Spring 2018

A comparison of linear and nonlinear ECG-based methods to assess pilot workload in a live-flight tactical setting

Christopher Patrick Reichlen
University of Iowa

Recommended Citation
Reichlen, Christopher Patrick. "A comparison of linear and nonlinear ECG-based methods to assess pilot workload in a live-flight tactical setting." MS (Master of Science) thesis, University of Iowa, 2018.
https://doi.org/10.17077/etd.hqx7edns

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Industrial Engineering Commons

A COMPARISON OF LINEAR AND NONLINEAR ECG-BASED METHODS TO ASSESS PILOT WORKLOAD IN A LIVE-FLIGHT TACTICAL SETTING

by

Christopher Patrick Reichlen

A thesis submitted in partial fulfillment
of the requirements for the Master of Science
degree in Industrial Engineering in the
Graduate College of
The University of Iowa

May 2018

Thesis Supervisor: Professor Thomas Schnell

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

MASTER'S THESIS

This is to certify that the Master's thesis of

Christopher Patrick Reichlen

has been approved by the Examining Committee for the thesis requirement for the Master of Science degree in Industrial Engineering at the May 2018 graduation.

Thesis Committee: ___________________________________

Thomas Schnell, Thesis Supervisor

____

Priyadarshini Pennathur

____

Daniel McGehee

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Tom Schnell, and his team at the OPL for providing me a one-of-a-kind educational experience. I could not have been more impressed with the professionalism and expertise of this research team.

I would also like to thank my thesis committee, Dr. Priyadarshini Pennathur and Dr. Daniel McGehee for their mentorship through this process. In addition, I would like to thank all members of the Industrial Engineering faculty for their contributions to my educational development over the last two years.

I would like to thank USAF and AFIT for affording me to opportunity to pursue a graduate degree at the University of Iowa. I would also like to thank the 711 HPW and AFRL for funding and allowing me to contribute to the spatial disorientation research which motivated this thesis.

Lastly, I would like to thank my wife Cat for her love, support, and most of all, patience throughout the course of my graduate studies.

ABSTRACT

This research compares methods for measuring pilot mental workload (MWL) from the electrocardiogram (ECG) signal. ECG-based metrics have been used extensively in MWL research. Heart rate (HR) and heart-rate variability (HRV) exhibit changes in response to varying levels of task demand. Classical methods for HRV analysis examine the ECG signal in the linear time and frequency domains. More contemporary research has advanced the notion that nonlinear elements contribute to cardiac control and ECG signal generation, spawning development of analytical techniques borrowed from the domain of nonlinear dynamics (NLD). Applications of nonlinear HRV analysis are substantial in clinical diagnosis settings; however, such applications are less frequent in MWL research, especially in the aviation domain. Specifically, the relative utility of linear and non-linear HRV analysis methods has not been fully assessed in pilot MWL research.

This thesis contributes to aforementioned research gap by comparing a non-linear HRV method, utilizing transition probability variances (TPV), to classical time and frequency domain methods, focusing the analysis on sensitivity and diagnosticity. ECG data is harvested from a recent study characterizing spatial disorientation (SDO) risk amongst three candidate off-boresight (OBS) helmet-mounted display (HMD) symbologies in a tactically relevant live-flight task. A comparative analysis of methods on this dataset and supplemental workload analysis for the HMD study are presented. Results indicate the TPV method may exhibit higher sensitivity and diagnosticity than classical methods. However, limitations of this analysis warrant further investigation into this question.

PUBLIC ABSTRACT

This research compares methods for measuring pilot mental workload (MWL) from the electrocardiogram (ECG) signal. The ECG signal reflects changes in heart rhythms associated with the human response to task demands. Classical analytical methods employ basic statistical summaries and spectrum estimates. More robust contemporary methods leverage the complexity of the ECG signal. Applications of these more complex methods are substantial in clinical diagnosis settings; however, they are largely underutilized in MWL applications, especially in aviation settings. Specifically, the relative utility of linear and non-linear HRV analysis methods has not been fully assessed in pilot MWL research.

