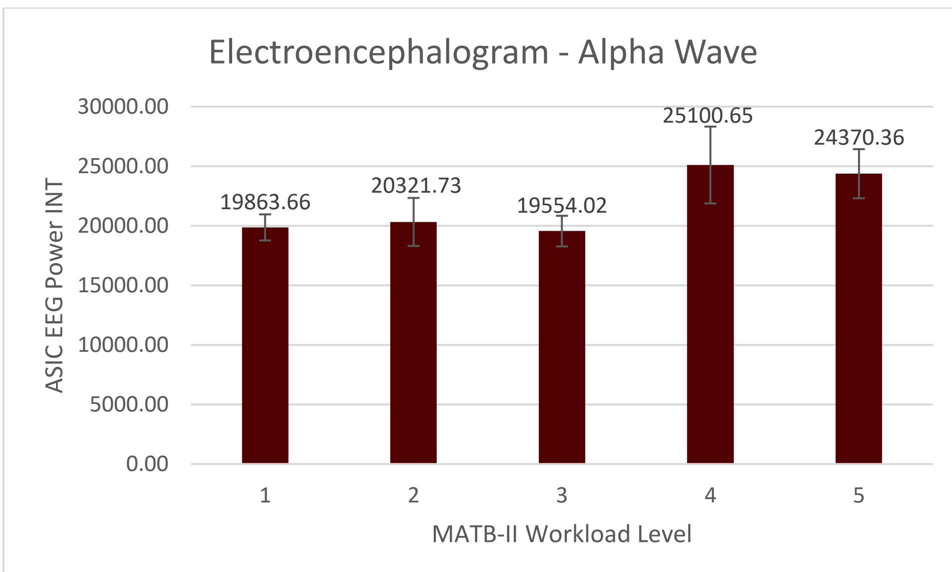


Figure 22: Electroencephalogram alpha wave average. Error bars show standard error.

Subjective Measures

The overall score change for the SSSQ was 0.031, which was calculated using the formula in the Methods section. A t-test determined that the overall score change was not significantly different from 0.

NASA-TLX results (Figure 23) showed the workload level is significantly affected by increased cognitive workload levels ($F(4, 37)= 77.49$, $p = <0.001$).Tukey post-hoc results found level 1 significantly different from workload levels 3 ($P < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 2 was also significantly different from workload levels 3 ($P < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 3 was not significantly different from workload level 4 ($p = 0.999$), but it was significantly different from workload level 5 ($p = 0.005$). Level 4 was also found to be significantly different from level 5 ($p = 0.01$).

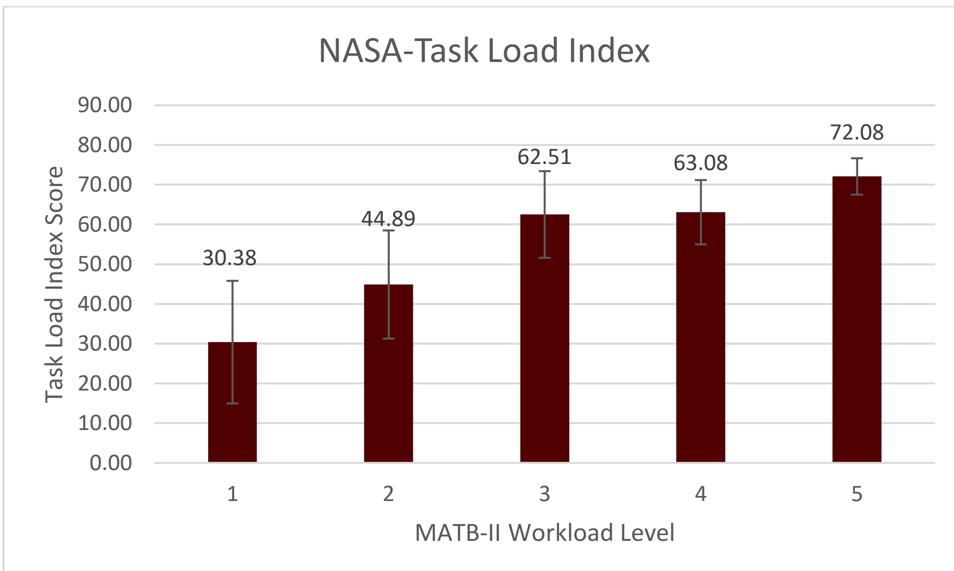


Figure 23: NASA Task Load Index. Error bars show standard error.

Because the data obtained from the Workload Profile Index questionnaire are integer values (discrete data), this dataset was analyzed using Friedman's test, a non-parametric statistical test used to detect differences in treatment across different tests attempts. Results indicated WPI (Figure 24) gives a pattern of results that is concurrent with other measures, is a predictor of the difficulty in the workload levels ($\chi^2 = 35.29$, $p < 0.001$).

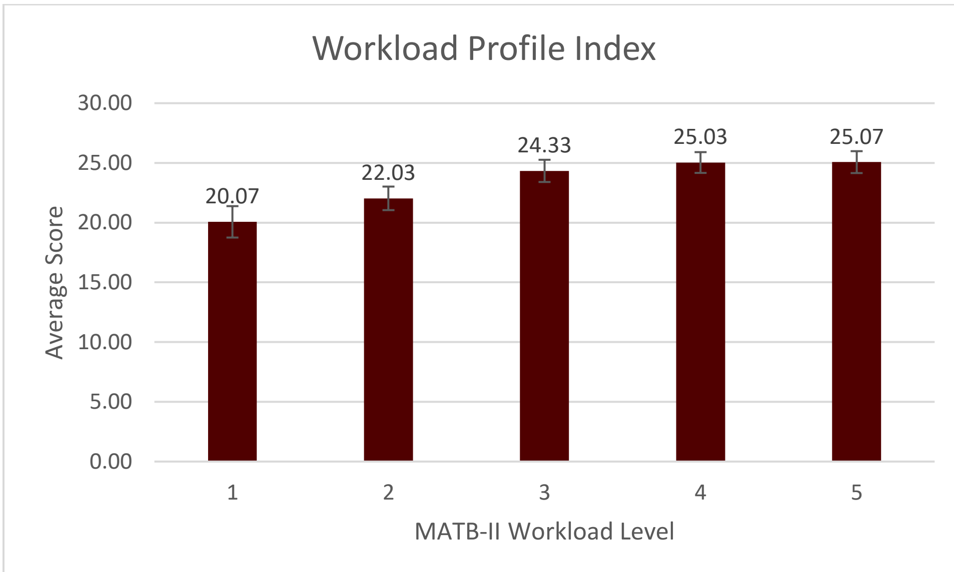


Figure 24: Workload profile index. Error bars show standard error.

Performance Measures

Performance was calculated for each of the MATB-II tasks, as described in Table 3.

Results for the tracking task (TRACK) (Figure 25) indicated performance decreased as the workload levels increased in difficulty ($F(4,116)=39.52$, $p<0.001$). Tukey post-hoc tests found level 1 to be significantly different from levels 3 ($p<0.001$), 4 ($p<0.001$), and 5 ($p<0.001$). Level 2 is significantly different from levels 4 ($p<0.001$), and 5 ($p<0.001$). There was no significant difference between levels 3, 4 and 5.

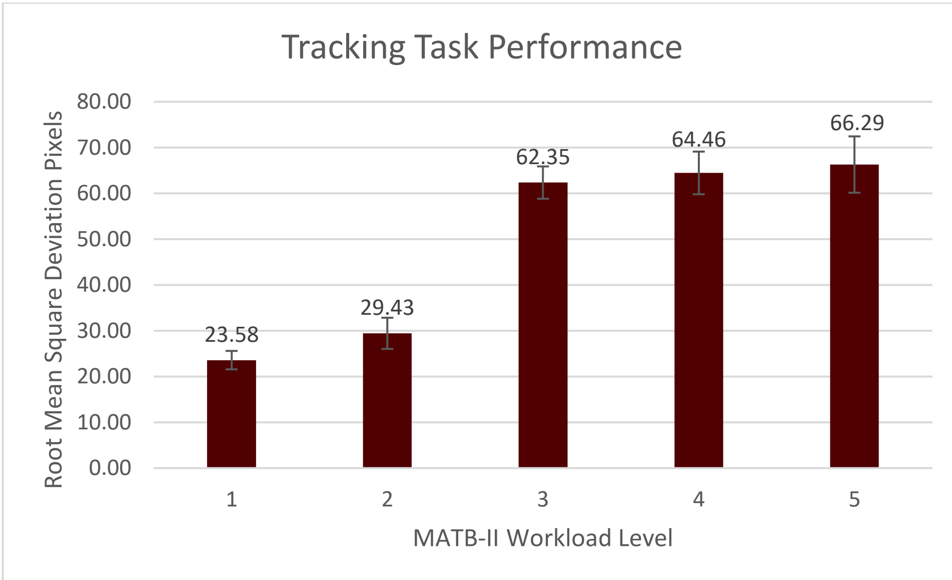


Figure 25: Tracking performance. Error bars show standard error.

Results for the resource management (RESMAN) task (Figure 26) showed the workload level had a significant effect on the deviation from the fuel target ($F(4,116) = 16.54$, $p<0.001$). Tukey post-hoc tests showed that level 1 is significantly different from level 3 ($p = 0.0081$), level 4 ($p = 0.0034$), and level 5 ($p<0.001$). Level 2 is significantly different from level 3 ($p = 0.0016$), level 4 ($p = 0.0006$), and level 5 ($p<0.001$). Level 3 is only significantly different from level 5 ($p<0.001$). There was no significant difference between levels 4 and 5.

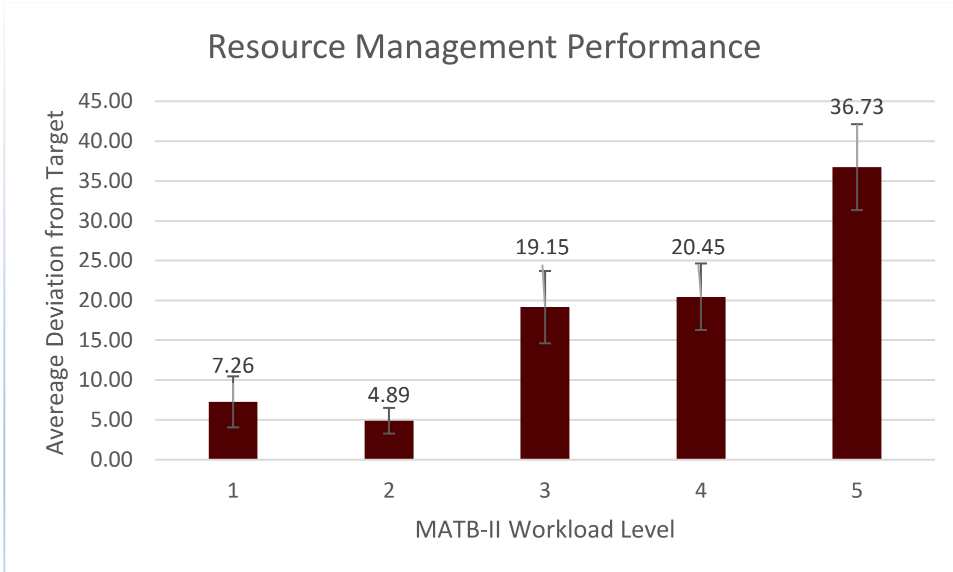


Figure 26: Resource management performance. Error bars show standard error.

Results for the system monitoring (SYSMON) task (Figure 27) found a significant effect of the MATB-II workload level on the percentage of time the system was in a failed state ($F(4,116) = 37.42$, $p<0.001$). A Tukey post-hoc test found level 1 to be significantly different from level 2 ($p = 0.0024$), level 3 ($p<0.001$), level 4 ($p<0.001$), and level 5 ($p<0.001$). Level 2 is significantly different from level 3 ($p<0.001$), level 4 ($p<0.001$), and level 5 ($p<0.001$). There was no significant different between levels 3, 4, and 5.

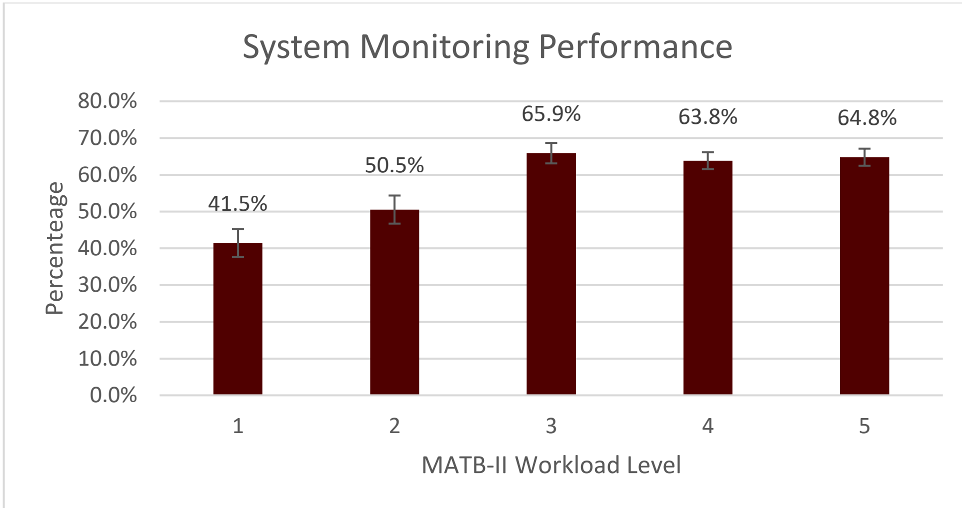


Figure 27: System monitoring performance. Error bars show standard error.

Results for the communication (COMM) task indicated the workload level had a significant effect on the percentage of time the communication system was in a failed state ($F(4,116) = 7.77$, $p<0.001$). A Tukey post-hoc test found level 1 to be significantly different from level 3 ($p = 0.0242$), and from level 5 ($p = 0.0020$). Level 2 is significantly different from level 3 ($p<0.001$), level 4 ($p = 0.0013$), and level 5($p<0.001$). There was no significant difference between levels 3, 4, and 5.

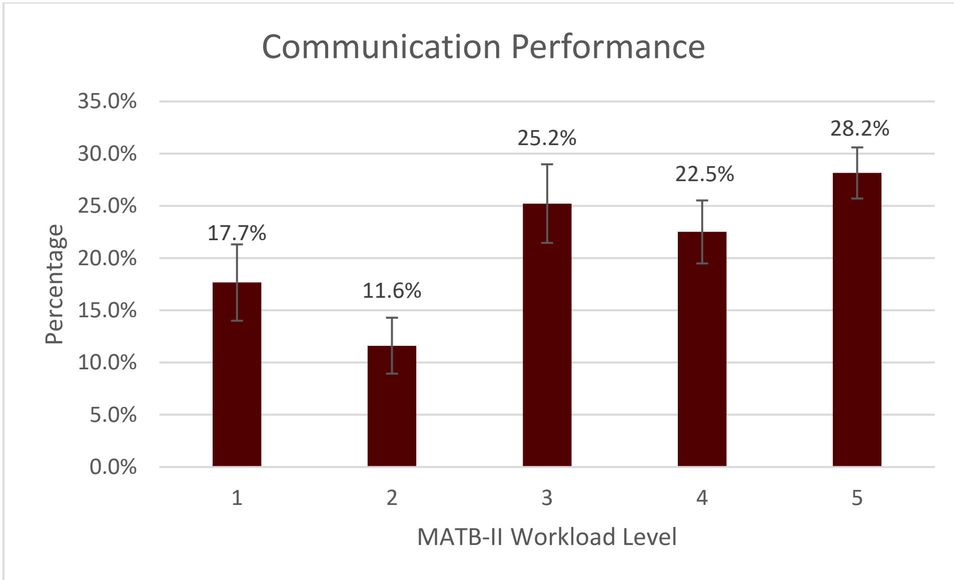


Figure 28: Communication performance. Error bars show standard error.

Discussion

Several physiological measures were analyzed in this study to determine their sensitivity near the redline of cognitive workload. Two of the measures, heart rate and breathing rate, did not show statistically significant differences among the workload levels, however, heart rate and breathing rate showed an asymptotic pattern, which was expected as an indicator that the participants approached their redline thresholds. From the proposed definition in chapter 1, physiological measures were expected to stop fluctuating and reach an asymptotic pattern.

Skin conductance response was not statistically significant, but similar to the other measures, it also shows the asymptotic pattern. Only EEG (alpha wave) and HRV were statistically significant. The asymptotic pattern found in this study is similar to the event-related potential asymptotic pattern experienced by participants under increasing workload conditions (Vogel & Machizawa, 2004).

The alpha wave increase can be interpreted as a lack of cognitive engagement (Kamzanova et al., 2011; Offenloch & Zahner, 1990). However, because levels 4 and 5 imposed very difficult workload levels on the participants, another way to interpret these results would be to analyze the task shedding behavior of the participants. While this data was not collected during the study, future studies could investigate the different strategies participants use to try to maintain their performance.

Based on ANS activity, the SNS will remain activated to attend to the "threat" presented in the fight or flight response. PSNS activity should remain low while the body prepares for the "threat," in this case, each of the workload tasks. Thus, there was elevated SNS activity in all of the scenarios, based on the changes, but the asymptote pattern presented by levels 3, 4, and 5 indicated that physiologically, there was no more change. This is the point where the participants approached their redline threshold.

The physiological results were supported by the results from subjective measures and performance. The results from SSSQ showed no difference between the pre- and post-test scores. This result indicated that stress was not a factor in the effects analyzed in the study, rather, the changes observed come mainly from the workload.

NASA-TLX show that participants felt levels 3 and 4 to be very similar, while level 5 was rated higher. However, the scenarios were designed to get progressively more difficult. A possible explanation is that the experimenter and pilot test subjects were MATB-II expert users, while participants were novel users. When participants were overwhelmed by the amount of imposed cognitive workload, they may not have been able to discriminate the difference between the last 3 difficulty levels, as their mental resources were already heavily in use. Experienced users, or users that are more familiar with the MATB-II interface, such as pilots, may provide different results.

Subjectively, participants felt that the workload kept increasing until the third workload level, based on NASA-TLX results. At the third level, physiological measures started to stabilize into the asymptote pattern, which is the workload level at which performance also started to degrade. The NASA-TLX levels 3 and 4 were rated by participants as very similar, while participants felt that level 5 was slightly more difficulty. While the experimenter and pilot subjects did iterative testing to ensure an equivalent magnitude of change between the levels, results validate the design of the 5 levels of difficulty, and help to make the case about physiological measures stabilizing and reaching a "plateau" pattern when the person is approaching their redline.

Among one of the limitations of the study, the use of MATB-II was perhaps the most limiting one. While participants were trained on how to correctly operate the software, MATB-II was designed to incorporate tasks similar to those found in an aircraft cockpit (Comstock Jr & Arnegard, 1992), which not many participants experience in their daily lives. While the scenarios were developed by expert MATB-II users to ensure an equivalent magnitude of change, participants still rated the last three difficulty scenarios as being very similar in difficulty. Future studies could use other, more relatable tasks to similar to daily activities and gather a clear understanding of the cognitive workload demands experienced in everyday life by participants.

Additionally, of all the possible physiological measures, only a few of the physiological measures were collected. Individual differences are prevalent when analyzing physiological measures, and while considerations were taken, such as including the baseline, future studies need to address this. Additionally, because the redline threshold is a fairly new concept, future studies, especially those in naturalistic setting, may need to account for external factors, such as expertise and motivation, that could influence how much cognitive workload a person can withstand before approaching their redline.

Future studies could incorporate additional measures, such as eye-related measures. Pupillometry has also been linked to changes in workload (Backs & Walrath, 1992; Iqbal et al., 2004). It would be interesting to compare electrodermal responses and pupillometry, as these measures tend to react faster than others to changes in stimuli.

It is important that future studies also compare the results among all three methods of data collection (physiological measures, subjective measures, and performance), especially when someone is near their threshold. Understanding how these measures compare can depend the understanding of how overload can occur to implement ways to minimize performance decrements.

CHAPTER III

REDLINE OF COGNITIVE WORKLOAD MEASURES

Introduction

The previous study investigated which physiological measures are more sensitive to changes in workload when the person is approaching their redline of cognitive workload. Based on the results from study 1, cardiovascular measures, particularly the pNN50 measure of HRV, and electrodermal activity are the most sensitive in detecting when the person is approaching their redline based on the asymptotic pattern. This second study builds on these results by introducing a new measure, pupillometry, and investigating the changes among physiological measures, performance, and subjective measures as the person experiences increasing workload levels and approach their redline.

One of the first study's limitations is that it mostly considered physiological measures based on electrical signals (e.g. cardiovascular measures, EEG, and EDA, were collected using electrical signals); a limitation which this study sought to address.

One question remaining from the previous findings was to determine if multiple physiological measures would provide similar results to identify the redline, and if performance and subjective measures could be correlated to further investigate the association between performance decrements, physiology, and workload to enhance the robustness of the redline of cognitive workload identification.