This thesis contributes to aforementioned research gap by comparing a novel method, utilizing transition probability variances (TPV), to classical methods, focusing the analysis on sensitivity and diagnosticity. ECG data is harvested from a recent study characterizing spatial disorientation (SDO) risk amongst three candidate helmet-mounted display (HMD) symbologies in a tactically relevant live-flight task. A comparative analysis of methods on this dataset and supplemental workload analysis for the HMD study are presented. Results indicate the TPV method may exhibit higher sensitivity and diagnosticity than classical methods. However, limitations of this analysis warrant further investigation into this question.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ECG – Electrocardiogram

CAS – Close Air Support

CDF – Current Display Format

DFR – Distributed Flight Path Reference

ETM – Ergodic Transition Matrix

HF – High Frequency

HMD – Helmet-Mounted Display

HRV – Heart Rate Variability

LF – Low Frequency

MWL – Mental Workload

NDFR – Non-Distributed Flight Path Reference

OBS – Off-boresight

RMSSD – Root Mean Square of Successive Differences

SDNN – Standard Deviation of the NN Interval

SDO – Spatial Disorientation

TPV – Transition Probability Variance

vHUD – Virtual Heads-Up Display

CHAPTER 1 – INTRODUCTION

Synopsis

This thesis examines physiologic measurement of pilot mental workload (MWL) in a live-flight, tactically relevant flying task. The specific objective is to compare analytical methods for MWL measurement using metrics derived from the electrocardiogram (ECG) signal in the linear and nonlinear domains. A newer, nonlinear method is compared to classical linear time and frequency domain methods on the same ECG dataset. The analysis focuses on sensitivity and diagnosticity of the methods and assesses their utility in a live-flight operational setting.

Background and Motivation

Mental workload (MWL) is an extensively researched topic in human factors and has become increasingly important as the introduction of new technology continues to impose greater cognitive demands (M. S. Young, Brookhuis, Wickens, & Hancock, 2015). While many definitions exist, mental workload may be defined as the costs a human operator incurs in the performance of a task (Kramer, 1991), or costs incurred accomplishing mission requirements (Hart, 2006). It is a concept of obvious importance in a high-stakes domain such as aviation, particularly in modern fighter aircraft, where it has direct implications for safety of flight and mission effectiveness.

The modern fighter pilot must balance multiple concurrent tasks such as navigation, communication, threat avoidance, and weapons employment. As such, managing cockpit tasks inherently places high demands on a pilot's finite attentional resources. If the cost of meeting such demands exceeds available mental resources, safety and mission effectiveness suffer. Non-optimal levels of MWL are inherently linked to situation awareness (SA) related errors, which often have costly and deadly consequences (Wickens, 2002b). MWL considerations are of growing importance in this context given the evolution of technology apparent in 5th generation fighter aircraft, such as the F-35, compared to legacy fighters. Sensor and information capabilities are more robust than ever. The preponderance of air-to-ground and beyond visual range (BVR) air-to-air employment has made shifted demands of the fighter pilot more into the cognitive than physical domain. Further, smaller inventories funnel more tactical information and SA requirements to a single aircraft and its pilot. Such factors have the potential to increasingly push human MWL limits and must be effectively managed.

Given these considerations, MWL measurement is naturally an area of interest in aircraft systems development during which there exists a need to quantify the mental cost of performing tasks and predict system performance (Cain, 2007). Nearly a half-century of research has generated a wide range of MWL measurement techniques, which fall under three general categories: subjective measures, task performance measures (primary and secondary), and physiologic measures (Wierwille & Eggemeier, 1993). Subjective methods are based on the operator's perception of task demands and are usually administered via rating scales and questionnaires. Performance measures, as the name implies, characterize the operator's performance in specific task, capturing ability to perform the primary task and the amount of spare mental capacity via the secondary task performance. Physiologic methods measure biological signals to infer nervous system activity known to vary as a function of an operator's response to task demands.

With so many available methods, there exists a challenge selecting a method, given no single one is universally considered superior in every setting. In most cases, selection of the appropriate method is context dependent. Suitability may vary based on the mental resource type associated with the task in question (e.g. working memory, attention), characteristics of the operator (e.g. expert vs. novice), and the scope and limitations of the research (M. S. Young et al., 2015). In general, MWL measurement methods must be adequately sensitive to changes in MWL levels and diagnostic with respect to workload-driving elements of the task. Further, these methods must be suitable and effective in both laboratory (i.e. simulator) and field (i.e. live flight) environments to characterize MWL in both controlled experiments and operationally relevant settings.

Physiologic methods constitute just one category of methods, but the category of specific interest for the research herein. These methods appeal to practitioners since they provide continuous objective measurement with little to no impact on task execution. Cardiovascular measures, most derived from the ECG signal, constitute the most widely-used set of physiologic MWL indices (Scerbo, Freeman, Parasuraman, Di Nocero, & Prinzel, 2001). Other methods include electroencephalogram (EEG), galvanic skin response (GSR), and eye-related measures. Given current technology, ECG remains one of the more appealing methods for applied MWL measurement given the accuracy, reliability and low footprint of the sensors (Fahrenberg & Wientjes, 1999). As such, ECG has been used successfully for this purpose in real-world settings for many years (Wilson, 1992).

Given their suitability for live-flight research, the focus of this thesis is further narrowed to ECG-based methods. The ECG signal reflects the dynamic regulation of heart rhythms by the autonomic nervous system (ANS) and thus provide numerous metrics that correlate to MWL (Kramer, 1991). Just as in the greater picture of MWL measurement, there exists controversy as to which ECG metric is superior (Henelius, Hirvonen, Holm, Korpela, & Muller, 2009). These metrics primarily include heart rate (HR) and heart rate variability (HRV). HRV is generally defined as the variance between successive cardiac cycles, or the normal-normal (NN) interval between QRS complexes in the ECG signal. As MWL increases, HR generally increases while HRV decreases (Wilson, 2002a).

HRV analysis is applied in multiple domains of science to include medicine and psychophysiological research. As a result, a wide variety of analysis techniques have been developed which leverage different characteristics of the signal (Kuo & Chen, 1998). These methods may be grouped in three general categories: time domain, frequency domain, and nonlinear domain. Metrics in each of these domains have demonstrated utility for various scientific objectives. Time and frequency domain methods constitute what is referred to herein as "classical" methods and have been by far more widely utilized in MWL research. More contemporary methods are based on the notion that ECG signal exhibits chaos as nonlinear elements contribute to cardiac control (Glass, 2009; Voss, Schulz, Schroeder, Baumert, & Caminal, 2009). Given this assumption, the ECG signal dynamics, and therefore the underlying system dynamics, are obscured in the linear domain. This has spawned the development of newer metrics based on analytical techniques borrowed from the domains of nonlinear dynamics and information theory (Voss et al., 2009; H. Young & Benton, 2015). It has been suggested that nonlinear HRV methods may complement, if not outperform, classical time and frequency domain methods (H. Young & Benton, 2015).

A number of studies have compared linear and nonlinear HRV metrics (Francesco et al., 2012; Schneider et al., 2017; H. Young & Benton, 2015). The majority, however, have been related to HRV analysis in clinical diagnosis (Francesco et al., 2012). A smaller number of studies have assessed the relative utility of these methods as compared to time and frequency domain methods in mental MWL research (Heine et al., 2017; Schneider et al., 2017; H. Young & Benton, 2015). Few, however, have pursued this question in aviation-specific research.

Statement of Problem

A nonlinear method developed at the University of Iowa Operator Performance Laboratory (OPL), based on the transition probability variance (TPV) of the ECG, has been used successfully used in multiple previous studies (J. Engler & Schnell, 2012). The method shows a promising capability to assess MWL with potentially higher sensitivity and diagnosticity than classical HRV methods. Further, these previous studies have shown it can be successfully employed in live-flight settings and capture MWL variations in near real time (Schnell, Reuter, & Cover, 2017). However, to date, the TPV method has not been compared to classical HRV methods on the same dataset.