While different physiological indicators of the redline of cognitive workload are affected by sympathetic arousal, these measures express different sensitivities to different stressors or stimuli; thus, the physiological patterns indicative of the stimuli change will be observed over different time windows. For example, HRV spectral analysis is more meaningful when it is analyzed over a longer time window, as the recommendation is to analyze at least 5 minutes (Task Force of the European Society of Cardiology, 1996), though shorter time windows have been studied as well, such as 30 seconds (Salahuddin et al., 2007; Yentes et al., 2013). Thus, a measure that shows faster response time was chosen for this second study, pupillometry, a measure of change in pupil diameter.

For the aforementioned reasons, pupillometry was chosen as an additional measure in the study. Pupillometry is the indicator of pupil dilation (i.e., widening of the pupil), and previous studies have associated it with changes in mental workload (Beatty, 1982), where larger mean values are associated with more complicated tasks (Iqbal et al., 2004). The pupil dilates rapidly, in the order or milliseconds, when a task imposes cognitive workload. Pupil changes can be observed on up to 0.5 mm (Klingner et al., 2008), and the pupil tends to constrict quickly to normal size after the completion of the task that triggered the dilation. Pupil dilation can also be measured in a non-intrusive way, and can provide real-time information on the cognitive workload the person is experiencing. However, pupil dilation is also sensitive to light changes and the gaze angle (Pomplun & Sunkara, 2003). Thus, the experiment must be performed in a controlled environment.

While pupillometry was incorporated into the study, EDA was kept to serve as a comparison point, since results from the previous study show that EDA is sensitive to detect changes in workload, and also shows an asymptotic pattern when the person is approaching their redline. Additionally, EDA is only modulated by the SNS (Boucsein, 2012), while pupillometry is enervated by both branches of the SNS (Mathôt, 2018).

This study also incorporated other measures of cognitive workload, particularly subjective ratings and performance-based measures. Similar to the previous study, NASA-TLX ratings, WPI, and SSSQ were used, while and performance measures were derived from the software used. All these measures were used to estimate when the person is approaching their redline of cognitive workload.

This study aims to answer the second research question: **To what extent do changes in these sensitive physiological variables compare with other standard measures of task performance and subjective measures?** Two physiological measures were collected throughout the study, EDA and pupillometry. EDA was selected based on the results from the previous study, while pupillometry was incorporated as a new measure. Subjective ratings and performance-based measures were also collected to serve as a "ground truth" of when participants approached their redline, based on previously established research finding a relationship between subjective measures and performance decrements (Colle & Reid, 2005; Reid & Colle, 1988).

Findings from this research can add to the results found from the first study, by identifying the association among subjective measures, performance decrements, and physiological measures. The information obtained from this study can be incorporated in future naturalistic studies where physiological measures can be collected in real-time to identify when someone is approaching their redline before the performance degradation occurs.

Methods

Twenty-two participants, who were at least 18 years old, completed the IRB-approved study (IRB approval: IRB2014-0499D). Other inclusion criteria consisted of participants having normal or corrected-to-normal vision, and no physical impairments that prevented them from using a joystick, mouse, and keyboard.

Participants completed a series of tasks set in a multitasking environment, similar to those tasks required to pilot an aircraft. Through the study, physiological data were collected to infer how sensitive these measures are to changes in workload. The difficulty levels were controlled, and presented in a counterbalanced order.

Procedure

Participants were briefed on the study when they arrived at the laboratory, and signed the informed consent form. Then, participants completed a background questionnaire, which sought information about their experience with videogames and flight simulators, caffeine consumption, and demographics. After completing the background questionnaire, participants filled the pre-experiment section of the Short Stress State Questionnaire (SSSQ) (Helton, 2004).

Then, with help from the experimenter, participants proceeded to put on the physiological devices, which consisted of Shimmer GSR (Realtime Technologies Ltd, 2017) and Pupil by Pupil Labs (PupilLabs, 2019), as shown in Figure 29. The experimenter helped the participants to put on the devices. The experimenter then calibrated the devices to ensure the proper streaming and storing of data.

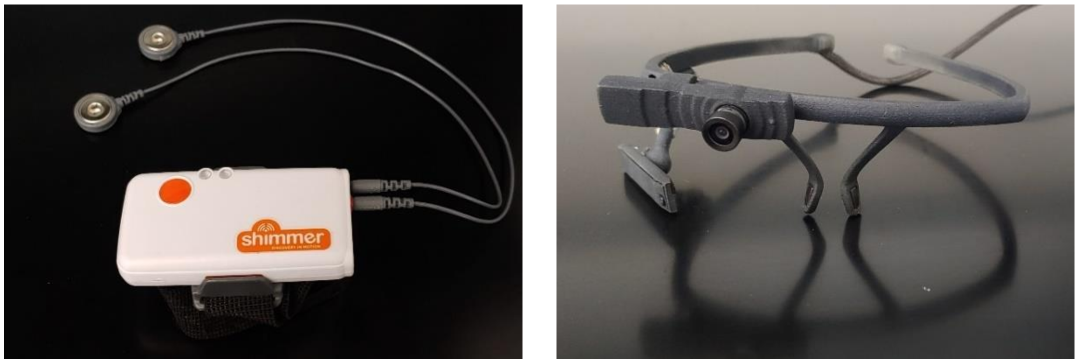


Figure 29: Physiological devices used in the study. The Shimmer GSR system is on the left, and Pupil by Pupil Labs on the right.

Then, the data corresponding to the physiological baseline were collected (see Chapter 1 for more details), which were used as covariate in the statistical analysis. The baselining process consisted of 10 minutes of paced breathing which allows for the body pressures to stabilize (Kergoat & Faucher, 1999), instructions for which were presented on the Paced Breathing app (TrexLLC, 2014), consisting of a pattern of inhalations and exhalations. The paced breathing followed a breathing ratio of 1:2 inhalation to exhalation, which has been shown to decrease blood pressure and heart rate (Modesti et al., 2010). Nature sounds (johnnielawson, 2013) played through noise-cancelling earphones to provide a relaxing environment. The objective of the baselining procedure is to obtain physiological data when then person is in a relaxed state with minimal cognitive activity, which can serve to compare when the person is relaxed or in a state of high cognitive activity, thus why it is considered as a covariate. Both physiological measures (pupillometry and EDA) investigated in this study were collected throughout the baselining procedure.

Then, participants proceeded with the study, which consisted of the same MATB-II scenarios and procedure similar to the study presented in Chapter 2.

Participants completed the Multi-Attribute Task Battery (MATB-II) (Santiago-Espada et al., 2011) training to reduce the likelihood of learning effects (Prinzel et al., 2000). The training consisted of an explanation on how the software works using a PowerPoint presentation (see Appendix B), and the explanation that all four tasks were equally important. The PowerPoint slides consisted of screenshots of the software, and the instructions were provided verbally. Participants were encouraged to ask any questions they may have. Then, participants had to complete 2 MATB-II training scenarios, each one was 3-minutes long. The training difficulty was similar to the medium and hard scenarios, which are explained in Table 2 from Chapter 2. At the end of the second training scenario, participants confirmed that they did not have any other questions regarding the functioning of MATB-II before proceeding with the study.

Participants then proceeded to take a 3-minute resting break, following a similar procedure to the baselining data collection, following paced breathing (TrexLLC, 2014) exercises and listening to nature sounds (johnnielawson, 2013). The purpose of the resting break was to minimize the risk of a carryover effect between scenarios, by allowing physiological measures to revert back to the relaxed state, where the participant experienced minimal cognitive activity. This way, any changes reflected while performing the study were due to the current MATB-II scenario and not to the previous one.

Participants then completed 5 MATB-II scenarios of varying workload levels, ranging from very easy to very difficult (see Table 1 on Chapter 2) presented in a counterbalanced order, and each scenario was followed by the NASA-TLX and WPI questionnaires. Two subjective workload assessments were collected to capture different information. For example, NASA-TLX provides diagnostic information based on 6 workload dimensions and an aggregate workload score (Hart, 2006; Hart & Staveland, 1988), while WPI collects data on which cognitive resources were used (Rubio et al., 2004) (see Chapter 1 for more information on these measures).

After completing the fifth and final scenario, participants proceeded to complete the NASA-TLX pairwise comparison survey, which consisted of ranking the 6 workload categories presented in pairs, and ranking the one they thought provided the highest workload (Hart & Staveland, 1988). These weights were then used to obtain the aggregate workload score. After the NASA-TLX pairwise comparison, participants completed the post-test section of the SSSQ (Helton, 2004). Participants then removed the physiological devices to complete the study. A summary of the full study procedure is shown in Figure 30.

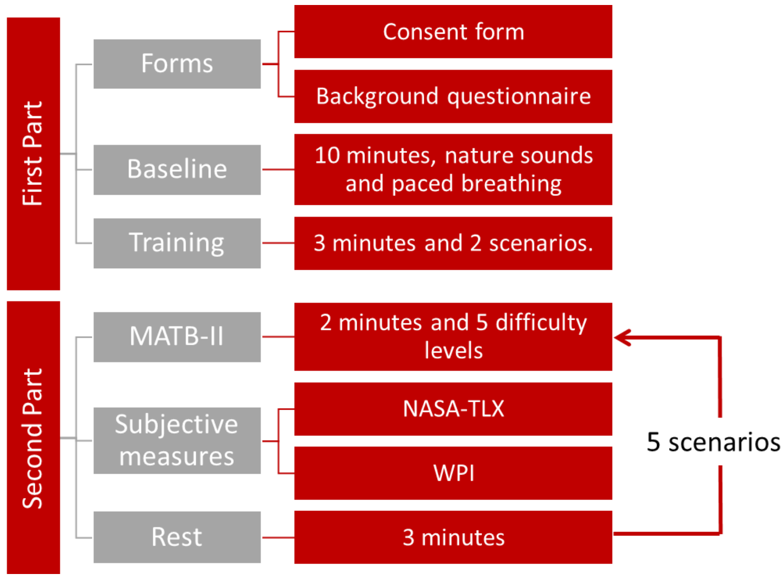


Figure 30: Experimental procedure.

Apparatus

This section provides an overview of the different physiological sensors, questionnaires, and software used during the data collection. A summary of all the dependent measures collected throughout the study is provided in Table 6.

Physiological Devices

Two devices (Shimmer GSR and Pupil by Pupil Labs) were used to collect physiological measures in real-time while participants performed the study activities, including the physiological baselining procedure, experimental trials, questionnaires, and resting periods. Before starting the study activities, the devices were calibrated following the standard for each device to ensure proper data streaming and storing.

Electrodermal activity (EDA) was collected using the Shimmer GSR system (Realtime Technologies Ltd, 2017), as shown in Figure 31. The Shimmer GSR system consisted of an amplifier, worn around the wrist with an adjustable strap, and two leads connected to isotonic gelled electrodes. The electrodes were placed in the thenar and hypothenar eminences of the palm (Dawson et al., 2017). The amplifier is worn around the wrist with an adjustable strap to prevent discomfort.

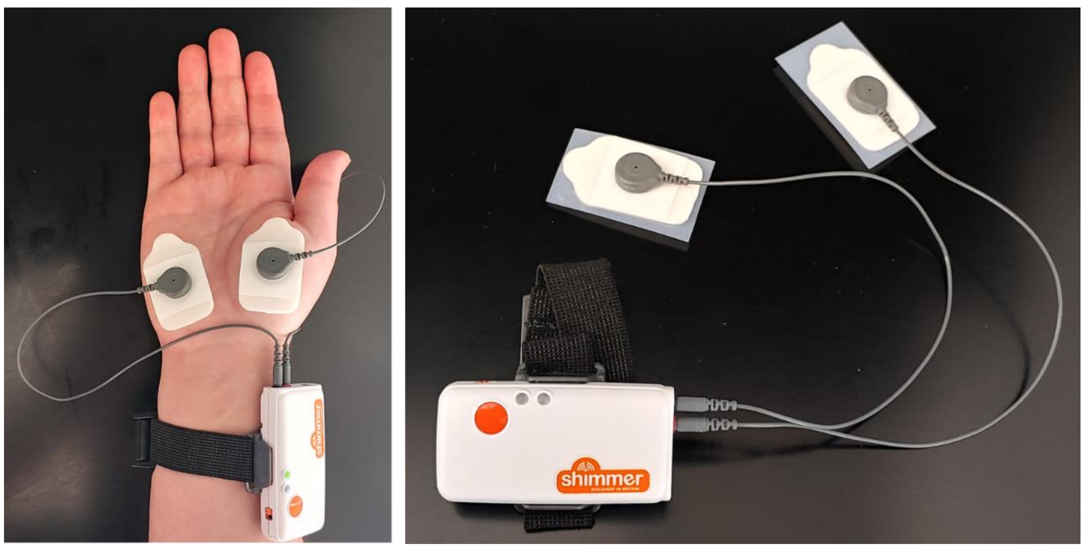


Figure 31: The Shimmer GSR system. On the left is the system with the correct lead placement on the thenar and hypothenar palm eminences. The figure on the right shows the amplifier with the leads connected to the pre-gelled electrodes.

The ShimmerGSR system collected the EDA data from the dominant hand, even though EDA is traditionally collected from the non-dominant hand (Dawson et al., 2017). In this study, the non-dominant hand controlled the mouse and keyboard, and therefore, had more movement, increasing the risk for motion artifacts. Extensive testing by the experimenters found less motion artifacts when the Shimmer GSR systems was worn on the dominant hand, as it was only controlling the joystick. Collecting EDA from the dominant hand in this study decreases the risk of motion artifacts and missing data values.

The raw GSR data collected was filtered and processed using Ledalab (Karenbach, 2005), a Matlab plug-in. The skin conductance response (SCR) was extracted after the filtering process, and these data were later used in the statistical analysis.

Pupil dilation (pupillometry) was collected using Pupil by Pupil Labs (PupilLabs, 2019), as shown in Figure 32. The Pupil device was worn like a pair of glasses, and it had a small camera that is adjusted to capture the participant's pupil.

Figure 32: Pupil by Pupil Labs.

Because each person's face anthropometry is different, the glasses had to be adjusted and calibrated before each study. The first adjustment consisted of the experimenter adjusting the camera facing the participant's eye to better capture the pupil. A red circle appeared when the camera was correctly tracking the pupil (see Figure 33). The second step consisted of calibrating the device by looking into the red dot at the center of a bullseye target, composed of black and white concentric circles (see Figure 33). The red dot blinked and turned green before moving to a different area of the screen. This calibration procedure was repeated before each experimental trial to ensure the device was still properly calibrated despite head movement.

Figure 33: On the left, Pupil by pupil labs. The camera that faces the pupil is shown. On the right, is the target that was used to calibrate the Pupil.

The video data captured by Pupil was processed using Pupil Play, a software developed by Pupil Labs. The change in pupil size was then extracted for each time segment, and these data were used in the statistical analysis.

Software to Manipulate Mental Workload

Similar to study 1 (Chapter 2), the Multi-Attribute Task Battery-II (MATB-II) was used to manipulate mental workload conditions. Once more, all four tasks (SYSMON, TRACK, COMM, and RESMAN) were used. See chapter 2 for the full information.

Subjective Questionnaires

Several subjective questionnaires were used to capture different parameters. Similar to the first study, SSSQ, NASA-TLX, and WPI were used. See chapter 2 for more information.

Experimental Variables

The dependent variables collected in the study are based on the three different types of data collected in this study: physiological, subjective, and performance-based measures. A summary of the variables is listed in Table 6.

Table 6: Dependent Variables

Workload Measure Type	**Measure Collected**	**Device or Method Collected**	**Unit**
Physiological	Pupil size	Pupil by Pupil Labs	Pixels
Physiological	Electrodermal activity (EDA), (skin conductance response)	Shimmer GSR	Microsiemens
Subjective	Short Stress State Questionnaire (SSSQ)	Survey	Percent change
Subjective	Workload Profile Index (WPI)	Survey	Average score for all categories (on a scale from 0 to 5)
Subjective	NASA-Task Load Index (NASA-TLX)	MATB-II, pairwise comparison weights	TLX Score
Performance	TRACK	MATB-II	Pixels
Performance	SYSMON	MATB-II	% time the system is in a failure state
Performance	COMM	MATB-II	% time the system is in a failure state
Performance	RESMAN	MATB-II	Fuel units

Study Design

Similar to the first study presented in Chapter 2, the study trials were presented in a counterbalanced order based on the difficulty levels of the MATB-II scenarios. The counterbalanced order was implemented to reduce the likelihood of learning and carryover effects. The completed counterbalanced order is for n=10, where the first 5 trials were presented in a balanced order, and the other 5 are presented in a mirrored order. The counterbalanced order is presented in

Table 5 (Chapter 2). Each scenario was saved with a color-based code to prevent participants from guessing the difficulty levels, and thus having an influence on the subjective measures.

Results

Twenty-two participants (mean age = 26.5 years old, standard deviation = 4.6 years, 13 male and 9 female) completed the IRB-approved study (IRB approval: IRB2014-0499D).

All data were analyzed using SAS Studio version 9.4 (SAS Institute Inc., 2021) statistical software. The PROC MIXED code was used to analyze physiological measures, performance, and NASA-TLX, in order to reduce the maximum likelihood estimates. For the analysis, the significance level was set at $\alpha = 0.05$. The SSSQ was analyzed using PROC TTEST, and the WPI was analyzed using PROC FREQ.

A within-subjects ANCOVA was used for the analysis of physiological measures (EDA and pupillometry), where the data collected during the physiological baselining procedure were used as the covariate. Additionally, post-hoc tests involved in standard t-tests were used to determine the main effect between the scenarios.

Physiological Measures

The results from pupil size (Figure 34) found a significant effect of the MATB-II level on the pupil size in pixels (F (4, 29 = 8.94, p<0.001). Post-hoc results using the least squares means found no difference between levels 1 and 2 (p = 0.1418). Level 1 was significantly different from levels 3 (p = 0.0078), 4 (p<0.001), and 5 (p<0.001). Level 2 was not significantly different from level 3 (p = 0.2066), but was significantly different from levels 4 (p = 0.0031) and 5 (p = 0.0020). Level 3 was significantly different from levels 4 (p = 0.0485) and 5 (p = 0.0366). Level 4 was not significantly different from level 5 (p = 0.9603).

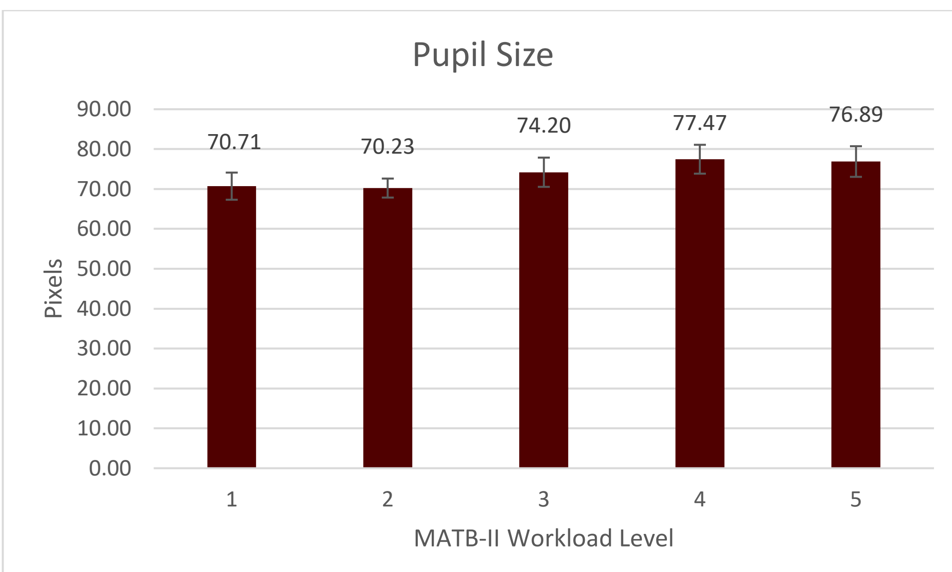


Figure 34: Pupil size average. Error bars show standard error.

The results from skin conductance response (Figure 35) showed that there was no interaction between the workload level and electrodermal activity (F (4, 33 = 0.40, p = 0.8044).

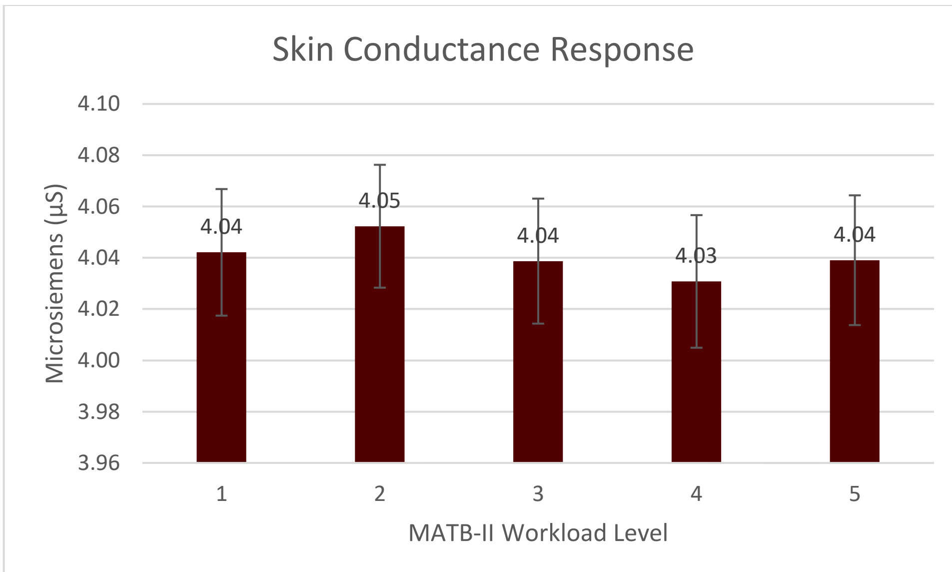


Figure 35: Skin conductance response. Error bars show standard error.

The results from skin conductance level (Figure 36Figure 35) showed that workload had no significant impact (F (4, 33 = 0.46, p = 0.7630).

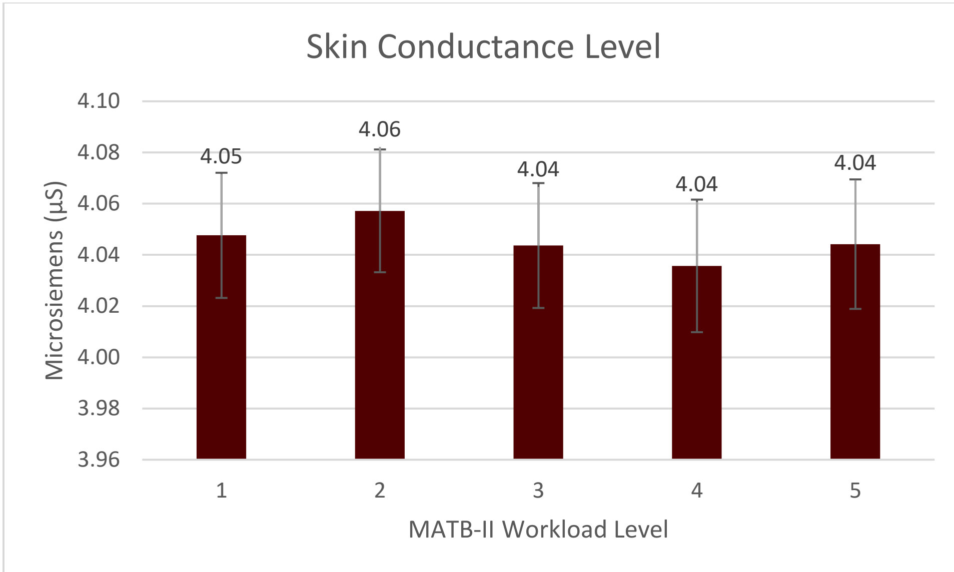


Figure 36: Skin conductance level. Error bars show standard error.

Subjective Measures

The overall SSSQ change score was 0.077, calculated using the formula described in Chapter 2. A t-test score determined that the change score was not significantly different from 0.

NASA-TLX results (Figure 37) showed a significant effect of the MATB-II workload level on the TLX score (($F(4, 84)= 53.33$, $p = <0.001$). Post-hoc results using the least squares means found that level 1 is significantly different from workload levels 3 ($P < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 2 was also significantly different from workload levels 3 ($P < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). There was no significant different from levels 3, 4, and 5.

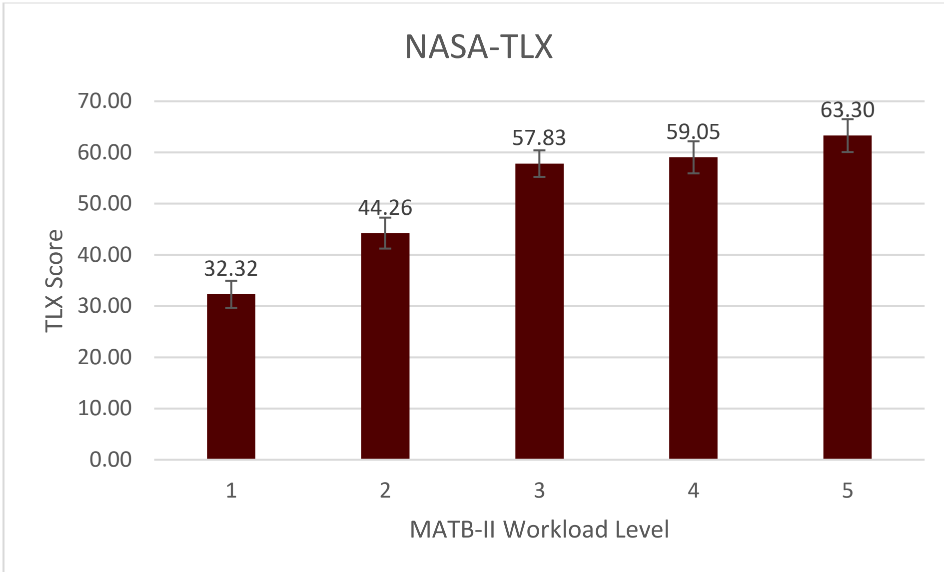


Figure 37: NASA Task Load Index. Error bars show standard error.

The Workload Profile Index questionnaire (Figure 38) collected data in the form of integer values; thus, this dataset was analyzed using a non-parametric test, Friedman's test, which is used to detect differences in treatments across different groups. Results showed that the MATB-II workload level has a significant effect on the WPI score (($\chi 2 = 13.0451$, $p = 0.01$).

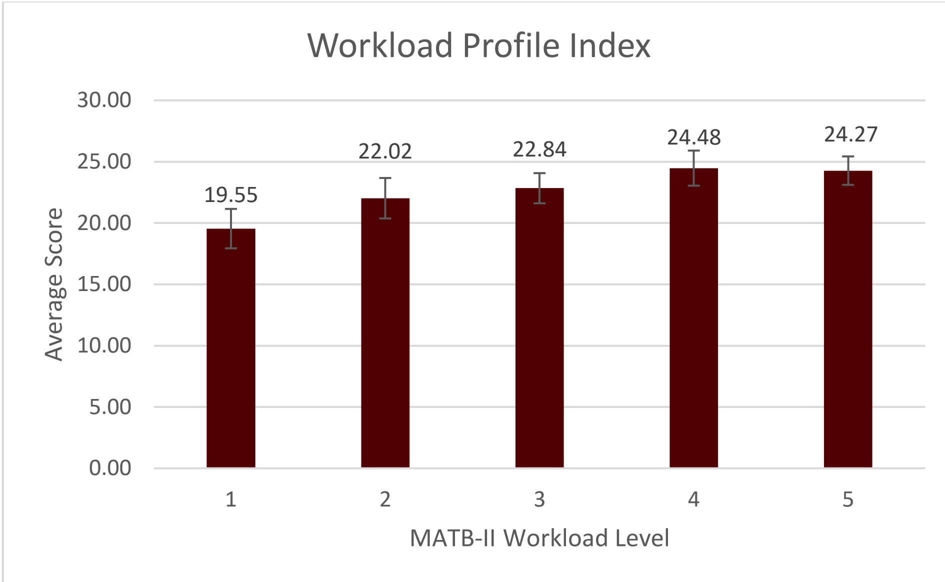


Figure 38: Workload profile index. Error bars show standard error.

Performance Measures

Performance was calculated for each of the MATB-II tasks, as described in Chapter 2.

Results for the tracking task (TRACK) (Figure 39) showed the MATB-II workload level had an impact on the performance. Performance got progressively worse based on the workload level ($F(4, 83)= 63.77.52$, $p<0.001$). Least squares means post-hoc results found level 1 to be significantly different from level 2 ($p = 0.0239$), and from levels 3 ($p<0.001$), 4 ($p<0.001$), and 5 ($p<0.001$). Level 2 is significantly different from levels 3 ($p<0.001$), 4 ($p<0.001$), and 5 ($p<0.001$). Level 3 is not significantly different from level 4 ($p = 0.1083$), but is significantly different from level 5 ($p = 0.0230$). Level 4 is not significantly different from level 5 ($p = 0.4846$).

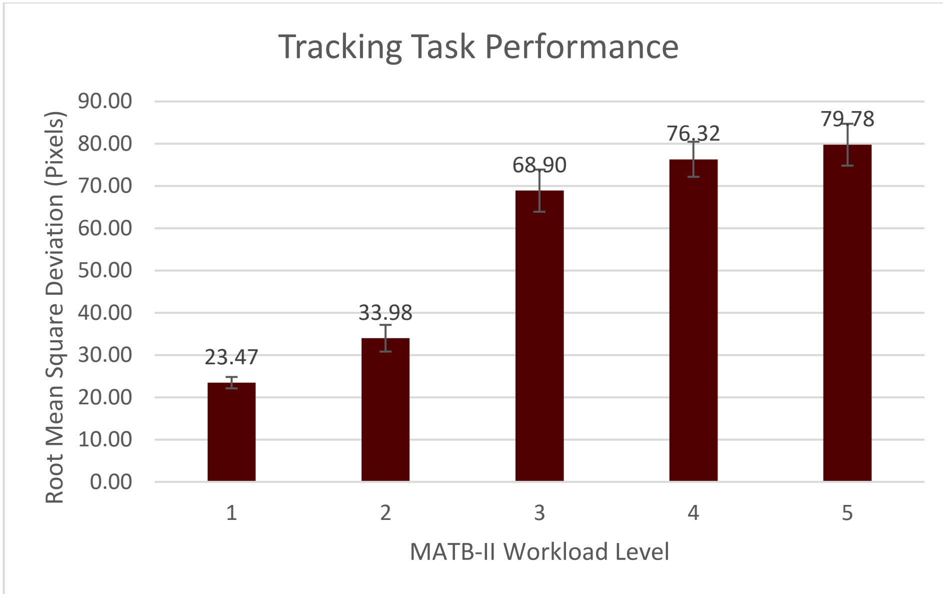


Figure 39: Tracking performance. Error bars show standard error.

Results for the resource management task (RESMAN) (Figure 40) showed the MATB-II workload level had an impact on the performance ($F(4, 83)= 7.42$, $p<0.001$). Least squares means post-hoc results found no significant difference between levels 1 and 2 ($p = 0.6281$), levels 1 and 3 ($p = 0.1627$), but found significant difference between levels 1 and 4 ($p = 0.0112$) and 1 and 5 ($p < 0.001$). Level 2 is not significantly different from level 3 ($p = 0.3590$) but level 2 is significantly different from levels 4 ($p = 0.0379$) and 5 ($p < 0.0001$). Level 3 is not significantly different from level 4 ($p = 0.2386$), but is significantly different from level 5 ($p = 0.0009$). Level 4 is significantly different from level 5 ($p = 0.0268$).

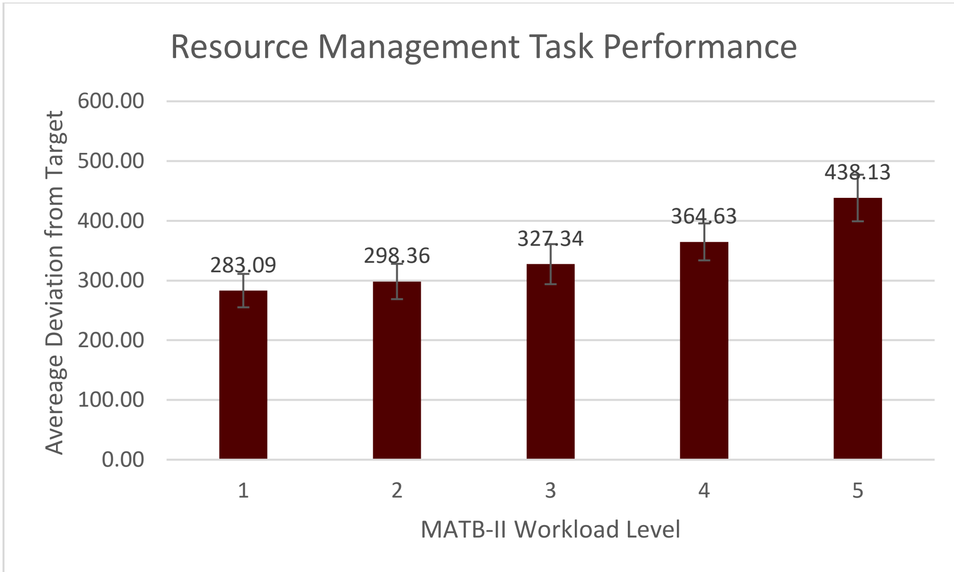


Figure 40: Resource management performance. Error bars show standard error.

Results for the system monitoring (SYSMON) (Figure 41) showed the MATB-II workload level had an impact on the performance ($F(4, 83)= 24.83$, $p<0.001$). Least squares means post-hoc results found no significant difference between levels 1 and 2 ($p = 0.4166$), but found significant difference between levels 1 and 3 ($p < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 2 was significantly different from levels 3 ($p < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 3 was not significantly different from level 4 ($p = 0.5638$) nor from level 5 ($p = 0.9326$). Level 4 was not significantly different from level 5 ($p = 0.5121$).

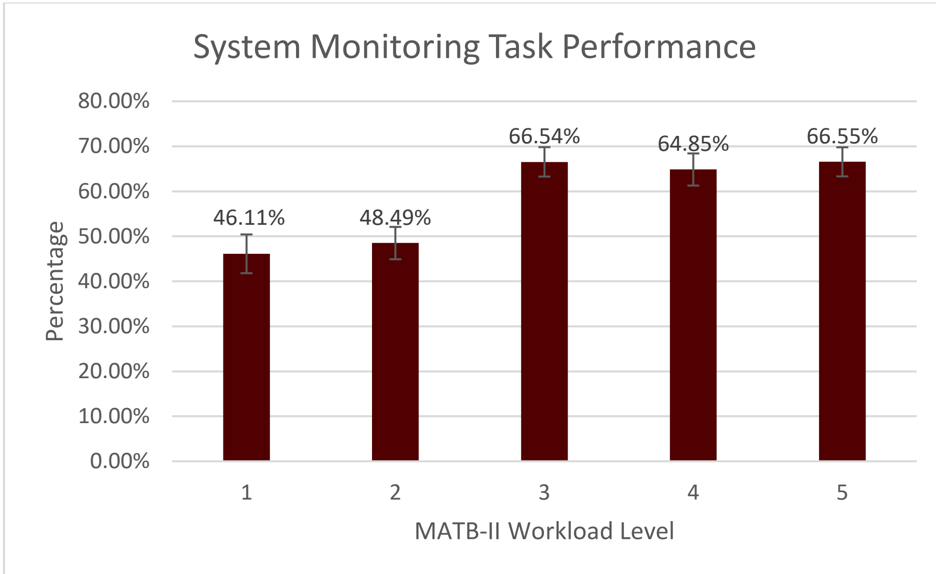


Figure 41: System monitoring performance. Error bars show standard error.

Results for the communication task (COMM) (Figure 42Figure 41) showed the MATB-II workload level had an impact on the performance ($F(4, 83)= 10.40$, $p<0.001$). Least squares means post-hoc results found no significant difference between levels 1 and 2 ($p = 0.2038$), but found significant difference between levels 1 and 3 ($p = 0.0055$), 4 ($p = 0.0150$), and 5 ($p < 0.001$). Level 2 was significantly different from levels 3 ($p < 0.001$), 4 ($p = 0.003$), and 5 ($p < 0.001$). Level 3 was not significantly different from level 4 ($p = 0.7130$) nor from level 5 ($p = 0.1299$). Level 4 was not significantly different from level 5 ($p = 0.06131$).

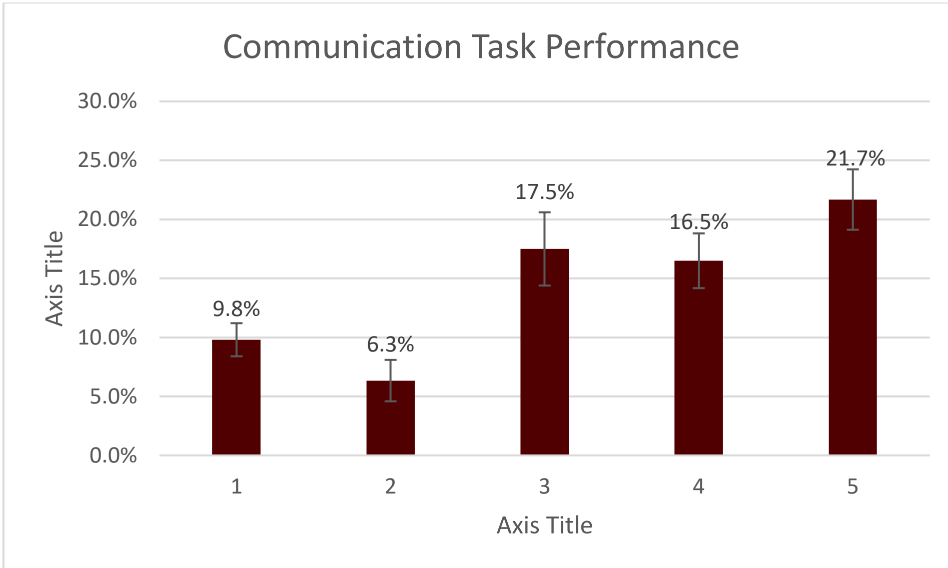


Figure 42: Communication performance. Error bars show standard error.

Discussion

Detecting the cognitive redline of workload could potentially identify when the person's performance may degrade, thus the importance of studying this topic. Research to date has primarily identified the cognitive redlines retroactively via subjective measures and performance, rather than in real-time. Physiological measures may provide data in real-time to detect the redline of cognitive workload, as described in recent studies (Mehler et al., 2009; Vogel & Machizawa, 2004). Heart rate variability and EDA were a few of the physiological measures studied in the previous study that show potential as indicators of the cognitive redline. Building upon these previous findings, the current study incorporated pupillometry and EDA as indicators of workload and characteristic patterns in the data indicate that the cognitive redline may be identified. Unlike HRV or other physiological measures, pupillometry is a relatively nonintrusive data source, with the hardware causing little discomfort or awareness beyond what eyeglasses might provide (Palinko et al., 2010; Recarte & Nunes, 2003).

The pattern of NASA-TLX was similar to the findings in the previous study (Chapter 2), providing further evidence of operators meeting and exceeding their individual redlines. The TLX score increases significantly as the task level changing from easy to medium, and to medium hard (i.e., from level 1 to level 3). The normal hard and very hard levels (i.e. levels 3 and 4) had very similar results. The extremely hard level (i.e. level 5) is nearly on the same level as 3 and 4. The findings show that the participants perceived differences in workload demand until level 3, which acted as a difficulty threshold. After level 3, additional difficulty was not registered, despite the extensive iterative pilot testing that was done to ensure each of the MATB-II levels had a similar magnitude of change between them. This is especially strong evidence, suggesting participants reached their redline, as the levels were presented to the participants in a counterbalanced order.

The results from the pupillometry showed that difficult tasks trigger a higher pupillary response, which is consistent with previous findings (Iqbal et al., 2004). Workload levels 1 and 2 (easy and medium workload difficulty) are nearly the same, representing loads that were sufficiently easy and so did not require a measurable amount of additional mental effort. A stable increasing pattern was identified for moderate workload (between levels 2 and 4). However, when the workload level reaches the difficult levels, pupillary response plateaus between levels 4 and 5, corresponding to the very hard and extremely hard levels of workload. The shape of the graph might indicate the reaching of a stable level (Mehler et al., 2009, Vogel & Machizawa, 2004).

The results of pupillometry, SSSQ and NASA-TLX suggested that pupillometry can be used to indicate the cognitive redline. Since SSSQ did not indicate an increase in stress levels, it can be inferred those changes in pupil diameter were assumed to be driven by task workload as a main factor. Interestingly, the plateau patterns differed slightly between NASA-TLX and pupil data, but both the objective (pupillometry) and subjective (NASA-TLX) show that a stable pattern with high workload levels, which lend support to this pattern being indicative of the cognitive redline.

Mental workload increases when the task loads increase, and it is important to detect when performance may degrade to ensure proper performance in safety-critical multitasking environments. Being able to detect when someone is approaching their redline may prevent performance degradation and decreasing safety risks.

A limitation for the current study that can be addressed in future studies is the use of wearable devices and using more real-world tasks to improve the ecological validity of these results. Additionally, the devices used in this study are non-intrusive, lightweight, and easy to use, which may have led to a tradeoff in data quality. The processing and analysis of the skin conductance response and level data was carefully administered, but the results still did not show a significant effect. Better, more reliable sensors could provide better data for analysis. Also, because some of the devices might not be suited for everyday wear, future studies will look at other physiological measures that may be more robust or less intrusive to operators. Additionally, more lab studies in a control environment can be used using more robust devices to detect physiological patterns associated with sympathetic and parasympathetic arousal due to cognitive workload.

Future studies need to address the individual differences that may affect the redline threshold. While the current study controlled for some factors, such as including the SSSQ to ensure the observed physiological changes were due to workload and not stress, and keeping the environment the same for all participants, future studies need to implement additional measures to account for individual differences. It would be interesting to understand how expertise may influence how soon someone reaches their redline threshold, which could be done by comparing novices to experts in certain tasks. Additionally, motivation may be an influencing factor as to how much the person tries to stay on the task as they approach their redline. These individual differences could show different performance, subjective, or even physiological changes.