This research compares the TPV-based method to classical time and frequency domain HRV methods with ECG data harvested from a recent helmet-mounted display (HMD) study.

The study compared performance impacts of three candidate HMD symbologies in a live-flight, tactical Close Air Support (CAS) scenario. The primary analysis presented herein focuses on the ECG data in this study and seeks to provide insight into the relative utility of various indices of MWL. The end goal is to provide useful insights for risk mitigation in the design of current and future systems.

Research Goals and Hypothesis

This thesis aims to address the aforementioned research gap by comparing the TPV method to classical HRV methods on the same ECG data in a live-flight, operationally relevant task. The specific aim is to explore how well the TPV method can characterize MWL in a tactically relevant flying task. Further, it aims to explore how well various HRV methods can elucidate the MWL-driving elements of a task.

It is hypothesized herein that the TPV-based MWL method will show increased sensitivity and diagnosticity compared to classical time and frequency domain methods in a short duration task with highly dynamic MWL characteristics.

Contributions

This research provides several theoretical and practical contributions. It extends theory of HRV by exploring a relatively new nonlinear analytical method based on TPV. The TPV method has shown promising results in previous studies and the analysis herein provides additional validation. While these findings may pertain more to MWL research, further exploration of this method could prove useful in other domains, such as clinical diagnosis.

Because of the applied study used as the basis for the analysis, this research weighs more heavily on practical contributions. The utility of the TPV method is demonstrated in a live-flight operationally relevant context. Thus, it has implications for research and flight test domains in which system performance in the "real-world" must be evaluated. The HMD study focuses on SDO prevention. The live-flight element is crucial for SDO research to produce the full spectrum of novel sensory inputs generated by linear and rotational accelerations which contribute to SDO. It also explores a method by which to infer MWL from the ECG in near real-time. Not only does this provide a useful analytical method for research settings, it may provide a useful method for modern applications of psychophysiological signals such as real-time operator state assessment and adaptive automation (Scerbo et al., 2001). While this research does not explore operator state modeling, it validates that ECG signal may be a useful element to these efforts.

Thesis Structure

This thesis is structured as follows. Chapter 2 expands on the theoretical background of MWL and its measurement with a focus on ECG-based methods. Classical time and frequency domain measures are reviewed. Nonlinear methods are introduced and the TPV method is explained in detail. Chapter 3 provides a summary of the HMD study which is the basis of the HRV analysis to follow. Key results are included which bear relevance to the research questions in this thesis. Chapter 4 discusses the methodology for HRV analysis. The final two chapters present results, discussion, and conclusions.

CHAPTER 2 – BACKGROUND AND LITERATURE REVIEW

Mental Workload

Mental workload (MWL) constitutes an enormous body of research spanning a half-century in multiple human factors domains to include aviation, ground transport, and medicine (Wickens, 2017). Despite its wide presence in the literature, there is no universally accepted definition. It may be defined simply as the costs a human operator incurs in the performance of a task (Kramer, 1991), or the costs incurred accomplishing mission requirements (Hart, 2006). These "costs" refer specifically to the cognitive resources of the operator, which are finite and limited (Wickens, 2008). When operator resources invested do not match task demands, performance is impacted. Therefore, MWL is naturally an important consideration in a high-stakes domain such as aviation (Wickens, 2002a).

The theoretical underpinnings of the relationship between tasks and operator resources can be explained by Multiple Resource Theory (MRT) (Wickens, 1984). MRT asserts that humans do not simply have one central pool of cognitive resources; rather, there are multiple pools that are utilized in parallel (Wickens, 2008). Conflict, or interference, occurs when multiple tasks require the same resource, such as visual or auditory attention, or the tasks exceed demands of one or more resources. Thus, MWL considerations are especially important in multitask environments such as in the cockpit of a single-pilot aircraft.