CHAPTER IV

AUTONOMIC INDICES OF WORKLOAD

Introduction

The previous two studies investigated physiological patterns when the human is approaching their cognitive redline, and how these patterns correlate with performance and subjective measures. Based on the results from studies 1 and 2 (see Chapters 2 and 3), physiological measures reach an asymptote and performance degrades when the human experiences increasing levels of cognitive workload. The asymptote pattern could be related to the redline threshold, which will be investigated in more detail in this third study.

The physiological patterns represent autonomic activity, including both the sympathetic (SNS) and the parasympathetic (PSNS) nervous system for most physiological measures (see Chapter 1). To understand when the person may be approaching their redline, it is important to investigate the differences between the SNS and PSNS, and how these patterns reflect workload close to the redline threshold. While the SNS is related to “fight or flight”, previous studies have found SNS activity to increase with increasing cognitive workload (Backs, 1995; Backs & Seljos, 1994; Cinaz et al., 2013; Lean & Shan, 2012; Veltman & Gaillard, 1998). The PSNS has been studied as part of the recovery aspect, usually post-exercise (e.g.) (Chen et al., 2011; Pierpont et al., 2000). However, if the PSNS provides information about recovery, a similar analysis could be applied to investigate the physiological patterns associated with PSNS recovery after cognitive workload.

This study incorporated the use of virtual reality (VR), as it provides a fully-immersive experience, thus, allowing to observe physiological patterns associated with high levels of workload while still being in a controlled environment. Recently, VR has been used and implemented in highly cognitively demanding environments. For example, VR applications and training have been implemented in mining (Van Wyk & De Villiers, 2009; Zhang, 2017), oil and gas (Brasil et al., 2011; Lin & Liu, 2012), healthcare (de Ribaupierre et al., 2014; Mantovani et al., 2003), and aviation (Eschen et al., 2018; Marion et al., 2007).

Physiological data have been collected to assess other constructs such as stress or emotions in virtual reality (e.g.) (Annerstedt et al., 2013; Kotlyar et al., 2008; Macedonio et al., 2007; Yu et al., 2018). However, few studies have attempted to measure cognitive workload through physiology in simulated VR environments (e.g.)(Parsons & Reinebold, 2012; Tremmel et al., 2019), creating a need to investigate this aspect further. This is one of the research gaps that will be addressed with the present study.

This third study aims to answer the third and final research question postulated in the larger body of research: **To what extent do physiological measures indicate the redline of cognitive workload based on patterns of the different branches of the autonomic nervous system?** In this study, HRV data were collected using a research-grade electrocardiograph (ECG) system to ensure greater accuracy when performing the analysis to investigate autonomic activity related to cognitive workload.

The third study investigated the autonomic patterns (SNS, PSNS) related to when the human is approaching their redline of cognitive workload. Similar to the previous two studies, physiological data analysis was compared to the performance and subjective measures data. Participants played a videogame in virtual reality while ECG data were collected.

Methods

Due to COVID-19, data collection activities involving human subjects were restricted. Per regulations, only the participant and 1 experimenter were allowed in the lab space. In addition, both the participant and the experimenter had to wear masks throughout the study. All devices and workspaces used throughout the study were sanitized prior to each participant's arrival for the study, and hand sanitizer was provided. Appendix C has additional information regarding the COVID-19 protocol.

Fifteen participants (average age = 26.3, standard deviation = 5.5, 9 male and 6 female) completed the IRB-approved study (IRB approval: IRB2019-1311D). Participants were at least 18 years old, had normal or corrected-to-normal vision, and no physical conditions that would significantly challenge their ability to use the bimanual controllers (i.e. Oculus VR Touch Controllers, see Figure 43 and Figure 44) to complete the game tasks in the virtual environment (i.e. Cubism videogame). Upon their arrival at the laboratory, participants were briefed on the COVID-19 guidelines and study protocol

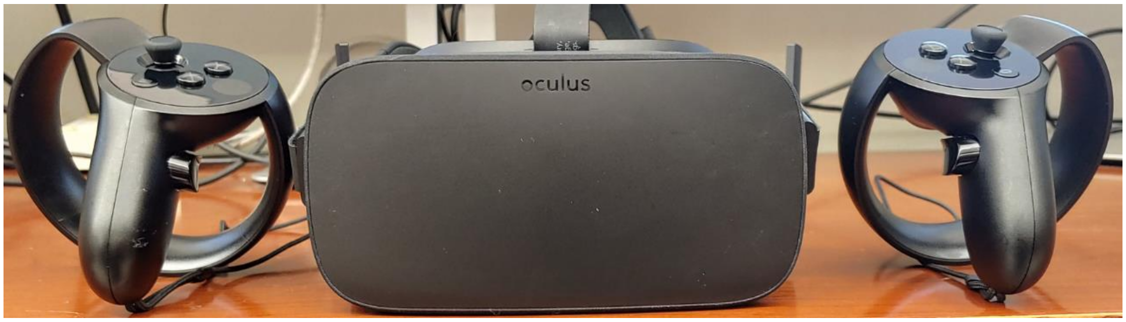


Figure 43: Oculus Rift VR system, with the headset in the center and the Touch controllers at the sides.

Figure 44: This figure shows how participants wore the Oculus VR system.

Participants were instructed to sit at a desk close the lab entrance, where they read and signed the consent form. Participants then completed a background questionnaire, which inquired about demographics, caffeine consumption, and participants' experience with videogames, including the hours per week played, systems used (e.g. type of gaming system, such as PC or console, and type of games played). In addition, the survey inquired whether participants had experience with VR, and if they did, they were asked to elaborate about their VR experiences. The complete questionnaire is included in Appendix C.

The physiological devices data collection devices used in this study included the Biopac BioNomadix RSPEC-R transmitter (ECG) (see Figure 45). The Biopac BioNomadix receiver was paired with the Biopac MP160 system, as seen in Figure 45.The Biopac AcqKnowledge software was used to collect data (see Figure 46). According to the COVID-19 protocol (see Appendix C), all physiological devices were disinfected prior to each participant's arrival for the study.

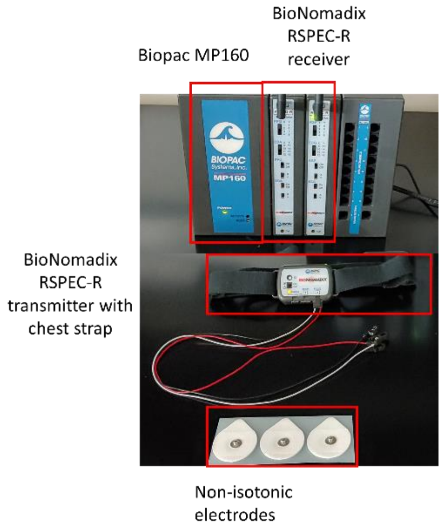


Figure 45: BioPac MP160 system with the BioNomadix receiver module (RSPEC-R). The BioNomadix transmitter is shown with the electrodes.

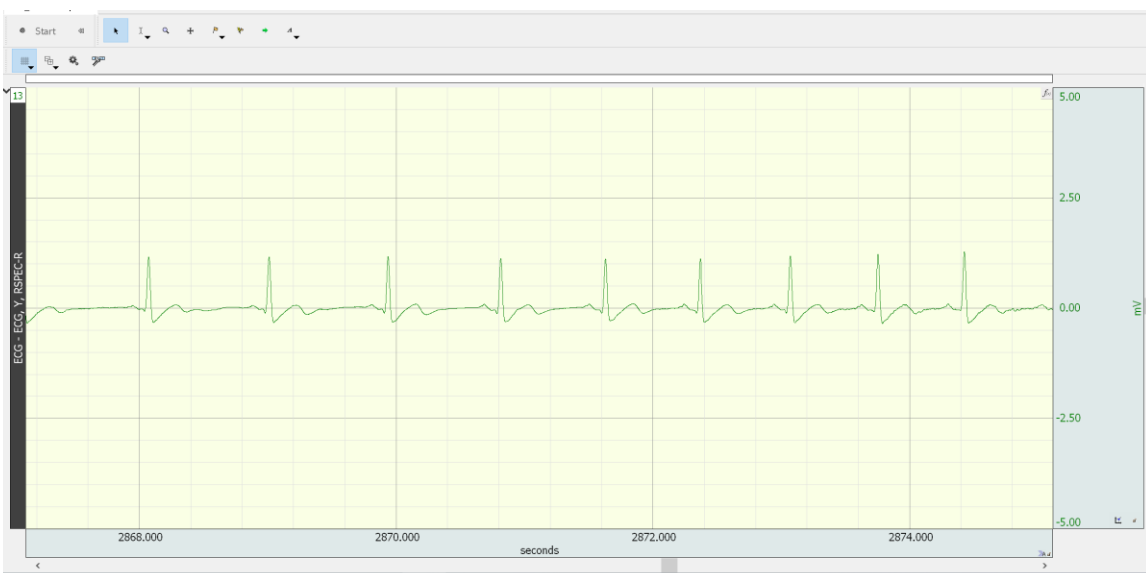


Figure 46: Biopac AcqKnowledge software showing ECG data.

Participants were then instructed how to put on the device used to collect the physiological data. A PowerPoint presentation contained a diagram that showed the correct electrode placement for the ECG BioNomadix, while the experimenter provided verbal instructions. The diagram is illustrated in Figure 53. Participants were then given a printed version of the ECG electrode placement diagram (Figure 53) and instructed to go to the restroom to put on the ECG transmitter (RSPEC-R) and the non-isotonic electrodes, which contain a hypertonic solution to provide better conductivity, as the device went under their shirt. Once they arrived back at the laboratory, participants took a seat to collect the physiological baseline.

Then, the physiological baseline data were collected following the procedure described in Chapter 1. As the baseline had been used in the previous 2 studies as a covariate, no further modifications were made to the baselining procedure. Participants were instructed to relax for 10 minutes while they listened to nature sounds and performed "paced breathing" by inhaling and exhaling in time with a visual cue presented on an iPad App, Breathe Easy (Inquiry Health LLC, 2017), see Figure 47. All physiological signals (ECG) collected during the baseline condition were processed as the dependent measures derived from physiological measures for analysis. The baseline condition data were used as covariates in the statistical analysis, as the data represent when the human is in a relaxed state, and experiencing minimal cognitive workload.

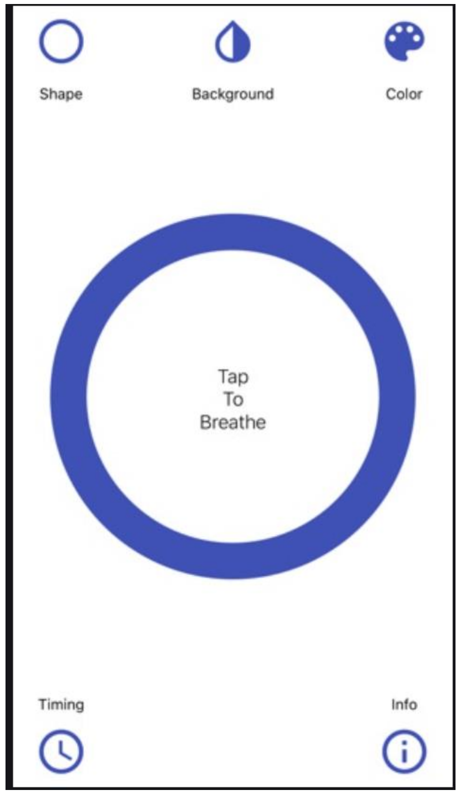


Figure 47: Breathe Easy (Inquiry Health LLC, 2017) provided a visual cue in the form of a circle. The circle would expand and contract, and participants were instructed to inhale and exhale by following this pattern.

After the baseline condition, participants proceeded to watch an instructional video (see Figure 48) which demonstrated how to properly put on and adjust the Oculus Rift VR headset. The video was necessary as the experimenter, following the COVID-19 protocol, was not able to physically interact with the participant and help to adjust the headset. The video was filmed in the laboratory by the experimenters.

Figure 48: Screenshot of the training video. Participants had to watch the training video to understand how to put on and adjust the Oculus Rift headset and the Touch controllers.

Participants were then instructed to sit in the Oculus playing area, which was in a different part of the lab. Participants remained in this part of the lab until the end of the study. A 6-feet distance was always kept between the participant and the experimenter due to the COVID-19 protocol, for example in Figure 49.

Figure 49: The Oculus VR playing area is denoted by masking tape on the floor. Participants were seated in this area, and the experimenter kept a 6-feet distance due to the Covid 19 protocol.

Participants were then instructed to hold in each hand the Oculus Rift Touch controllers (Figure 43), secure them by wearing the wrist straps and locking them securely around their wrists. Participants then put on the Oculus Rift VR headset (see Figure 44).

After adjusting the headset to fit satisfactorily, participants proceeded to open the Cubism videogame (see Figure 50 for more details). On-screen instructions prompted participants to press the back trigger buttons on the Touch controllers, which calibrated the playarea for the game to ensure participants can easily reach the figures and configuration settings in the tri-dimensional (3D) space. Then, participants were instructed to reach out and grab the green rectangle by using the back trigger button on the Touch controller (Figure 50a), and place it within the gray rectangle (see Figure 50b). Physically moving the Touch controllers moves the cursors on the screen in a tri-dimensional space. Pressing the back-trigger button on the Touch controllers performs a grabbing motion. Completing this first puzzle opened the Cubism main screen (see Figure 50), and effectively introduced participants to the game, consisting of completing 3-dimensional puzzle pieces.

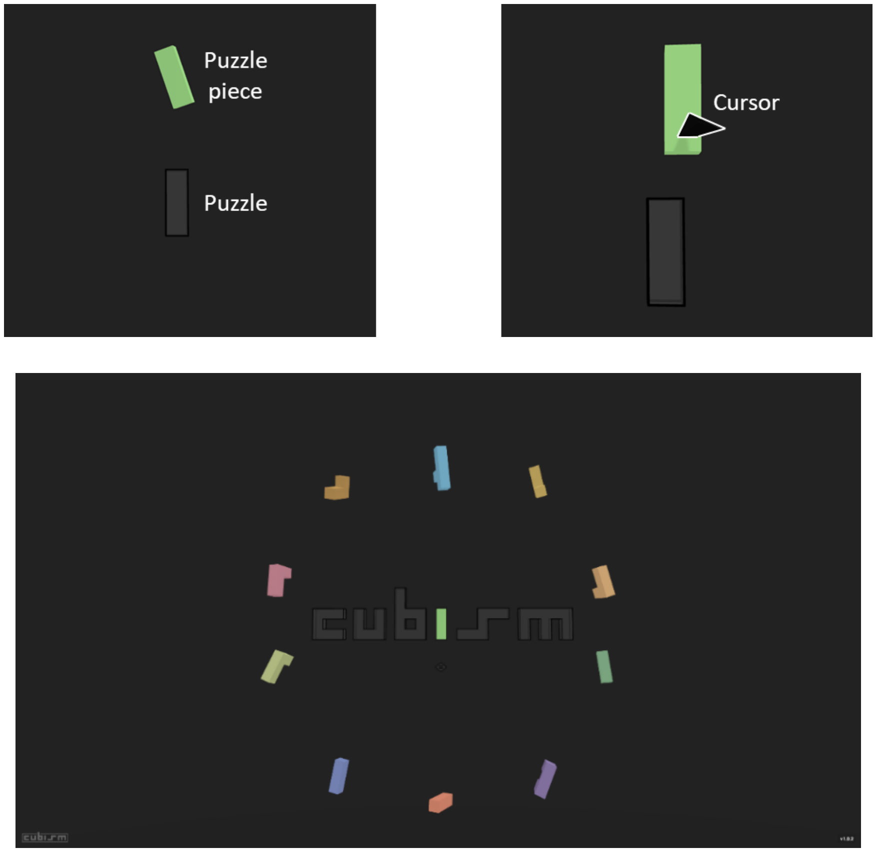


Figure 50: (a) A green puzzle piece and a grey puzzle appeared upon calibrating the play area. **(b)** Participants more the Touch controller and press the back trigger button to grab the figure. **(c)** Completing the puzzle presents the Cubism main screen.

Participants then proceeded to complete a training procedure to become more familiarized with the Oculus Rift touch controllers, as well as the game dynamics. Training consisted of 3 different puzzles (see Figure 51), as training has been used before to decrease the effect of the learning effect (Prinzel et al., 2000). Participants were coached on how to move the pieces throughout the 3 puzzles, but during the third puzzle, the n-back task was also introduced. All participants completed all the training puzzles, that is, they placed all the geometric shapes inside the gray figure (see Figure

51). Participants received periodic feedback about the n-back task while completing the final training puzzle.

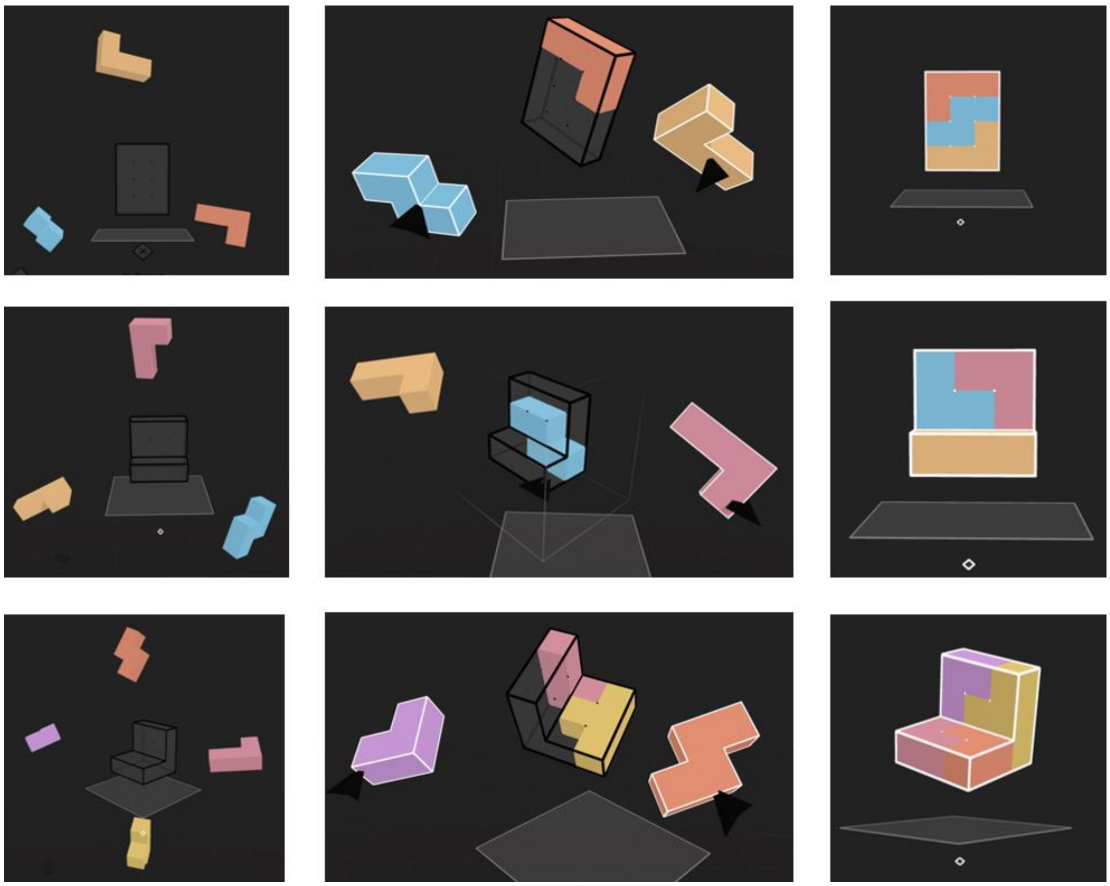


Figure 51: Cubism training puzzles. The left column shows the figure configuration when the puzzle is opened. The middle column shows the participant's progress in completing the puzzles. The right column shows the completed puzzle.

After completing the training procedure, participants removed the headset and took a 3-minute break, the duration which was based on the previous two studies. Pilot testing was done to determine if the resting break duration was sufficient in the VR environment, and physiological data and subjective reports were collected. Based on the subjective reports and physiological pilot data, the break duration remained at 3-minutes. During this break, they relaxed while listening to nature sounds and followed the paced breathing exercise on the iPad, similar to the baseline procedure. This break allowed participants to revert to a relaxed state with minimal cognitive workload, and for physiological measures to reset to levels similar to those of the baseline condition.

After this break, participants were instructed to put the Oculus Rift headset back on, and reminded of the proper way to adjust the headset based on the instructional video. Participants then proceeded to work on the first scenario. Each scenario, presented in a counterbalanced order, consisted of a puzzle and an n-back task (see Table 9). The maximum time allowed for the puzzles was 5 minutes, but the scenario was finished early if participants completed the puzzle. The scenario difficulty was defined according to the puzzle complexity and the n-back task, and ranged from very easy to very hard. See the next section for details of tasks and difficulty characterizations. At the end of each scenario, participants completed a NASA-TLX questionnaire, which included the weighted scales. NASA-TLX was selected as it provides the aggregate workload for both tasks (i.e. primary and secondary task). The full cycle of 3-minute rest. Scenario, and NASA-TLX was repeated a total of 5 times. Once participants completed the last scenario, they were instructed in how to remove the physiological devices and place them aside to be sterilized, and throw away the electrodes. A summary of the experiment study is provided in Figure 52.