External task demands and operator resources are the two primary determinants of MWL (Vidulich & Tsang, 2012). Task demands may be driven by factors such as task complexity, time pressure, or environmental factors. Operator resources consist of the available resources to support cognitive processes such as attention, memory, planning, and decision making (Vidulich & Tsang, 2012). This resource-demand interaction is further moderated by operator skill and effort put forth while attending to tasks (Cain, 2007).

The primary rationale for studying MWL is its relationship to performance. The relationship is complex, in that increases in MWL do not necessarily result in decreased performance. As task demand increases, the operator may invest additional resources (e.g. effort, more efficient strategy) to cope with increased demand (Hockey, 1997). It is at suboptimal levels of MWL, rather, that performance failures occur (M. S. Young et al., 2015). Overload results in distraction, insufficient processing capacity, and divided attention, while underload results in inattention or reduced alertness (Brookhuis & de Waard, 2010). Thus, there is an optimum range of MWL associated with highest performance (Hancock, 1989). This relationship between performance and MWL is shown in Figure 1. "Activation level" represents to the level to which the operator invests resources to cope with task demands.

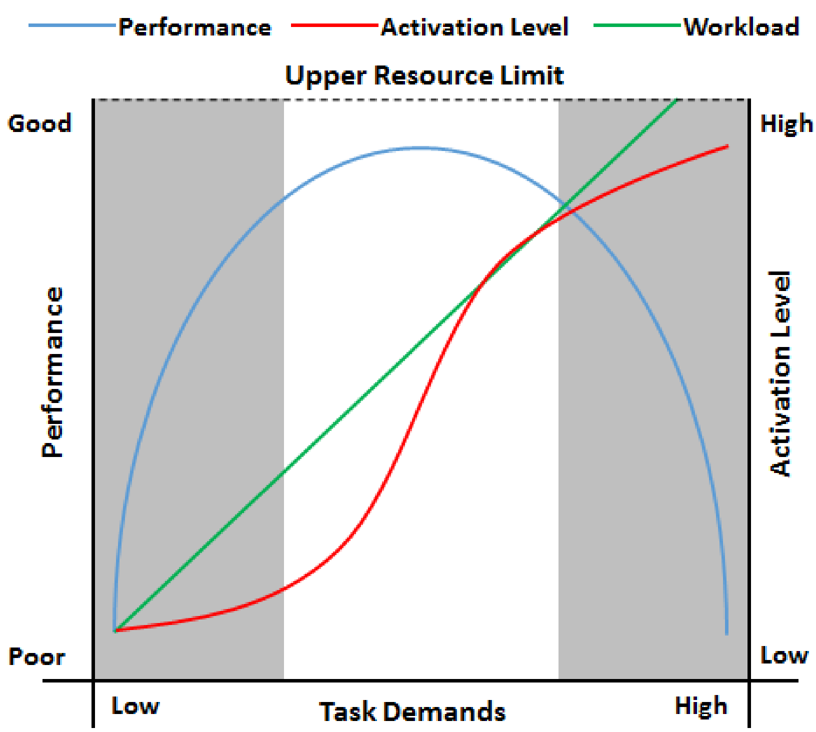


Figure 1. Relationship between Activation Level, Workload, and Performance[1].

[1] Adapted from De Waard (1996) and M. S. Young et al. (2015)

Measurement

This section provides a broad overview of common MWL assessment methods and their applications in aviation. As the focus of this thesis is ECG-based methods, more detail on these methods will be described in the sections to follow.

Design, evaluation, and optimization of aircraft systems require assessment of MWL. Though a seemingly simple endeavor, this comes with several challenges. First, because it is an abstract construct, MWL must be inferred rather than directly observed. Unlike the aircraft they fly, humans are not equipped with convenient "test points" to facilitate such measurement (Wilson, 1992). Secondly, definitive acceptable limits of MWL have not been well defined. The "redlines" indicating overload and underload continue to elude researchers to this day, such that even when MWL is quantified, it is difficult to say whether it is too much or too little (M. S. Young et al., 2015). Thirdly, the vast inventory of measurement methods generated over the past half-century varies greatly in practicality, reliability, and validity, making selection of a method in any given context difficult (Wierwille & Eggemeier, 1993). Lastly, measurements are impacted by skills, abilities, and effort put forth by the human operator. A task may induce unacceptable MWL for a novice yet fall within acceptable limits for an expert (Casner & Gore, 2010). Despite these challenges, MWL assessment is a useful tool for designers to elucidate MWL characteristics of tasks and design alternatives. MWL assessment methods fall into three general categories: subjective measures, primary and secondary task performance, and physiologic measures (Vidulich & Tsang, 2012).

The most straightforward and widely-used method is to quantify task performance. This can be accomplished by measuring performance in the primary task or measuring performance in a secondary task to infer spare capacity. Primary task performance measures generally assume that speed and accuracy of performance will decrease with increased levels of MWL (Wierwille & Eggemeier, 1993). Thus, measurements assume an acceptably low level of error and high level of efficiency (M. S. Young et al., 2015). In aviation research, primary task performance is often measured as error or accuracy with respect to flight technical parameters (e.g. airspeed, altitude, or bank angle, pitch rate, roll rate) or tactical parameters (e.g. bomb accuracy, missile tracking) (Gawron, 2008). Secondary tasks (e.g. tracking, detection, mental math) may be included with the primary task, providing more diagnosticity than the primary task alone (Cain, 2007).

Researchers have noted several limitations of primary task performance measures. As the operator compensates for increased task demands, stable performance can be maintained, reducing the sensitivity of the measure to variations in MWL (Wierwille & Eggemeier, 1993). Additionally, failures may impact the operator's perceptions of MWL (Hancock, 1989). Criticisms of the secondary task method include intrusion or interference with the primary task and variance in skill levels in performance of the primary, secondary, or combination of the two tasks (Casner & Gore, 2010).

Subjective measures gauge the operator's perception of task demands. These have the advantage of low-cost, high face-validity, and adequate sensitivity to changes in MWL levels (Wierwille & Eggemeier, 1993; Zhang, Zheng, Duan, Meng, & Zhang, 2015). The primary disadvantage is that subjective measures tend to be more widely variable based on individual differences and are subject to bias. Additionally, they may be considered intrusive, as task performance must be interrupted to measure.

The most widely used subjective method is the NASA Task-Load Index (TLX) (Hart, 2006). The NASA-TLX is a multi-dimensional scale gathering operator estimates of perceived MWL during or immediately following task performance using a set of six variables which may be weighted based on operator perceptions. A similar multidimensional method, the Subject Workload Assessment Technique (SWAT) technique, has also been commonly used in flight test settings (Reid, Potter, & Bressler, 1989). Single-dimension, decision-tree based methods constitute another commonly used class of subjective scales. Common scales include the modified Cooper-Harper scale and the Bedford Workload scale (Roscoe & Ellis, 1990).

The use of physiological variables in MWL research is based on the neurophysiological response to increase task demand, which is discussed in more detail in the following section. The principle advantage of physiologic methods is their objectivity and the ability to provide continuous, uninterrupted measurement over a given time interval (Longo, 2015). However, they can often be expensive, analytically complex, and subject to confounding environmental variables (Longo, 2015).

Cardiovascular measures, typically gathered via ECG, comprise the most commonly used physiological indices of MWL due to both validity, practicality, and ease of implementation compared to other physiological indices (Meshkati, Hancock, Rahimi, & M. Dawes, 1995; Scerbo et al., 2001). Other common methods include electroencephalogram (EEG), eye-related measures (e.g. pupil diameter, blink rate), and galvanic skin response (GSR) (Hsu, Wang, Chen, & Chen, 2015). Additionally, other measures of the brain's hemodynamic and electromagnetic activity, such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and functional near infra-red (fNIR) spectroscopy have been investigated (Ayaz et al., 2012).