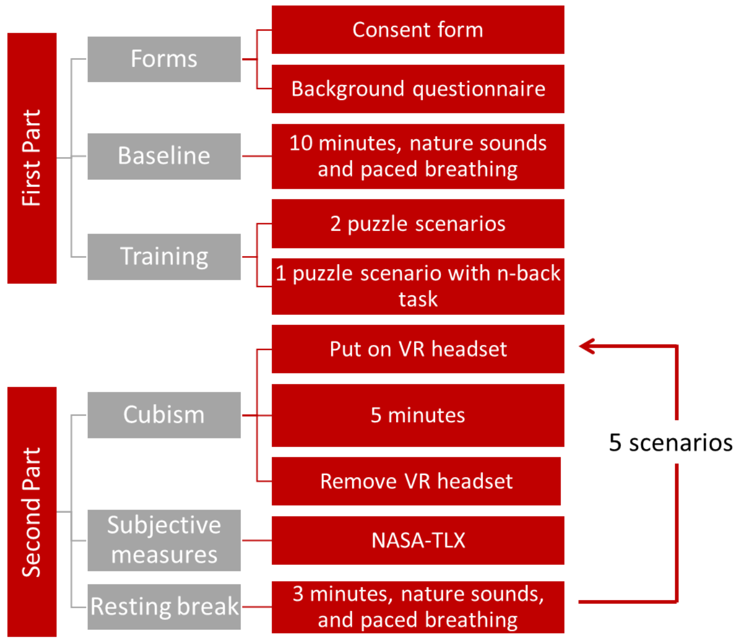


Figure 52: Experimental procedure.

Apparatus

The different physiological devices and questionnaires used during the data collection are explained in this section. Additionally, a summary of the dependent and independent measures is provided in Table 7.

Physiological Sensors

The Biopac system was used to collect ECG data. The MP160 module was connected with the RSPEC-R BioNomadix transmitters via Bluetooth to collect ECG data (see Figure 45).

The RSPEC-R transmitter collected ECG recordings via a lead II configuration, where the electrodes are placed on the chest wall equidistant from the heart (Abi-Saleh & Omar, 2010). The negative lead was placed below the right clavicle, the positive lead on the lower rib on the left side, and the ground lead was placed below the left clavicle (see Figure 53). The electrodes used were nonisotonic.

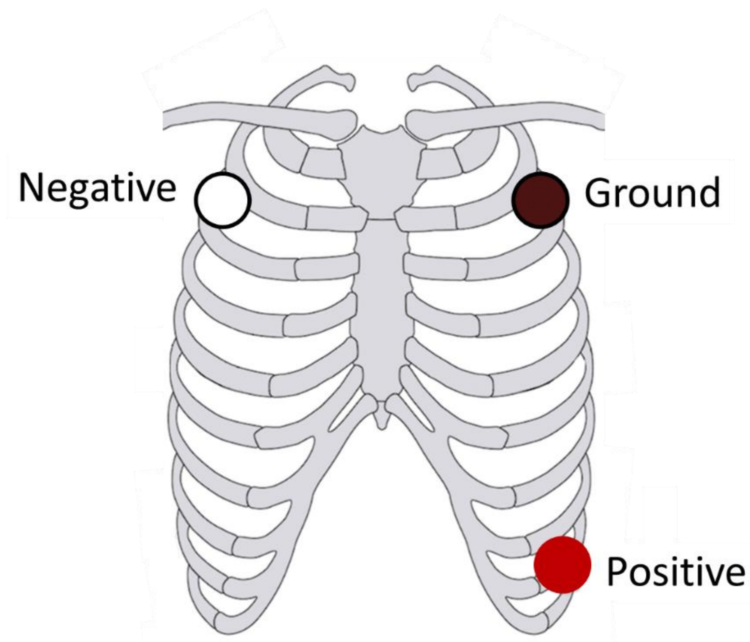


Figure 53: Lead-II configuration. The negative electrode is placed below the right clavicle, the ground electrode is below the left clavicle, and the positive electrode is placed on the lower left rib.

The ECG data was analyzed using the BioPac AcqKnowledge software, by using the hemodynamics analysis to extract several measures from the ECG data, including heart rate, and the R-R interval.

Virtual Reality Device

The Oculus Rift VR system was used, and consists of the headset, 2 Touch controllers, and 2 sensors. The Oculus Rift was chosen as it provides a fully-immersive experience, and videogames that provide a greater realism have been found to elicit greater increases in skin conductance levels (Ivory & Kalyanaraman, 2007). The Oculus Rift headset features a customized optical configuration which allows for a wide field of view with high visual fidelity (Facebook Technologies, LLC, 2021). The Touch control system uses two hand-encompassing remote controllers tracked through the sensors, providing intuitive control in VR. The two sensors tracked an array of infrared (IR) LEDs on the front faceplate of the headset to map the player's position in relation to the headset and controllers (see Figure 54).

Figure 54: Oculus VR system. The headset is shown in the top figure, and the Touch controllers on the bottom figure.

Cognitive Workload Manipulation in Virtual Reality

Cognitive workload conditions were manipulated in Cubism (Van Bouwel, 2020), a virtual reality game consisting of 3D puzzles based on geometric shapes. Cubism provides a challenge to spatial reasoning, as each of the puzzle pieces can be rotated and must fit within the larger 3D piece. Previous studies have manipulated cognitive workload with puzzles (e.g. Guastello et al., 2015; Whitaker et al., 1997), and more recently, using 3D puzzles (Maior et al., 2018; Milla et al., 2019; Werrlich et al., 2018). The Cubism puzzles are easy to interact with, as the Touch controllers are intuitive, but still provide a challenging cognitive task. As the participant moves their hands, they can see the cursors moving, and pressing on the back trigger button creates a grabbing motion to touch the figures. The cognitive challenge is set in understanding how the different 3D geometric pieces fit within the larger 3D puzzle.

The difficulty level was varied based on several parameters, such as the number of 3D puzzle pieces per puzzle, and size and complexity of the puzzle, and refined with extensive pilot testing made by the experimenters (
Figure 55).

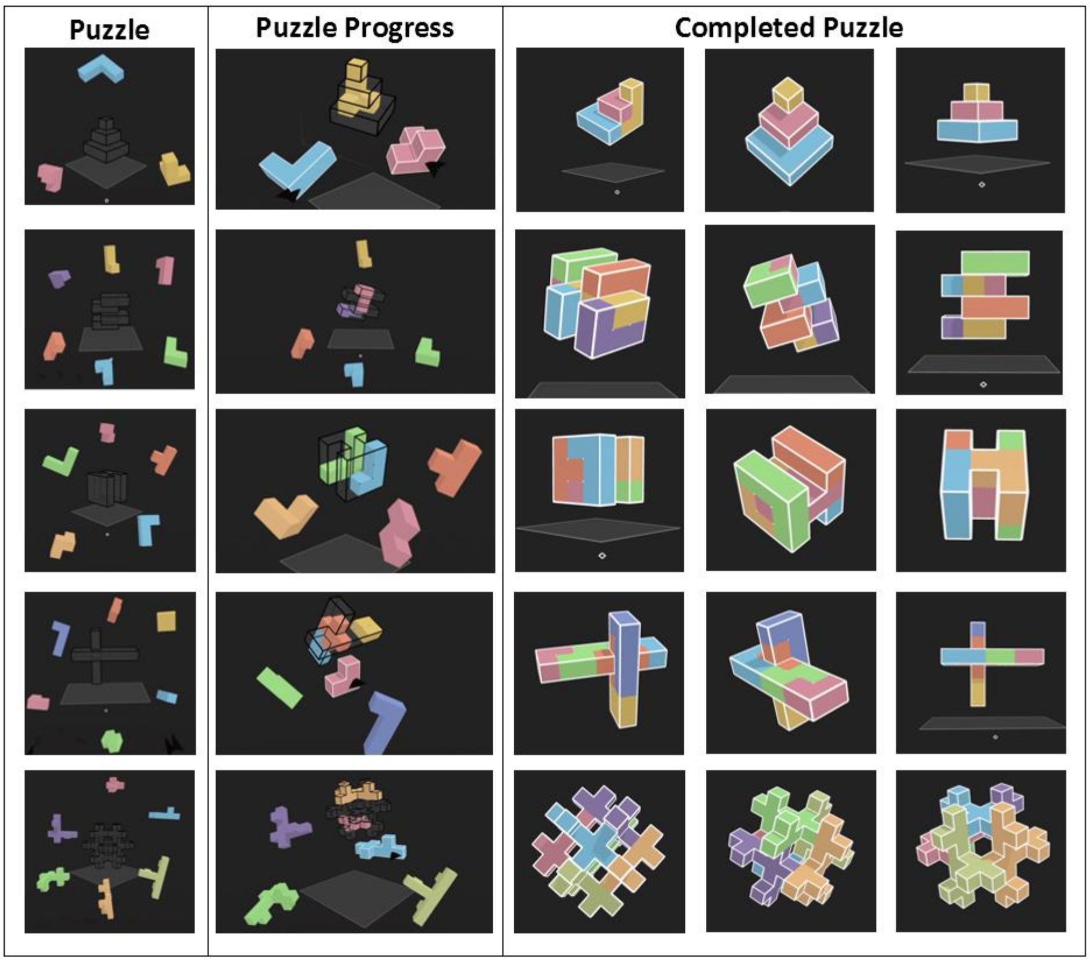


Figure 55: The five selected Cubism puzzles. The first column shows the unsolved puzzle, the middle column shows the puzzle in different stages of completion, and the last column shows 3 different views of the completed puzzle.

Secondary Task

An auditory n-back task, consisting of a sequence of colors, was selected as the secondary task, as it minimally interfered with the visual, manual, and spatial reasoning required by the primary task, according to Wickens' Multiple Resource Theory (see Chapter 1) (Wickens, 2008). Colors were selected instead of other random words, words referencing color add additional cognitive workload in differentiating the colors from the primary task and the secondary task. Previous studies have used the n-back task has been used in previous studies to assess cognitive workload (Brouwer et al., 2012; Herff et al., 2014; Pergher et al., 2019; Wang et al., 2015). The n-back task consisted of a sequence of colors randomly generated (see Figure 56 for an example), and read through Panopreter (Panopreter, 2021), a text-to-speech software, using Microsoft's David Voice. The sequence of colors lasted for approximately 20 seconds, then an alarm sound would ring to indicate the participant needed to verbally respond with the n-back answer, which participants were informed if it was 1-back, 2-back or 3-back before the beginning of the scenario. After the alarm sound, participants had around 5 seconds to reply before the next sequence of colors would begin. The easy levels were paired with 1-back, the next two levels were paired with 2-back, while the most difficult level was paired with 3-back.

Figure 56: Example of the sequence of colors presented during the n-back task. These were read out loud to participants.

Experimental Variables

The dependent variables are based on the three methods for detecting changes in cognitive workload (i.e. subjective methods, performance metrics, and physiological measures). A summary of the variables collected is listed in Table 7.

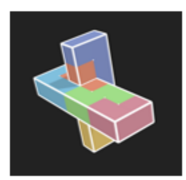
Table 7: Dependent Variables

Workload Measure Type	**Signal or Measure Collected**	**Device or Method Collected**	**Unit**
Physiological	Electrocardiogram (ECG)	Biopac BioNomadix RSPEC-R transmitter	Milliseconds (ms)
Subjective	NASA-Task Load Index (NASA-TLX)	NASA-TLX plus weights from pairwise comparison	TLX Score
Performance	Percent correct (correctly-placed puzzle pieces)	Cubism	Percentage (%)
Performance	Percent correct	n-back task	Percentage (%)

Experiment Design

A total of 5 scenarios were designed for the study to impose workload. The scenarios are a combination of Cubism puzzle and n-back task, and ranged in difficulty from very easy to very difficult. The difficulty level for the Cubism puzzles was selected based on the number of pieces and piece complexity (e.g. how many angles per piece), and refined with extensive pilot testing by the experimenters. The N-Back task was then selected based on an auditory color sequence, as it minimally interrupts the vision, spatial, and manual controls, while adding additional cognitive workload in differentiating the colors between the primary and secondary task. Table 8 contains the final design for the scenarios.

Table 8: Scenarios designed for the study

Scenario	N-Back Audio	N-Back	Cubism Puzzle	
1	1	1-Back	3-1	
2	2	1-Back	4-2	
3	3	2-Back	5-3	
4	4	2-Back	7-3	
5	5	3-Back	6-5	

The scenarios were presented in a counterbalanced order to minimize the carryover effect during the statistical analysis, as seen in Table 9.

Table 9: Counterbalanced Order

Participant	1	2	3	4	5
1	Audio 1 1-Back Cubism 3-1	Audio 2 1-Back Cubism 4-1	Audio 5 3-Back Cubism 6-5	Audio 3 2-Back Cubism 5-3	Audio 4 2-Back Cubism 7-3
2	Audio 2 1-Back Cubism 4-1	Audio 3 2-Back Cubism 5-3	Audio 1 1-Back Cubism 3-1	Audio 4 2-Back Cubism 7-3	Audio 5 3-Back Cubism 6-5
3	Audio 3 2-Back Cubism 5-3	Audio 4 2-Back Cubism 7-3	Audio 2 1-Back Cubism 4-1	Audio 5 3-Back Cubism 6-5	Audio 1 1-Back Cubism 3-1
4	Audio 4 2-Back Cubism 7-3	Audio 5 3-Back Cubism 6-5	Audio 3 2-Back Cubism 5-3	Audio 1 1-Back Cubism 3-1	Audio 2 1-Back Cubism 4-1
5	Audio 5 3-Back Cubism 6-5	Audio 1 1-Back Cubism 3-1	Audio 4 2-Back Cubism 7-3	Audio 2 1-Back Cubism 4-1	Audio 3 2-Back Cubism 5-3
6	Audio 4 2-Back Cubism 7-3	Audio 3 2-Back Cubism 5-3	Audio 5 3-Back Cubism 6-5	Audio 2 1-Back Cubism 4-1	Audio 1 1-Back Cubism 3-1
7	Audio 5 3-Back Cubism 6-5	Audio 4 2-Back Cubism 7-3	Audio 1 1-Back Cubism 3-1	Audio 3 2-Back Cubism 5-3	Audio 2 1-Back Cubism 4-1
8	Audio 1 1-Back Cubism 3-1	Audio 5 3-Back Cubism 6-5	Audio 2 1-Back Cubism 4-1	Audio 4 2-Back Cubism 7-3	Audio 3 2-Back Cubism 5-3
9	Audio 2 1-Back Cubism 4-1	Audio 1 1-Back Cubism 3-1	Audio 3 2-Back Cubism 5-3	Audio 5 3-Back Cubism 6-5	Audio 4 2-Back Cubism 7-3
10	Audio 3 2-Back Cubism 5-3	Audio 2 1-Back Cubism 4-1	Audio 4 2-Back Cubism 7-3	Audio 1 1-Back Cubism 3-1	Audio 5 3-Back Cubism 6-5

COVID-19 Protocol

As part of the study, a COVID-19 reopening protocol was developed (see Appendix C for the full protocol). Certain precautions had to be considered to ensure a safe environment for the study. The experimenter disinfected all tables and devices before each of the participant's arrival. Throughout the stud, a social distance of at least 6 feet was maintained at all times. In addition, both the participant and experimenter wore masks while in the lab.

Results

Fifteen participants (male = 9, female = 6) completed the IRB-approved study (IRB approval: IRB2019-1311D). The mean age for participants was 24.3 years old, (standard deviation = 3.7).

All data were analyzed using SAS Studio statistical software, with the significance level was set at $\alpha = 0.05$. A within-subjects ANCOVA analysis was used to analyze physiological measures, and the covariate was the data collected during the baseline procedure to account for individual differences. All other measures (subjective and performance) were analyzed using within-subjects ANOVA. Post-hoc tests involved standard t-tests to determine where the differences between groups occurred.

Physiological Measures

Heart rate (Figure 57) was not significantly affected by workload ($F(4, 36) = 1.80$, $p = 0.1511$).

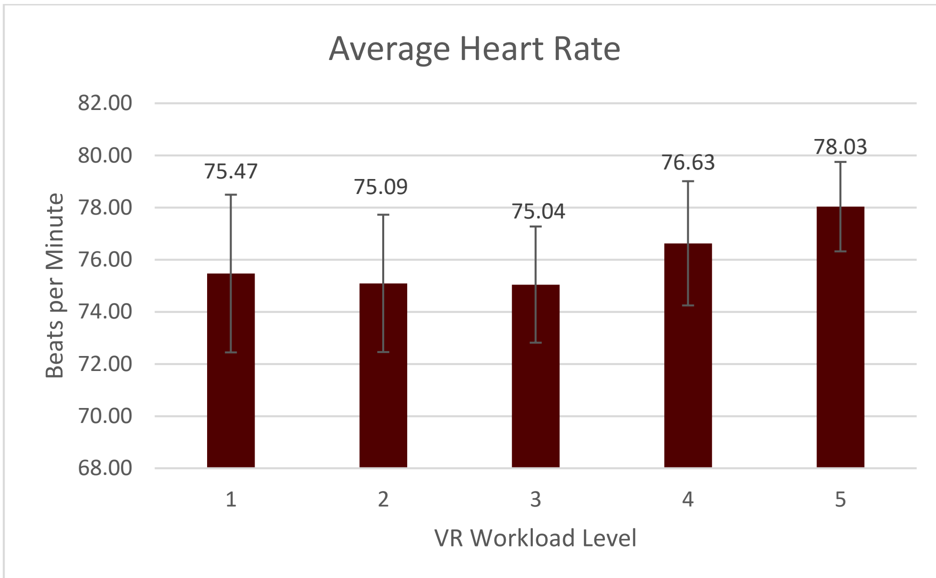


Figure 57: Average heart rate. Error bars show standard error.

Heart period (Figure 58) was not significantly affected by the workload level ($F(4, 36) = 1.99$, $p = 0.1174$).

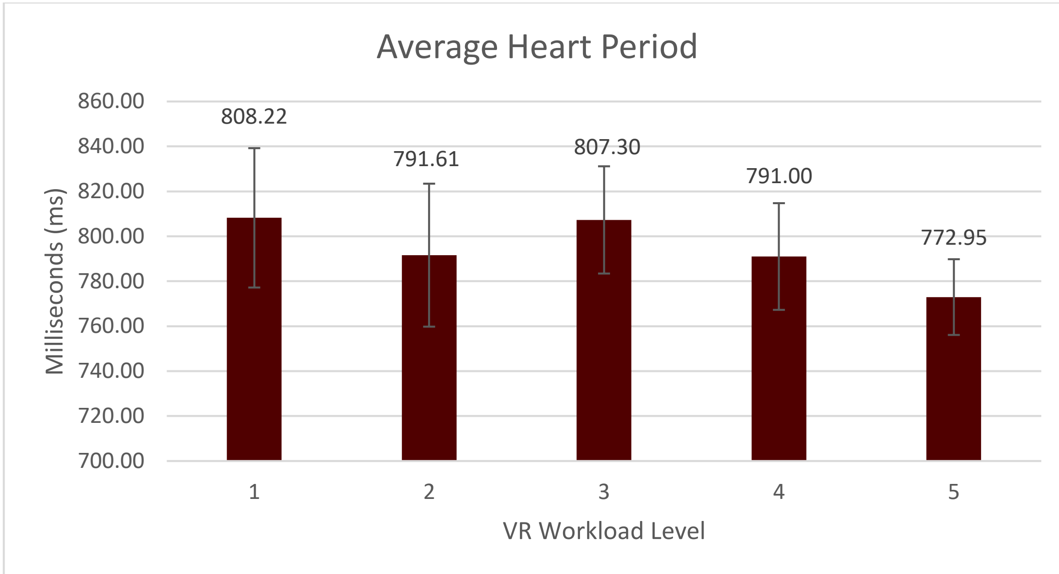


Figure 58: Average heart period. Error bars show standard error.

Heart rate variability (HRV) (Figure 59) was analyzed based on spectral analysis. The workload level did not have a significant effect on the low frequency ($F(4, 36) = 1.48$, $p = 0.2299$).

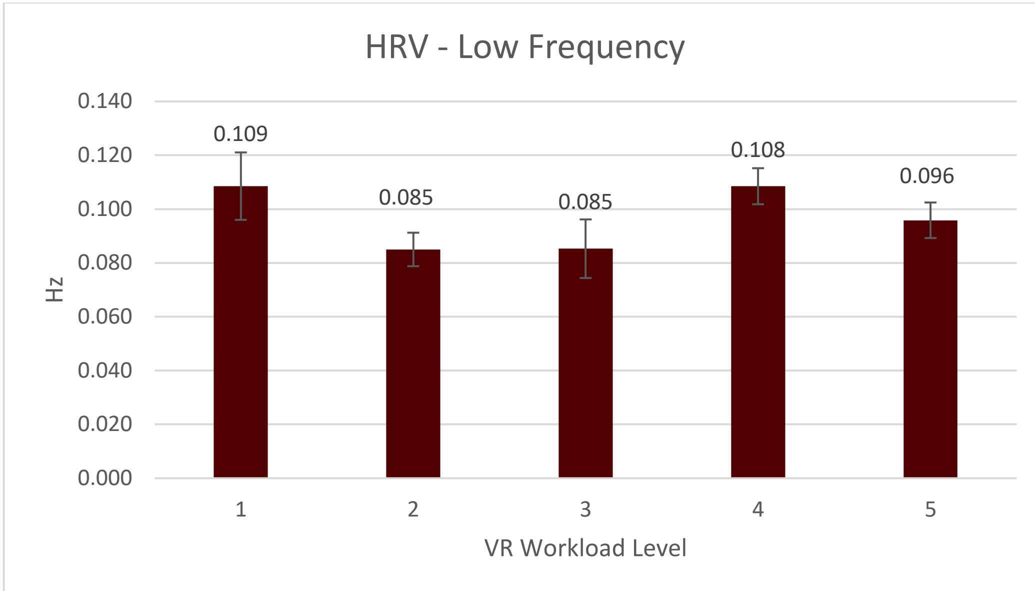


Figure 59: Low frequency analysis of heart rate variability. Error bars show standard error.

The heart rate variability (HRV) high frequency (Figure 60) was not significant ($F(4, 36) = 1.48$, $p = 0.2299$).

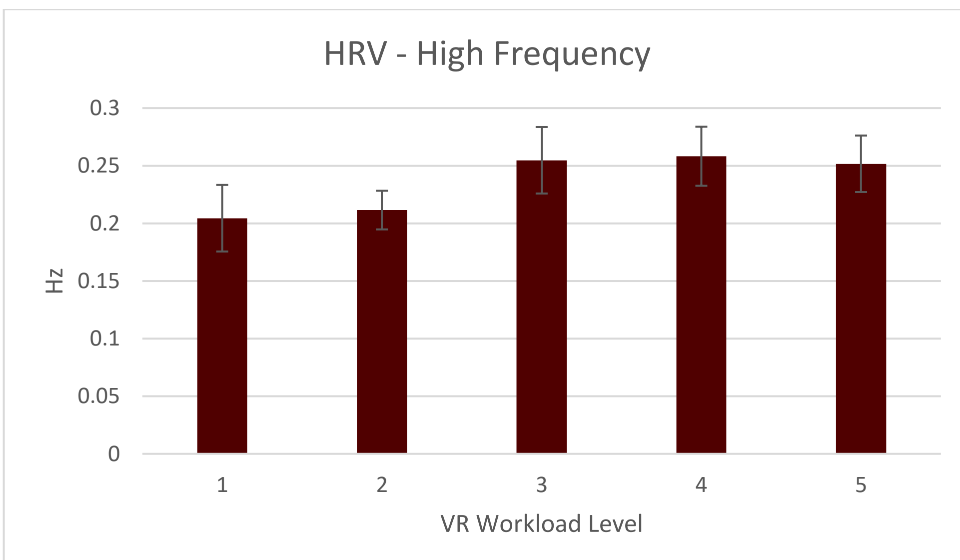


Figure 60: High frequency analysis of heart rate variability. Error bars show standard error.

Subjective Measures

The total NASA-TLX (Figure 61) score showed that workload had an impact on the subjective ratings ($F(4, 56) = 17.39$, $p < 0.0001$). Post hoc results using least squares means showed that level 1 was not significantly different from level 2 ($p = 0.0861$), but level 1 was significantly different from levels 3 ($p < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 2 was significantly different from levels 3 ($p < 0.001$), 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 3 was not significantly different from level 4 ($p = 0.7742$), nor from level 5 ($p = 0.5843$). Level 4 was not significantly different from level 5 ($p = 0.7942$).

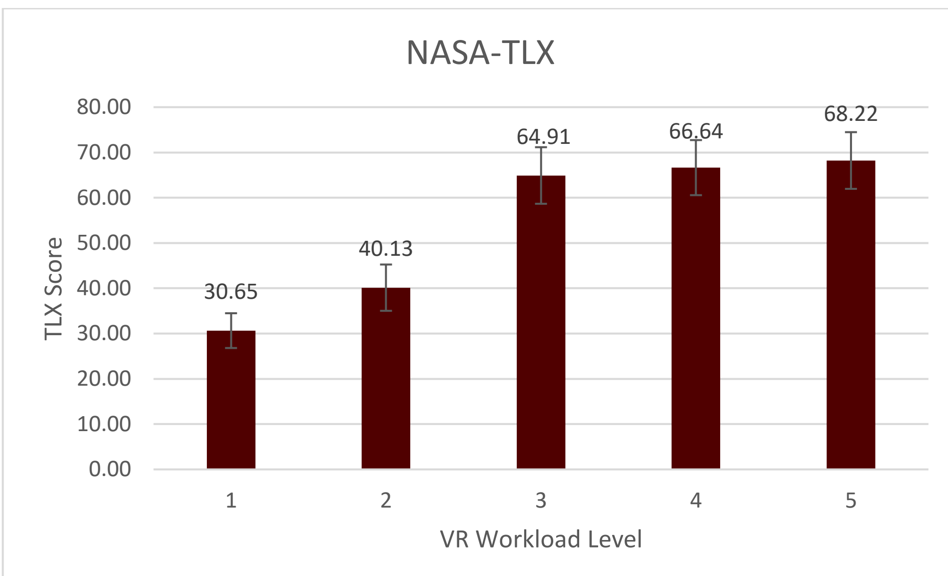


Figure 61: NASA-TLX. Error bars show standard error.

Performance

Performance data for the primary task (Figure 62) was calculated based on the total score for each puzzle, consisting of the number of correctly placed puzzle pieces at the end of 5 minutes. The puzzle difficulty had a significant effect on the total score ($F(4, 42) = 8.49$, $p < 0.0001$). Post hoc results using least squares means showed that level 1 was not significantly different from level 2 ($p = 0.9669$), but level 1 was significantly different from levels 3 ($p = 0.0034$), 4 ($p = 0.0019$), and 5 ($p < 0.001$). Level 2 was significantly different from levels 3 ($p = 0.0044$), 4 ($p = 0.0026$), and 5 ($p < 0.001$). Level 3 was not significantly different from level 4 ($p = 0.8432$), nor from level 5 ($p = 0.1850$). Level 4 was not significantly different from level 5 ($p = 0.2590$).

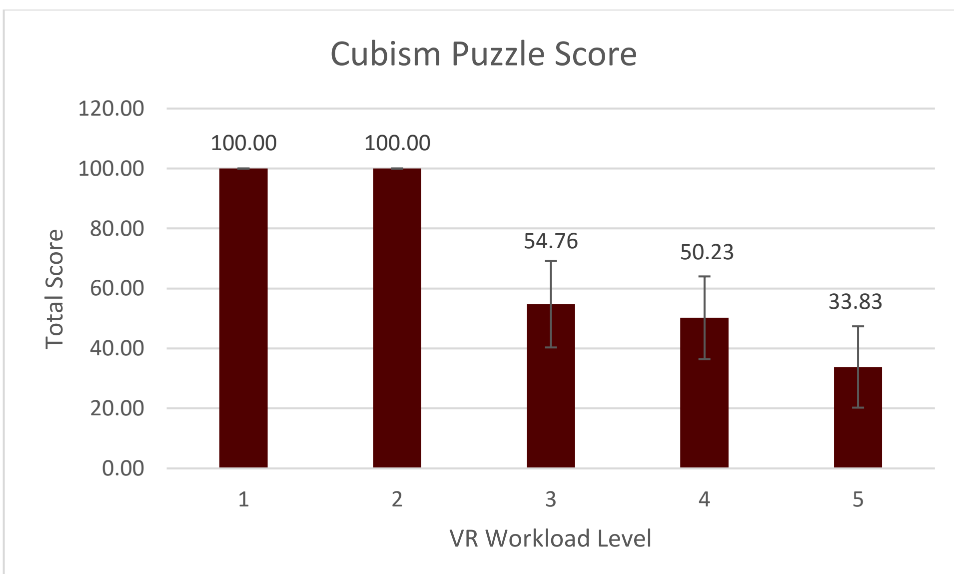


Figure 62: Cubism puzzle score. Error bars show standard error.

The secondary task (Figure 63) was calculated based on the percent correct of the n-back task. The workload level had an impact on the n-back task score $F(4, 48) = 15.49$, $p < 0.0001$). Post hoc results using least squares means showed that level 1 was not significantly different from level 2 ($p = 1.00$), nor level 3 ($p = 0.0603$), but level 1 was significantly different from levels 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 2 was not significantly different from level 3 ($p = 0.0603$), but was significantly different from level 4 ($p < 0.001$), and 5 ($p < 0.001$). Level 3 was not significantly different from level 4 ($p = 0.0190$), but was significantly different from level 5 ($p < 0.001$)). Level 4 was not significantly different from level 5 ($p = 0.0529$).

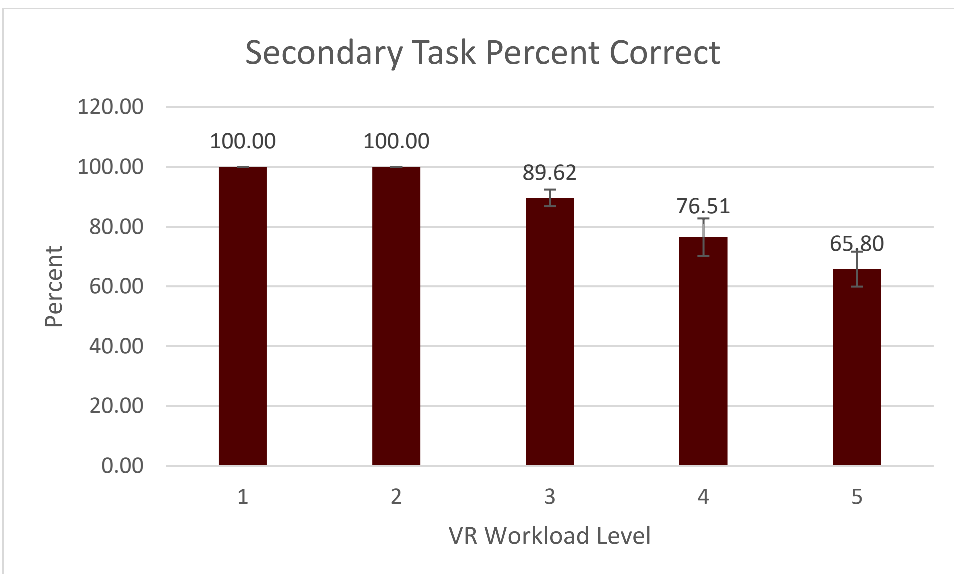


Figure 63: Secondary task performance. Error bards show standard error.

Discussion

The cognitive redline of workload is an interesting and important but difficult to detect phenomenon in the human factors domain. Previous research has focused on identifying cognitive redlines based on performance and subjective ratings, rather than potentially detecting the redline in real-time. Physiological indicators of sympathetic and parasympathetic arousal can provide insight in the cognitive redline, and some recent studies suggest that dynamic patterns exhibit by physiology could be indicative of humans approaching their redline (Mehler et al., 2009; Vogel & Machizawa, 2004). Building upon findings from the previous two studies, the current study used cardiovascular measures as an indicator of workload and characteristic physiological patterns that may identify the cognitive redline.

The subjective ratings (NASA-TLX) show a similar pattern to the findings presented in the previous two studies, where participants subjectively felt the workload kept increasing, providing further evidence of participants approaching their redline. The NASA-TLX score was significantly affected by the workload level. These findings show that participants felt each of the tasks got steadily more complicated, requiring more cognitive resources. This is especially strong evidence as the different puzzles representing different workload levels were presented in a counterbalanced order.

The results from physiology were not significantly affected by the workload level, but there are unique patterns that approach a trend, especially the low and high frequency analysis of heart rate variability. The low frequency represents sympathetic arousal, and the high frequency represents parasympathetic arousal. While sympathetic arousal was fluctuating, parasympathetic activity was higher for the last three scenarios, which could mean that the parasympathetic was getting activated.

Performance measures for the primary and secondary task show stable performance for the first and second workload levels. However, as the difficulty of the scenarios increases, performance degrades. Puzzle completion performance for levels 1 and 2 was 100% correct, but performance dropped starting at level 3. Performance for levels 3, 4, and 5 was significantly different from level 1, which was the easiest one, thus to be taken as the baseline. For the secondary task, levels 1 and 2 had a 100% score, and level 3 was not significantly different from level 1. However, levels 4 and 5 were statistically significant from level 1, which means that, starting at level 4, performance precipitously dropped. The results indicate that the results from subjective measures and performance could be used to indicate the redline threshold, but future studies need to investigate the role of physiological patterns with more subjects.

As cognitive workload increases, the person is able to maintain stable performance until a critical point, where their performance starts to degrade. Being able to detect this redline threshold of workload provides several potential benefits, such as improving safety in the workplace by decreasing the risk of performance degradation. Some performance degradation mitigation techniques consist of redesigning the task , changing procedures, or additional training (Grier et al., 2008; Reid & Colle, 1988).

Several limitations exist when planning a data collection involving physiological devices. First and foremost, it is important to consider the environment and the conditions under which the data will be collected. For example, high temperatures will impact sweat production, and previous studies have shown that changes in ambient temperature impact non-specific skin conductance fluctuations (Doberenz et al., 2011). Similarly, cold temperatures will decrease the effectiveness of optical sensors to collect heart rate-based measures (Khan et al., 2016), and sunlight will impact the accuracy of pupillometry (Beatty, 1982).

Additionally, there is a tradeoff between wearable devices and data quality. Medical-grade, highly accurate devices may be easily used in controlled laboratory settings, but these devices are less than ideal to collect data in a naturalistic setting. The studies presented in Chapters 2 and 3 provide a wide range of possible physiological devices, which may be applicable in different situations. Physiological data collection is not a one-size-fits-all situation, rather, each situation and data collection need to be carefully assessed to ensure the proper devices are used to collect the correct data for the study.

A limitation of the current study that can be addressed in future studies is a small sample size, as data was collected during the COVID-19 pandemic. Data from additional participants can help to expand the analysis that can be done with the physiological measures to detect different patterns associated with the redline of cognitive workload. Additionally, it is uncertain how much stress participants were already experiencing due to the pandemic, which may have influenced their physiological response. Additionally, ECG was used in this study, but other wearable sensors may also be implemented in the future, which are less intrusive to participants.

Future studies could investigate the role of individual differences that may affect how soon people reach their redline threshold. For example, incidental stress, motivation, or expertise. While certain controls were implemented to mitigate individual physiological differences, such as using the baseline data as covariate, other individual differences not related to physiology may influence how soon people perceive or actually reach their threshold.

CHAPTER V

CONCLUSIONS

Volatile, uncertain, complex, and ambiguous (VUCA) environments are often in safety-critical domains, where specific performance levels are required to maintain safety standards. In these environments, experiencing cognitive overload can lead to increased safety risks (Carayon & Alvarado, 2007; Endacott, 2012; Milczarek et al., 2007), and approximately 68% of the American workforce feels overloaded while performing their daily activities (IT Business Edge, 2014), and constantly feeling overloaded can lead to burnout (Portoghese et al., 2014; Sweeney & Summers, 2002; Van Bogaert et al., 2013). Every year, an approximate $125 to $190 billion are spent in healthcare costs associated to employee burnout (Blanding, 2015; Garton, 2017). While people feeling overloaded is concerning, the problem arises when the person approaches their redline of cognitive workload, which is the threshold where the available mental resources are depleted, so the person no longer has residual cognitive resources to continue performing the current tasks, or take on additional tasks. At the redline threshold, the task demands surpass the person's available mental resources, and usually results in performance degradation, increasing the risk for errors.

Previous research has investigated the point where performance degrades, by studying the relationship between performance and subjective measures based on varying levels of cognitive workload. For example, (Reid & Colle, 1988) provided the score of 40+/-10 for the Subjective Workload Assessment Test (SWAT) as a range where performance degradation would occur. Similarly, a score of 60 or more in the

Improved Performance Research Integration Tool (IMPRINT) was considered as the range for performance degradation (Mitchell et al., 2003). However, a limitation of these studies is that only performance-based measures and subjective metrics were used; no previous study has incorporated the use of physiological measures to detect the redline of cognitive workload.

The research presented in this dissertation investigated several physiological measures as objective data measures to detect the redline of cognitive workload threshold. Physiological measures, which are modulated by the autonomic nervous system (ANS), reflect changes in cognitive workload (Cinaz et al., 2013, 2010; Luque-Casado et al., 2016; Mehler et al., 2011). The ANS system, which controls certain body activities (e.g. blood pressure, heart rate, digestion), has two main branches, the parasympathetic nervous system (PSNS), and the sympathetic nervous system (SNS) (Jansen et al., 1995; Waterhouse & Campbell, 2008). The PSNS maintains homeostasis (Jänig & Häbler, 2000; Pichon & Chapelot, 2010; Recordati, 2003), while SNS acts as the "fight or flight" response (Bers & Despa, 2009; Curtis & O'Keefe Jr, 2002; Jansen et al., 1995). In most cases, the SNS and PSNS have opposite reactions (i.e. one is inhibited, while the other is activated), though in certain situations both may be activated simultaneously.

When the person experiences changes in cognitive workload or stress, the SNS gets activated, and the PSNS may withdraw to allow the body to enter into the "fight or flight" response (Ruscio et al., 2017), causing changes in the physiological systems of the body, such as respiratory and cardiovascular. For example, under high workload levels, heart rate and electrodermal activity will increase, while heart rate variability tends to decrease.

Physiological measures offer several advantages over the other two methods to measure cognitive workload (subjective measures and performance), such as being able to provide data in real-time, and the ability to provide information of the overall state of the person. For example, physiological measures can be analyzed in terms of SNS and PSNS to detect when the systems are activated, and how this is reflected in terms of workload. Furthermore, PSNS is the system used for recovery, and SNS can be activated by changes in workload or stress. Understanding the patterns between these two systems over an extended time period can provide information on the functioning of the ANS, as well as when the person may be reaching their limit.

This research investigated the role of physiological measures as indicators of the redline of cognitive workload threshold, that is, when the person approaches their overloaded state. Additionally, the subjective ratings, performance-based measures, and physiological indices were studied as the person approaches their redline. Finally, the role of the SNS was studied as indicator of when someone may be approaching their redline.

Research Questions

This section provides a summary of the results and discussion centered around each of the three research questions explored in this research.

Research Question 1

Which physiological measures are more sensitive to changes in workload at very high levels that are near or exceeding the redline?

The first research question was explored with the first study (Chapter 2), involving a multitasking environment similar to a flight cockpit. The Multi-Attribute Task Battery-II (MATB-II) environment was used to impose varying levels of workload, while physiological measures were collected throughout the study.

Findings from the first study suggest physiological measures will show an asymptote pattern as the person approached their redline threshold. Some measures were more sensitive than others, in this case, heart rate variability and skin conductance. Performance measures degraded around the same level as the physiological measures started the asymptotic pattern. Furthermore, subjective measures indicated that the participants felt the workload kept increasing. Together, these results show that there may be a similar pattern aspect among physiology, subjective measures, and performance. Future work needs to address the sensitivity of physiological measures at detecting the threshold, in addition to studying the similar patterns among physiological measures, performance, and subjective measures. Mostly, because the main advantage of physiology-based redline detection is the fact that these measures can be collected in real time, and can be used to track the person’s overall state, not just the workload.

Research Question 2

To what extent do changes in these sensitive physiological variables compare with other standard measures of task performance and subjective measures?

The second study presented in Chapter 3 addresses some of the future work proposed in Chapter 2. A similar study to the one presented in Chapter 2 to answer the first research question was performed. In this case, pupillometry was introduced as a physiological measure that does not depend on electrical signals, while electrodermal activity was kept from the previous study as a reference point. Similar to the previous study, subjective ratings and performance-based measures were also collected as additional data to detect the redline using different data sources.

Results showed a pattern between workload and performance, with increased workload resulting in performance degradation. While this study explored the relationship between the physiological asymptotic patterns and performance degradation as the person nears their redline threshold, future studies need to further explore the role of the ANS in the asymptotic pattern. Understanding the role of the ANS can provide additional information, such as if the person is recovering after the task (PSNS activation) or if the person's SNS is still highly activated.

Research Question 3

To what extent do physiological measures indicate the redline of cognitive workload based on patterns of the different branches of the autonomic nervous system?

The third and final question (Chapter 4) investigated in this body of work explored the redline from an ANS point of view. Particularly, EDA and ECG measures were collected, which then were analyzed based on their ANS activity. The study investigated the impact of different workload levels on physiological measures as the person approached their redline. The study involved solving 3D puzzles in a virtual reality (VR) environment.