As noted earlier, a considerable challenge in MWL assessment is the selection of the most appropriate method or combination of methods in a given context. Wierwille and Eggermeier (1993) outline several criteria that should be considered in assessing the value of MWL measurement techniques in flight test and evaluation: sensitivity, diagnosticity, intrusion, transferability, and implementation requirements.

Sensitivity is simply the degree to which the technique can distinguish different levels of MWL. For example, subjective measures may be sensitive in changes from low to moderate, yet insensitive from changes from moderate to high levels of MWL (Kramer, 1991). Other measures, such as performance-based, may be insensitive at low levels of MWL as the operator invests additional effort. Another important consideration noted by Kramer is the aspect of temporal sensitivity, or the ability of a measure to detect momentary changes in MWL.

Temporal sensitivity a valuable element of many physiologic and performance metrics which can be recorded continuously.

Diagnosticity is the ability of the method to identify the cause of MWL, whether is it a specific cognitive resource or specific element of the task. As it relates to MRT described above, this may refer to diagnosticity with respect to specific mental resource (e.g. visual or auditory) being utilized. It can also refer to the element of the task driving MWL levels. To this end, a measurement method with high temporal sensitivity (e.g. near real time), may also be diagnostic.

Intrusion refers to the impact of a MWL measure on performance in the primary task. In other words, an intrusive measure may produce artificially high measurements of MWL, a common criticism of, for instance, the secondary task method (Wierwille & Eggemeier, 1993). Transferability is the ability of a technique to be useful in various applications. Lastly, implementation requirements refer to the equipment, instrumentation, and data collection procedures required for a given technique. These, of course, must be appropriate for the task being evaluated.

Psychophysiology of Workload

Psychophysiology concerns the interrelationships between the physiological and psychological aspects of brain and behavior (Etzel, 2006). These relationships provide the rationale for inference of MWL through monitoring operator physiology (Wilson, 2002b). The human response to increased task demands elicits a predictable physiologic response. Activation level, or arousal, refers to a state of preparedness associated with heightened activity in the nervous system and explains the global neurophysiological response (Roscoe, 1992). Early research into the phenomenon of arousal, refined throughout the last century, suggested an "inverted-U" relationship between arousal level and performance (Yerkes & Dodson, 1908). As Roscoe (1992) notes, although the concept of arousal over simplifies complex neurophysiological mechanisms, it provides an adequate and functional explanation of the relationship between task MWL and the variance in physiologic state.

The nervous system has two basic anatomical divisions: the central nervous system (CNS) and peripheral nervous system (PNS) (Guyton, 2006). The CNS consists of the brain and spinal cord, while the PNS consists of the peripheral neural networks throughout the body (Guyton, 2006). Each of these plays a role in governing the physiologic response to MWL. Central control originates through the stimulation of the reticular activating system (RAS), which results in increased alertness, improved information processing, and shorter reaction times (Roscoe, 1992). Heightened arousal is then maintained through feedback mechanisms between the RAS, cortex, and hypothalamus (Roscoe, 1992). Central control enables MWL inferences from the EEG, MRI, TCD, and other instruments directly monitoring brain activity. The PNS provides inputs and executes actions initiated by the CNS (Wilson, 2002b). The PNS is further divided into autonomic (ANS) and somatic (SNS) branches.

The ANS is divided into sympathetic and parasympathetic branches. The sympathetic branch is responsible for what is commonly known as the "fight or flight" response, which generally results in an increase in activation level. The parasympathetic branch, or "rest and digest" mechanism, has the opposite effect. In a resting state, parasympathetic influence normally predominates. Both branches of the ANS influence cardiac function, eye activity, and electrodermal activity (Wilson, 2002b). Because of this, monitoring certain variables related to these functions allows inference to ANS activity and thus, the response to MWL variations.