Results indicate that SNS activity increases with added workload levels, but when the person is approaching their redline, the SNS remains active. This means that when the person is approaching their overloaded state, the SNS is already activated, but close to the redline threshold, no other changes are experienced, and based on the data analysis, heart automaticity no longer increases.

A limited body of research have found similar patterns in physiology, concurring with the physiological asymptotic pattern. For example, (Vogel & Machizawa, 2004) found event-related potentials reach an asymptote pattern when the visual working memory gets overloaded. There is a threshold to the available mental resources, and identifying this threshold can prevent performance degradation.

A limitation found in all three studies is that data were collected in a controlled, laboratory setting, and physiological measures were only analyzed for each of the experimental scenarios. Thus, it can be concluded that while physiological changes show an asymptotic pattern, in real-life situations this pattern may not be as distinguishable as physiological measures tend to contain noise. This limitation can be addressed by the future work proposed, where wearable devices can be implemented to obtain physiological data over a longer period of time.

Proposed Redline of Cognitive Workload Risk Matrix

The proposed redline diagram presented in chapter 1 (Figure 3) can be applied to each of the three studies to identify the redline. For example, taking into consideration the HRV results from Chapter 2, by workload level 5 physiological measures have stopped changing, and workload level 5 is significantly different from workload level 1, which is the condition for a level 3 in the physiological classification in the risk matrix. NASA-TLX results show a similar pattern, where by workload level 5, we can infer participants were feeling overloaded based on the TLX score as well as performance, and workload level 5 is significantly different from workload level 1, which would also grant a level 3 classification. The third classification, performance, also has the last workload level as significantly different from workload level 1, which would also grant a level 3 classification in the risk matrix. Based on the matrix classification, it would be in red zone, and the participant may be at risk of being very close to reaching their cognitive redline. Since the results from Chapter 2 and 3 are very similar, these results would also apply to Chapter 3.

However, the third study showed different results. In this case, a physiological classification of 2 would be awarded for the physiological classification, as the average heart rate measures in workload level 5 has increased from the baseline, but is not significantly different. The subjective classification can be assumed to be at level 3, as based on performance and the TLX score, participants may be feeling overwhelmed, and the TLX score is significantly different from workload level 1. Similarly, the performance classification can be assumed to be a level 3, as performance is statistically significant from workload level 1. This will place the classification on the yellow-red gradient, where the person is approaching their redline.

In these cases, additional changes may be needed to improve safety and decrease risks. A limitation of the model is that it was developed based on the three studies presented in this body of research, in a very specific context, hence why the model is color coded. The color coding will allow other domains and naturalistic studies to adapt the model to the data collected. If the model is to be implemented in further studies, then additional data will be needed to understand how the data collected in naturalistic studies fill fit into the cognitive redline risk matrix. Future studies need to implement the diagram while designing the tasks and collecting additional subjective ratings which specifically inquire about workload perception.

General Conclusions

Through the three studies, it was discovered that sympathetic arousal increases when the person experiences high levels of workload, which leads to an asymptotic physiological pattern when the person approaches their redline. Understanding these physiological patterns can lead to preventing performance decrements by providing different strategies. For example, tasks can be redesigned early in the design phase to ensure proper workload levels (i.e., not too high nor too low). Additionally, automation could be incorporated to facilitate certain aspects of the tasks in multitasking environments.

It is vital to stress the importance of studying the role of the parasympathetic nervous system. While most cognitive workload studies have emphasized the role of sympathetic arousal, and the PSNS system has mostly been overlooked, with only a few studies citing PSNS results (Dias et al., 2018; Hughes et al., 2019; Muth et al., 2012; Reimer & Mehler, 2011; Tjolleng et al., 2017). However, other studies involving recovery after sports have fully investigated the role of PSNS and overworking (Chen et al., 2011; Pierpont et al., 2000). Some of the recovery theories investigated after exercise could be applied to cognitive workload by investigating the role of the PSNS as an aspect of cognitive recovery after difficult or extended multitasking. Cognitive recovery is important, as it can prevent burnout. Measuring PSNS activity to investigate recovery will require additional physiological measurements, such as the trends as the person performs their daily activities. Such data may be provided by wearable devices.

Wearable devices are now becoming more common and affordable, and can collect physiological measures in real-time. As these are objective measures, meaning they can be observed directly from the subject (Orlandi & Brooks, 2018), they reflect the changes posed by changes in daily workload or stress. While some changes are immediate, for example, getting stressed about performing a task will immediately increase heart rate, there are other considerations to study. Additional studies are needed to really understand the connection between the short-term, task related changes, and how these will reflect in the ANS at a later time. For example, wearable devices also offer the advantage of being able to provide data over long periods of time, instead of just the data collected during a controlled laboratory study. Understanding every day activities, but also getting the recovery data can provide more effective ways of looking at everyday cognitive workload and stress, particularly when this is related to work environments with high turnover and burnout.

Long-term tracking of the ANS to determine SNS and PSNS activity, especially during sleep, can provide insight into the recovery aspect to determine how the stress and cognitive workload and stress experienced in daily activities is reflected.

Besides studying daily activities, studying and understanding the recovery aspect can provide insight into how well the person is resting between day-to-day activities. Mostly, as adequate rest and sleep also contribute to adequate job performance, which, when the person is not well-rested, may increase the risk for errors and performance degradation, as the person reaches their redline threshold.

Long-term data can also be used in future studies to understand how the redline shifts based on different factors, as the results from the three studies found the redline to be an asymptotic limit. For example, expertise could change the amount of workload someone may experience before reaching their threshold, while not sufficient rest may make someone more susceptible to reach their redline faster, as the person nears burnout.

Other factors may also affect how people perceive their redline. For example, intrinsic motivation, which is stronger than extrinsic motivation (Cerasoli et al., 2014), may have a delaying effect in how soon people reach their redline. Besides motivation, other personal factors, such as stress and need for cognition could influence the redline threshold for each person. These are just a few factors that can be studied in future studies.

Caution should be taken when collecting physiological data, especially in naturalistic settings, where the influence of environmental stressors can also influence changes in physiology. There are recommendations to collect data using at least 2 measures (Brookhuis & de Waard, 2001; Callan, 1998; Shakouri et al., 2018; Tsang & Velazquez, 1996). For example, if physiological data needs to be collected, additional data can be collected in the form of subjective measures or performance. Subjective questionnaires can inquire about other stressors or influencers that caused changes in workload, while performance can provide a better look at how well was the person able to handle the imposed workload. Collected data using all three methods can help with the data triangulation to ensure the collected data was actually measuring changes in workload. Additionally, the results from the three studies suggest an asymptotic pattern for physiological measures, so future studies can use planned contrasts for the physiological analysis.

Task engagement is another concern. While some participants may volunteer in the studies, there is no guarantee that they will be fully engaged in the task. A simple way to change this would be to add incentives to participate in the study (extrinsic motivation). However, if EEG data is one of the physiological measures that will be collected, then the alpha wave could be used as a way to measure engagement. Decreases in alpha power could be interpreted as lack of engagement or maybe due to task shedding in a multitasking context when the imposed cognitive workload is too high and participants feel cognitively overwhelmed. To avoid these confounding factors, subjective measures that inquire about participants' strategies to task shedding could be used to analyze and compare the data.

Because of the individual nature of the redline, if the plan is to only use physiological measures, then it is important to consider the metabolic baseline of the person, as well as the metabolic cost of the environment. For example, moving the arms or legs will increase heart rate. Thus, it is important to consider the metabolic rate as the person performs their daily activities. Calculating the metabolic rate can then be used as the baseline for energy expenditure as the person performs their daily activities. Then, any other observed changes, after subjective measures and performance are used to reduce the possibility of environmental stressors, will be related to changes in cognitive workload. The metabolic cost of the environment will be particular helpful in long-term studies, where the person's activities are tracked over several days, or in naturalistic studies, where there is less control. The metabolic rate (MR) can be calculated based on heart rate. The heart rate and metabolic rate are measured in a controlled environment to determine the relationship, and then the heart rate measurements are taken and converted to MR estimates to ensure the estimates are correct (Green, 2011).

Analyzing all of these factors, and how the redline shifts would require employing a systematic approach to study workload. This systematic approach would encompass studying the workload through daily tasks, and consider other factors that affect the redline threshold, such as intrinsic and extrinsic motivation, need for cognition, and expertise, among others. The systematic approach could use long-term monitoring to understand how the redline "shifts" depending on the influence of all the factors.

Overall findings of this body of research suggest that the redline of cognitive workload is a threshold, and can be detected based on physiological changes. This threshold is particularly important to monitor, as the results from this research suggests performance degrades when the person approached their redline.

The research efforts provided in this dissertation contribute to the body of knowledge and provide a better understanding of the redline threshold of cognitive workload. The contributions are based on the investigation of different physiological measures, including heart rate, heart rate variability, and electrodermal activity, to detect when participants are approaching their redline. These results were correlated with subjective-based ratings and performance measures. The importance of this research is based on the complex relationships and measurable patterns between each branch (SNS and PSNS) of the autonomic nervous system, as related to changes in workload as the person approaches their redline. This research can be applied to study employee workload, burnout, and turnover, as more than half of the American workforce experience overload (IT Business Edge, 2014).

REFERENCES

Abi-Saleh, B., & Omar, B. (2010). Einthoven's triangle transparency: A practical method to explain limb lead configuration following single lead misplacements. *Reviews in Cardiovascular Medicine*, *11*(1), 33–38.

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical and Biological Engineering and Computing*, *44*(12), 1031–1051.

Albuquerque, I., Tiwari, A., Gagnon, J.-F., Lafond, D., Parent, M., Tremblay, S., & Falk, T. (2018). On the analysis of eeg features for mental workload assessment during physical activity. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 538–543.

Alkahtani, M., Abidi, M. H., Ahmad, A., Darmoul, S., Samman, S., & Ghaleb, M. (2020). Human interruption management in workplace environments: An overview. *Engineering, Technology & Applied Science Research*, *10*(2), 5452–5458.

Alvarsson, J. J., Wiens, S., & Nilsson, M. E. (2010). Stress recovery during exposure to nature sound and environmental noise. *International Journal of Environmental Research and Public Health*, *7*(3), 1036–1046.

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., Hansen, Å. M. & Währborg, P. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest—Results from a pilot study. *Physiology & Behavior*, *118*, 240–250.

Bach, D. R. (2014). A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis. *Biological Psychology*, *103*, 63–68.

Backs, R. W. (1995). Going beyond heart rate: Autonomic space and cardiovascular assessment of mental workload. *The International Journal of Aviation Psychology*, *5*(1), 25–48.

Backs, R. W., & Seljos, K. A. (1994). Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. *International Journal of Psychophysiology*, *16*(1), 57–68.

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*(4), 243–254.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276.

Bennett, N., & Lemoine, G. J. (2014). What a difference a word makes: Understanding threats to performance in a VUCA world. *Business Horizons*, *57*(3), 311–317.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(Supplement 1), B231–B244.

Berntson, G. G., Quigley, K. S., Norman, G. J., & Lozano, D. L. (2017). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *The Handbook of Psychophysiology*. Cambridge University Press.

Bers, D. M., & Despa, S. (2009). Na/K-ATPase—An integral player in the adrenergic fight-or-flight response. *Trends in Cardiovascular Medicine*, *19*(4), 111–118.

Bilchick, K. C., & Berger, R. D. (2006). Heart rate variability. *Journal of Cardiovascular Electrophysiology*, *17*(6), 691–694.

Blanding, M. (2015). *National health costs could decrease if managers reduce work stress*. HBS Working Knowledge. http://hbswk.hbs.edu/item/national-health-costs-could-decrease-if-managers-reduce-work-stress

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Brasil, I. S., Neto, F. M. M., Chagas, J. F. S., de Lima, R. M., Souza, D. F. L., Bonates, M. F., & Dantas, A. (2011). An intelligent agent-based virtual game for oil drilling operators training. *2011 XIII Symposium on Virtual Reality*, 9–17.

Brookhuis, K. A., & de Waard, D. (2001). Assessment of drivers' workload: Performance and subjective and physiological indexes. *Stress, Workload and Fatigue*.

Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, *9*(4), 045008.

Cacioppo, J. T., Martzke, J. S., Petty, R. E., & Tassinary, L. G. (1988). Specific forms of facial EMG response index emotions during an interview: From Darwin to the continuous flow hypothesis of affect-laden information processing. *Journal of Personality and Social Psychology*, *54*(4), 592.

Callan, D. J. (1998). Eye movement relationships to excessive performance error in aviation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*, 1132–1136.

Carayon, P., & Alvarado, C. J. (2007). Workload and patient safety among critical care nurses. *Critical Care Nursing Clinics of North America*, *19*(2), 121–129.

Catterall, W. (2015). Regulation of cardiac calcium channels in the fight-or-flight response. *Current Molecular Pharmacology*, *8*(1), 12–21.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980.

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, *74*, 221–232.

Chen, J.-L., Yeh, D.-P., Lee, J.-P., Chen, C.-Y., Huang, C.-Y., Lee, S.-D., Chen, C.-C., Kuo, T. B., Kao, C.-L., & Kuo, C.-H. (2011). Parasympathetic nervous activity mirrors recovery status in weightlifting performance after training. *The Journal of Strength & Conditioning Research*, *25*(6), 1546–1552.

Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, *17*(2), 229–239.

Cinaz, B., La Marca, R., Arnrich, B., & Tröster, G. (2010). Monitoring of mental workload levels. *International Conference on E-Health. Sn: IADIS*, *189*, 193.

Colle, H. A., & Reid, G. B. (2005). Estimating a mental workload redline in a simulated air-to-ground combat mission. *The International Journal of Aviation Psychology*, *15*(4), 303–319.

Comstock Jr, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research*.

Curtis, B. M., & O'Keefe Jr, J. H. (2002). Autonomic tone as a cardiovascular risk factor: The dangers of chronic fight or flight. *Mayo Clinic Proceedings*, *77*(1), 45–54.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *The handbook of psychophysiology*. (4th ed.). Cambridge University Press.

de Munck, J. C., Gonçalves, S. I., Mammoliti, R., Heethaar, R. M., & Da Silva, F. L. (2009). Interactions between different EEG frequency bands and their effect on alpha–fMRI correlations. *Neuroimage*, *47*(1), 69–76.

de Ribaupierre, S., Kapralos, B., Haji, F., Stroulia, E., Dubrowski, A., & Eagleson, R. (2014). Healthcare training enhancement through virtual reality and serious games. In *Virtual, Augmented Reality and Serious Games for Healthcare 1* (pp. 9–27). Springer.

Dell'Agnola, F., Cammoun, L., & Atienza, D. (2018). Physiological characterization of need for assistance in rescue missions with drones. *2018 IEEE International Conference on Consumer Electronics (ICCE)*, 1–6.

Dias, R. D., Ngo-Howard, M. C., Boskovski, M. T., Zenati, M. A., & Yule, S. J. (2018). Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *The British Journal of Surgery*, *105*(5), 491.

Endacott, R. (2012). *The continuing imperative to measure workload in ICU: Impact on patient safety and staff well-being*. Springer.

Eschen, H., Kötter, T., Rodeck, R., Harnisch, M., & Schüppstuhl, T. (2018). Augmented and virtual reality for inspection and maintenance processes in the aviation industry. *Procedia Manufacturing*, *19*, 156–163.

Facebook Technologies, LLC. (2021). *Oculus Rift | Oculus*. https://www.oculus.com/rift/

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, *56*(2), 171–184.

Fox, J. R., Park, B., & Lang, A. (2007). When available resources become negative resources: The effects of cognitive overload on memory sensitivity and criterion bias. *Communication Research*, *34*(3), 277–296.

Freed, M. (2000). Reactive prioritization. *Proceedings of the 2nd NASA International Workshop on Planning and Scheduling for Space*.

Fuller, M. D., Emrick, M. A., Sadilek, M., Scheuer, T., & Catterall, W. A. (2010). Molecular mechanism of calcium channel regulation in the fight-or-flight response. *Science Signaling*, *3*(141), ra70–ra70.

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, *83*(3), 269–275.

Garton, E. (2017, April 6). Employee Burnout Is a Problem with the Company, Not the Person. *Harvard Business Review*. https://hbr.org/2017/04/employee-burnout-is-a-problem-with-the-company-not-the-person

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 113–131.

Ghaderyan, P., & Abbasi, A. (2016). An efficient automatic workload estimation method based on electrodermal activity using pattern classifier combinations. *International Journal of Psychophysiology*, *110*, 91–101.

Ghezeljeh, T. N., Nasari, M., Haghani, H., & Loieh, H. R. (2017). The effect of nature sounds on physiological indicators among patients in the cardiac care unit. *Complementary Therapies in Clinical Practice*, *29*, 147–152.

Green, J. A. (2011). The heart rate method for estimating metabolic rate: Review and recommendations. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *158*(3), 287–304.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & John, M. S. (2008). The red-line of workload: Theory, research, and design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*, 1204–1208. http://pro.sagepub.com/content/52/18/1204.short

Guastello, S. J., Shircel, A., Malon, M., & Timm, P. (2015). Individual differences in the experience of cognitive workload. *Theoretical Issues in Ergonomics Science*, *16*(1), 20–52.

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. A. (2014). Workload overload modeling: An experiment with MATB II to inform a computational model of task management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 849–853.

Hailstone, J., & Kilding, A. E. (2011). Reliability and validity of the Zephyr™ BioHarness™ to measure respiratory responses to exercise. *Measurement in Physical Education and Exercise Science*, *15*(4), 293–300.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Heard, J., Fortune, J., & Adams, J. A. (2019). Speech workload estimation for human-machine interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 277–281.

Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., & Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Applied Ergonomics*, *61*, 31–43.

Helton, W. S. (2004). Validation of a short stress state questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238–1242. https://doi.org/10.1177/154193120404801107

Helton, W. S., Fields, D., & Thoreson, J. A. (2005). Assessing daily stress with the Short Stress State Questionnaire (SSSQ): Relationships with cognitive slips-failures. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*, 886–890. http://pro.sagepub.com/content/49/10/886.short

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during n-back task—Quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, *7*, 935.

Herwig, U., Satrapi, P., & Schönfeldt-Lecuona, C. (2003). Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. *Brain Topography*, *16*(2), 95–99.

Hidalgo-Muñoz, A. R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., & El-Yagoubi, R. (2018). Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology*, *128*, 62–69.

Homan, R. W., Herman, J., & Purdy, P. (1987). Cerebral location of international 10–20 system electrode placement. *Electroencephalography and Clinical Neurophysiology*, *66*(4), 376–382.

Hubbard, J. (2012). *The peripheral nervous system*. Springer Science & Business Media.

Hughes, A. M., Hancock, G. M., Marlow, S. L., Stowers, K., & Salas, E. (2019). Cardiac measures of cognitive workload: A meta-analysis. *Human Factors*, *61*(3), 393–414.

Huikuri, H. V., Valkama, J. O., Airaksinen, K. E., Seppänen, T., Kessler, K. M., Takkunen, J. T., & Myerburg, R. J. (1993). Frequency domain measures of heart rate variability before the onset of nonsustained and sustained ventricular tachycardia in patients with coronary artery disease. *Circulation*, *87*(4), 1220–1228.

Hutter, O. F., & Trautwein, W. (1956). Vagal and sympathetic effects on the pacemaker fibers in the sinus venosus of the heart. *The Journal of General Physiology*, *39*(5), 715–733.

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction. *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, 1477–1480. http://dl.acm.org/citation.cfm?id=986094

IT Business Edge. (2014). Workplace productivity killers – and how to combat them. *Kelton Global*. https://www.keltonglobal.com/recognition/workplace-productivity-killers-and-how-to-combat-them/

Ivory, J. D., & Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players' feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication*, *57*(3), 532–555.

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.

Jänig, W., & Häbler, H.-J. (2000). Specificity in the organization of the autonomic nervous system: A basis for precise neural regulation of homeostatic and protective body functions. *Progress in Brain Research*, *122*, 351.

Jansen, A. S., Van Nguyen, X., Karpitskiy, V., Mettenleiter, T. C., & Loewy, A. D. (1995). Central command neurons of the sympathetic nervous system: Basis of the fight-or-flight response. *Science*, *270*(5236), 644–646.

johnnielawson. (2013, July 2). *Relax 8 hours-relaxing nature sounds-sleep-natural calming water sound-forest bird song-waterfall*. https://www.youtube.com/watch?v=eKFTSSKCzWA&t=13258s&ab_channel=johnnielawson

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., & Garrett, A. T. (2012a). BioHarness^{TM} multivariable monitoring device: Part. I: validity. *Journal of Sports Science & Medicine*, *11*(3), 400.

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., & Garrett, A. T. (2012b). Bioharness^{TM} multivariable monitoring device: Part. II: reliability. *Journal of Sports Science & Medicine*, *11*(3), 409.

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., Mitchell, A. C., & Garrett, A. T. (2012). Field based reliability and validity of the bioharness™ multivariable monitoring device. *Journal of Sports Science & Medicine*, *11*(4), 643.

Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology*, *34*(2–3), 237–257.

Kamzanova, A. T., Matthews, G., Kustubayeva, A. M., & Jakupov, S. M. (2011). EEG indices to time-on-task effects and to a workload manipulation (cueing). *International Journal of Psychological and Behavioral Sciences*, *5*(8), 928–931.

Karenbach, C. (2005). Ledalab-a software package for the analysis of phasic electrodermal activity. *Allgemeine Psychologie, Institut Für Psychologie, Freiburg, Germany, Tech. Rep*.

Karpinsky, N. D., Chancey, E. T., & Yamani, Y. (2016). Modeling relationships among workload, trust, and visual scanning in an automated flight task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 1550–1554.

Kennedy, L., & Parker, S. H. (2017). Making MATB-II medical: Pilot testing results to determine a novel lab-based, stress-inducing task. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *6*(1), 201–208.

Kergoat, H., & Faucher, C. (1999). Effects of oxygen and carbogen breathing on choroidal hemodynamics in humans. *Investigative Ophthalmology & Visual Science*, *40*(12), 2906–2911.

Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica*, 1(30).

Klingner, J., Kumar, R., & Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 69–72. http://dl.acm.org/citation.cfm?id=1344489

Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., & Zarjam, P. (2011). Measuring cognitive workload with low-cost electroencephalograph. *IFIP Conference on Human-Computer Interaction*, 568–571.

Kosch, T., Karolus, J., Ha, H., & Schmidt, A. (2019). Your skin resists: Exploring electrodermal activity as workload indicator during manual assembly. *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, 1–5.

Kotlyar, M., Donahue, C., Thuras, P., Kushner, M. G., O'Gorman, N., Smith, E. A., & Adson, D. E. (2008). Physiological response to a speech stressor presented in a virtual reality environment. *Psychophysiology*, *45*(6), 1034–1037.

Laguna, P., Caminal, P., Jané, R., & Rix, H. (1990). Evaluation of HRV by PP and RR interval analysis using a new time delay estimate. *[1990] Proceedings Computers in Cardiology*, 63–66.

Lean, Y., & Shan, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(3), 177–187.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for people: An introduction to human factors engineering*. CreateSpace.

Levy, M. N., & Martin, P. J. (1984). Neural control of the heart. In *Physiology and Pathophysiology of the Heart* (pp. 337–354). Springer.

Lin, L., & Liu, X. (2012). Design of training system of oil field simulation based on virtual reality technology. *Computer Technology and Development*, *10*.

Lubar, J. F., Swartwood, M. O., Swartwood, J. N., & O'Donnell, P. H. (1995). Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in TOVA scores, behavioral ratings, and WISC-R performance. *Biofeedback and Self-Regulation*, *20*(1), 83–99.

Luque-Casado, A., Perales, J. C., Cárdenas, D., & Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, *113*, 83–90.

Macedonio, M. F., Parsons, T. D., Digiuseppe, R. A., Weiderhold, B. A., & Rizzo, A. A. (2007). Immersiveness and physiological arousal within panoramic video-based virtual reality. *Cyberpsychology & Behavior*, *10*(4), 508–515.

Mackay, W. (2000). Responding to cognitive overload: Co-adaptation between users and technology. *Intellectica*, *30*(1), 177–193.

Mai, J. K., & Paxinos, G. (2011). *The human nervous system*. Academic press.

Maior, H. A., Wilson, M. L., Locke, C., & Swann, D. (2018). *Brain activity and mental workload associated with artistic practice*.

Malik, M. (1998). Heart rate variability. *Current Opinion in Cardiology*, *13*(1), 36–44.

Mantovani, F., Castelnuovo, G., Gaggioli, A., & Riva, G. (2003). Virtual reality training for health-care professionals. *CyberPsychology & Behavior*, *6*(4), 389–395.

Marion, N., Septseault, C., Boudinot, A., & Querrec, R. (2007). Gaspar: Aviation management on an aircraft carrier using virtual reality. *2007 International Conference on Cyberworlds (CW'07)*, 15–22.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1).

Matthews, G., & Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cognition and Emotion*, *24*(2), 357–373.

Matthews, G., Reinerman-Jones, L., Abich IV, J., & Kustubayeva, A. (2017). Metrics for individual differences in EEG response to cognitive workload: Optimizing performance prediction. *Personality and Individual Differences*, *118*, 22–28.

Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the dundee stress state questionnaire. *Psychology of Stress: New Research*, *1*, 49–90.

McCracken, J. H., & Aldrich, T. B. (1984). *Analyses of selected LHX mission functions: Implications for operator workload and system automation goals*. Anacapa Sciences Inc.Fort Rucker, AL.

McKendrick, R., Feest, B., Harwood, A., & Falcone, B. (2019). Theories and methods for labelling cognitive workload: Classification and transfer learning. *Frontiers in Human Neuroscience*, *13*, 295.

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12.

Mehler, B., Reimer, B., & Wang, Y. (2011). *A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload*.

Melo, H. M., Martins, T. C., Nascimento, L. M., Hoeller, A. A., Walz, R., & Takase, E. (2018). Ultra-short heart rate variability recording reliability: The effect of controlled paced breathing. *Annals of Noninvasive Electrocardiology*, *23*(5), e12565.

Milczarek, M., Brun, E., Houtman, I., Goudswaard, A., Evers, M., Bovenkamp, M., Roskams, N., Op de Beeck, R., Pahkin, K., & Berthet, M. (2007). *Expert forecast on emerging psychosocial risks related to occupational safety and health*.

Milla, K., Bakhshipour, E., Bodt, B., & Getchell, N. (2019). Does movement matter? Prefrontal cortex activity during 2D vs. 3D performance of the Tower of Hanoi Puzzle. *Frontiers in Human Neuroscience*, *13*, 156.

Mitchell, D. K., Samms, C. L., Henthorn, T., & Wojciechowski, J. Q. (2003). *Trade study: A two-versus three-soldier crew for the mounted combat system (MCS) and other future combat system platforms*. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD.

Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., & Yamamura, T. (2009). Physiological responses to workload change. A test/retest examination. *Applied Ergonomics*, *40*(6), 987–996.

Modesti, P. A., Ferrari, A., Bazzini, C., Costanzo, G., Simonetti, I., Taddei, S., Biggeri, A., Parati, G., Gensini, G. F., & Sirigatti, S. (2010). Psychological predictors of the antihypertensive effects of music-guided slow breathing. *Journal of Hypertension*, *28*(5), 1097–1103.

Muth, E. R., Moss, J. D., Rosopa, P. J., Salley, J. N., & Walker, A. D. (2012). Respiratory sinus arrhythmia as a measure of cognitive workload. *International Journal of Psychophysiology*, *83*(1), 96–101.

NeuroSky, Inc. (2014). *EEG band power values: Units, amplitudes, and meaning / development / knowledge base—NeuroSky—Home Page Support*. http://support.neurosky.com/kb/development-2/eeg-band-power-values-units-amplitudes-and-meaning

NeuroSky, Inc. (2015). *NeuroSky MindWave manual*.

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors*, *45*(4), 575–590.

Nixon, J., & Charles, R. (2017). Understanding the human performance envelope using electrophysiological measures from wearable technology. *Cognition, Technology & Work*, *19*(4), 655–666.

Offenloch, K., & Zahner, G. (1990). *Computer aided physiological assessment of the functional state of pilots during simulated flight*. Frankfurt University (Germany).

Ohsuga, M., Shimono, F., & Genno, H. (2001). Assessment of phasic work stress using autonomic indices. *International Journal of Psychophysiology*, *40*(3), 211–220.

Ohsuga, M., Terashita, H., Shimono, F., & Toda, M. (1995). Assessment of mental workload based on a model of autonomie regulations on the cardiovascular system. In *Advances in Human Factors/Ergonomics* (Vol. 20, pp. 771–776). Elsevier.

Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied Ergonomics*, *69*, 74–92.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 141–144.

Panopreter. (2021). *The text to speech software reads any text aloud, converts text to audio files.* https://www.panopreter.com/en/index.php

Parsons, T. D., & Reinebold, J. L. (2012). Adaptive virtual environments for neuropsychological assessment in serious games. *IEEE Transactions on Consumer Electronics*, *58*(2), 197–204.

Peltola, M. (2012). Role of editing of RR intervals in the analysis of heart rate variability. *Frontiers in Physiology*, *3*, 148.

Pergher, V., Wittevrongel, B., Tournoy, J., Schoenmakers, B., & Van Hulle, M. M. (2019). Mental workload of young and older adults gauged with ERPs and spectral power during N-Back task performance. *Biological Psychology*, *146*, 107726.

Phillips, J. B., & Boles, D. B. (2004). Multiple resources questionnaire and workload profile: Application of competing models to subjective workload measurement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(16), 1963–1967.

Pichon, A., & Chapelot, D. (2010). *Homeostatic role of the parasympathetic nervous system in human behavior*. Nova Science Publisher’s.

Pierpont, G. L., Stolpman, D. R., & Gornick, C. C. (2000). Heart rate recovery post-exercise as an index of parasympathetic activity. *Journal of the Autonomic Nervous System*, *80*(3), 169–174.

Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. *Proceedings of the International Conference on HCI*. http://www.cs.umb.edu/~marc/pubs/pomplun_hci2003.pdf

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, *40*(1–2), 187–195.

Porges, S. W. (1992). Vagal tone: A physiologic marker of stress vulnerability. *Pediatrics*, *90*(3), 498–504.

Portoghese, I., Galletta, M., Coppola, R. C., Finco, G., & Campagna, M. (2014). Burnout and workload among health care workers: The moderating role of job control. *Safety and Health at Work*, *5*(3), 152–157.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology*, *10*(4), 393–410.

PupilLabs. (2019). *Eye tracking technology—Gain insight into human behavior*. Pupil Labs. https://pupil-labs.com

Ranogajec, S., & Geršak, G. (2014). Measuring site dependency when measuring skin conductance. *Proceedings of the Twenty-Third International Electrotechnical and Computer Science Conference, Portorož, Slovenia*, 22–24.

Realtime Technologies Ltd. (2017). *Shimmer user manual*.

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.

Recordati, G. (2003). A thermodynamic model of the sympathetic and parasympathetic nervous systems. *Autonomic Neuroscience*, *103*(1–2), 1–12.

Reid, G. B., & Colle, H. A. (1988). Critical SWAT values for predicting operator overload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *32*, 1414–1418. http://pro.sagepub.com/content/32/19/1414.short

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics*, *54*(10), 932–942.

Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., & Tan, C. (2009). An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers. *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 115–118.

Rizzo, L., Dondio, P., Delany, S. J., & Longo, L. (2016). Modeling mental workload via rule-based expert system: A comparison with NASA-TLX and workload profile. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 215–229.

Rizzo, L., & Longo, L. (2017). Representing and Inferring Mental Workload via Defeasible Reasoning: A Comparison with the NASA Task Load Index and the Workload Profile. *AI^3@ AI* IA*, 126–140.

Rocchetti, M., Malfatto, G., Lombardi, F., & Zaza, A. (2000). Role of the input/output relation of sinoatrial myocytes in cholinergic modulation of heart rate variability. *Journal of Cardiovascular Electrophysiology*, *11*(5), 522–530.

Rodriguez Paras, C., Yang, S., Tippey, K., & Ferris, T. K. (2015). Physiological indicators of the cognitive redline. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*, 637–641. http://pro.sagepub.com/content/59/1/637.short

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, *34*(2–3), 259–287.

Rouvrais, S., Gaultier LeBris, S., & Stewart, M. (2018). *Engineering students ready for a vuca world? A design based research on decisionship*.

Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2016). Efficient workload classification based on ignored auditory probes: A proof of concept. *Frontiers in Human Neuroscience*, *10*, 519.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*(1), 61–86.

Rundmo, T., Hestad, H., & Ulleberg, P. (1998). Organisational factors, safety attitudes and workload among offshore oil personnel. *Safety Science*, *29*(2), 75–87.

Ruscio, D., Bos, A. J., & Ciceri, M. R. (2017). Distraction or cognitive overload? Using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accident Analysis & Prevention*, *103*, 105–111.

Salahuddin, L., Cho, J., Jeong, M. G., & Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. *2007 29th Annual International Conference of the Ieee Engineering in Medicine and Biology Society*, 4656–4659.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. (2011). *The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide*. http://ntrs.nasa.gov/search.jsp?R=20110014456

SAS Institute Inc. (2021). *SAS 9.4 software overview for the customer*. https://support.sas.com/software/94/

Schmidt, T. A., Kourdali, H. K., & Nixon, J. (2021). Evaluating process-based and crew-centred approaches to procedure design in aviation: Workload and performance changes in go-around manoeuvres. *Applied Ergonomics*, *90*, 103244.

Schmidt, & Walach, H. (2000). Electrodermal activity (eda)–state-of-the-art measurements and techniques for parapsychological purposes. *Journal of Parapsychology*, *64*(2).

Shakouri, M., Ikuma, L. H., Aghazadeh, F., & Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones. *International Journal of Industrial Ergonomics*, *66*, 136–145.

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data: Two field studies. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, 4057–4066. http://dl.acm.org/citation.cfm?id=2557068

Song, H.-S., & Lehrer, P. M. (2003). The effects of specific respiratory rates on heart rate and heart rate variability. *Applied Psychophysiology and Biofeedback*, *28*(1), 13–23.

Splawn, J. M. (2013). *Applying Hyperspectral Imaging to Heart Rate Estimation for Adaptive Automation*.

Stein, P. K., Bosner, M. S., Kleiger, R. E., & Conger, B. M. (1994). Heart rate variability: A measure of cardiac autonomic tone. *American Heart Journal*, *127*(5), 1376–1381.

Sweeney, J. T., & Summers, S. L. (2002). The effect of the busy season workload on public accountants' job burnout. *Behavioral Research in Accounting*, *14*(1), 223–245.

Sztajzel, J. (2004). Heart rate variability: A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly*, *134*, 514–522.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International Journal of Environmental Research and Public Health*, *16*(15), 2716.

Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, *17*, 354–381.

Temko, A. (2017). Accurate heart rate monitoring during physical exercises using PPG. *IEEE Transactions on Biomedical Engineering*, *64*(9), 2016–2024.

Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, J., & Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, *59*, 326–332.

Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y., & Krusienski, D. J. (2019). Estimating cognitive workload in an interactive virtual reality environment using EEG. *Frontiers in Human Neuroscience*, *13*.

TrexLLC. (2014). *Paced Breathing—Apps on Google Play*. TrexLLC. https://play.google.com/store/apps/details?id=com.apps.paced.breathing&hl=en_US&gl=US

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, *39*(3), 358–381.

Tsien, R. W., Bean, B. P., Hess, P., Lansman, J. B., Nilius, B., & Nowycky, M. C. (1986). Mechanisms of calcium channel modulation by β-adrenergic agents and dihydropyridine calcium agonists. *Journal of Molecular and Cellular Cardiology*, *18*(7), 691–710.

Van Bogaert, P., Clarke, S., Willems, R., & Mondelaers, M. (2013). Nurse practice environment, workload, burnout, job outcomes, and quality of care in psychiatric hospitals: A structural equation model approach. *Journal of Advanced Nursing*, *69*(7), 1515–1524.

Van Bouwel, T. (2020). *Cubism on Steam*. https://store.steampowered.com/app/804530/Cubism/

van den Hombergh, P., Engels, Y., van den Hoogen, H., van Doremalen, J., van den Bosch, W., & Grol, R. (2005). Saying 'goodbye'to single-handed practices; what do patients and staff lose or gain? *Family Practice*, *22*(1), 20–27.

Van Wyk, E., & De Villiers, R. (2009). Virtual reality training applications for the mining industry. *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, 53–63.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(3), 323–342.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656–669.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751.

Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2015). Using wireless EEG signals to assess memory workload in the $ n $-back task. *IEEE Transactions on Human-Machine Systems*, *46*(3), 424–435.

Waterhouse, J., & Campbell, I. (2008). Reflexes: Principles and properties. *Anaesthesia & Intensive Care Medicine*, *9*(5), 210–215.

Werrlich, S., Nguyen, P.-A., & Notni, G. (2018). Evaluating the training transfer of Head-Mounted Display based training for assembly tasks. *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 297–302.

Whitaker, L. A., Hohne, J., & Birkmire-Peters, D. P. (1997). Assessing cognitive workload metrics for evaluating telecommunication tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *41*(1), 325–329.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*. Psychology Press.

Wilhelm, F. H., Grossman, P., & Coyle, M. A. (2004). Improving estimation of cardiac vagal tone during spontaneous breathing using a paced breathing calibration. *Biomedical Sciences Instrumentation*, *40*, 317–324.

Yang, Y., Reimer, B., Mehler, B., & Dobres, J. (2013). *A field study assessing driving performance, visual attention, heart rate and subjective ratings in response to two types of cognitive workload*.

Yeh, & Wang, W.-J. (2008). QRS complexes detection for ECG signal: The difference operation method. *Computer Methods and Programs in Biomedicine*, *91*(3), 245–254.

Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, *30*(1), 111–120.

Yentes, J. M., Hunt, N., Schmid, K. K., Kaipust, J. P., McGrath, D., & Stergiou, N. (2013). The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering*, *41*(2), 349–365.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.

Young, M. S., Birrell, S. A., & Davidsson, S. (2011). Task preloading: Designing adaptive systems to counteract mental underload. *Contemporary Ergononomics and Human Factors*, 168–176.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17.

Young, M. S., & Stanton, N. A. (2002a). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, *3*(2), 178–194.

Young, M. S., & Stanton, N. A. (2002b). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, *44*(3), 365–375.

Yu, C.-P., Lee, H.-Y., & Luo, X.-Y. (2018). The effect of virtual reality forest and urban environments on physiological and psychological responses. *Urban Forestry & Urban Greening*, *35*, 106–114.

Zephyr Corporation. (2012). *BioHarness3 manual*.

Zephyr Technology. (2016). *BioHarness 3 Log Data Descriptions*. https://www.zephyranywhere.com/media/download/bioharness-log-data-descriptions-07-apr-2016.pdf

Zhang, H. (2017). Head-mounted display-based intuitive virtual reality training system for the mining industry. *International Journal of Mining Science and Technology*, *27*(4), 717–722.

Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175.

APPENDIX A

BACKGROUND QUESTIONNAIRE FOR STUDIES 1 AND 2

"Physiological Measurements of Cognitive Redline"
Background Questionnaire

Please fill out this short background questionnaire. Thank you for participating in our study!

1. Age: ____________

2. Sex: M F

3. Classification: (Please circle one)

   Undergraduate student

   Graduate student: Master's Ph.D.

   Professional (please describe):_____________________

4. Do you play videogames? Yes No

If you selected "yes" to question 4, please answer questions 5 to 8. Otherwise continue to question 9.

5. How many hours per week you usually play?

   Less than 1 hour 1-4 hours 5-8 hours 9-10 hours More than 10 hours

6. What system do you use?

Xbox Playstation WiiU PC
Other______________

7. Does it involve the use of a joystick? Yes No

8. What type of games you play? Please explain.

9. Do you have any experience with flight simulators or similar software/games?

Yes No

If yes, please explain.

10. Do you consume caffeine on a daily basis?

Yes No

11. How many cups of coffee or tea per day, or how many sodas do you usually drink?

12. How many cups of beverages containing caffeine have you consumed today before the study?

APPENDIX B

BACKGROUND QUESTIONNAIRE FOR STUDY 3

Cognitive Workload in Virtual Reality Environments

Thank you for participating in our study! Please fill out this short background questionnaire.

1. Age: ________________

2. Sex: M F

3. Classification: (Please circle one)

Undergraduate student

Graduate student

Master's

Ph.D.

Professional (please describe):___________________

4. Do you play videogames? Yes No

If you selected "yes" to question 4, please answer questions 5 to 7. Otherwise continue to question 8.

5. How many hours per week do you usually play?

Less than 1 hour 1-4 hours 5-8 hours 9-10 hours More than 10 hours

6. Which system do you use?

Xbox Playstation WiiU Switch PC Other

7. What type of games you play? Please explain

8. Do you have experience with virtual reality? Yes No

If yes, please explain

9. Do you consume caffeine on a daily basis? (Please circle one)

Less than once a week1-2 times a week 3-5 times a week Daily Other

If other, please specify

10. If you answered yes to the previous question, on average, how many cups per day?

11. Have you had any caffeinated drinks today? How long ago?

APPENDIX C

COVID-19 REOPENING PROTOCOL

COVID Reopening Plan

Cleaning Procedures

- Study personnel will clean and disinfect all surfaces in contact with the participants, which include tables, computer mouse and keyboard. The hardware needed for the study will also be disinfected, including the Oculus Rift headset and controllers, and the Biopac sensors. The electrodes attached to the Biopac sensors are disposable, so each participant will use new electrodes. This disinfecting process will occur before every participant to ensure everything is clean.
- Additionally, the door handles to the lab and the light switches will be disinfected every morning and at the end of the day.
- Hand sanitizer will also be available at the lab.

Personal Protective Equipment

- Experimenters and participants will wear a mask during the duration of the study. Hand sanitizer will be available.
- A distance of 6 feet will be kept at all times between the experimenters and the participant.

Schedule

- The study is expected to last no more than 150 minutes. Participants will be scheduled with at least 30 minutes in between participants to ensure enough time to clean and disinfect the equipment.
- Study participation requires only one visit per participant.