Regulation of Heart Rhythms

In the absence of external influence, the heart's sinoatrial (SA) node, commonly called the "pacemaker," generates heart beats at a regular interval (Guyton, 2006). The regulation of heart rhythms, however, is a complex and dynamic process which is, in large part, a function of ANS activity. Both the sympathetic and parasympathetic branches of the ANS innervate the heart tissue (Guyton, 2006). Sympathetic modulation is accomplished through nerves arising from the spinal cord. Release of norepinephrine in the sympathetic nerve endings results in increased rate of discharge in the SA node and increased excitability and conductivity of the heart tissue. Parasympathetic modulations flow through the vagus (tenth cranial) nerve and generate essentially the opposite effect (Etzel, 2006). Release of acetylcholine from the vagal nerve endings decreases the rhythm of the SA node and decreases excitability of the heart musculature thereby slowing the transmission of cardiac impulses (Guyton, 2006). The vagus nerve provides the primary ANS control of heart rhythms.

The propagation of electrical activity during the cardiac cycle is reflected in the ECG signal (Schumacher, 2004). ECG signals have been used in clinical and psychophysiological research for years and are well understood. Each element of the ECG waveform reflects a different physical stage of the heartbeat (see Figure 2). The P wave is generated by the depolarization of the SA node and contraction of the atria. The QRS complex represents the contraction of the ventricles. The peak of the QRS complex, or the "R peak," is a common signal feature used in HRV analysis. The R-R interval denotes the distance between successive heartbeats and is usually measured in milliseconds (ms). The T wave represents the heart recovery of baseline electrical activity and the refilling of the atria.

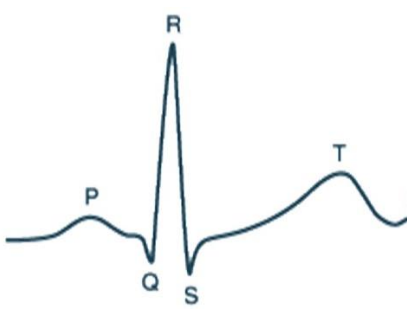


Figure 2. ECG Waveform[2].

There are numerous factors which produce constant variations in heart rhythms. To a certain extent, these processes are reflected in different frequency bands of the ECG signal,

[2] Adapted from Zhidong, Yi, and Qing (2011)

enabling spectral analysis of the signal. The two bands of most importance in psychophysiological research are the high frequency (HF) band (0.15-0.4 Hz) and low frequency (LF) band (0.05-0.15 Hz). Lower frequency components (<0.05 Hz) have been examined, but primarily in clinical applications, such as assessing risk of cardiac disease.

Variation in the HF band is known as the respiratory sinus arrhythmia (RSA) and reflects parasympathetic mediation of respiration-related activity through the vagus nerve. The relationship between ANS function and the RSA is well understood to be a reliable index of vagal control of the heart, and is the most conspicuous component of HRV (Tarvainen, Niskanen, Lipponen, Ranta-Aho, & Karjalainen, 2014). The LF band is thought to reflect a combination of sympathetic and parasympathetic influence.

While the ANS plays a predominate role in the regulation of heart rhythms, the cardiac system is influenced by other factors including thermoregulation, endocrine factors, blood pressure, and fitness level (H. Young & Benton, 2015). As such, a great deal of work in the past several decades has advanced the notion that the cardiac system, and consequently heart rhythm generation, exhibits deterministically nonlinear properties (Govindan, Narayanan, & Gopinathan, 1998; Owis, Abou-Zied, Youssef, & Kadah, 2002). This is an important consideration for HRV analysis, described in the following section. Despite the existence of nonlinear elements in the ECG signal, it has been common practice to use linear approximations, which may be unable to detect subtle nonlinear changes (H. Young & Benton, 2015).

ECG Methods for Workload Assessment

The two primary MWL measures derived from the ECG are heart rate (HR) and heart rate variability (HRV). HR is the simplest method and generally increases with increases in MWL. HRV is defined as the variability of normal-normal (NN) intervals, or variation between heart beats. As MWL increases, HRV generally decreases (Roscoe, 1992).

HRV analysis has been used to develop a wide range of metrics in both clinical and psychophysiological applications (Task Force, 1996). These methods generally fall into three categories: time domain, frequency domain, and nonlinear methods. Time and frequency domain methods have been more commonly used in MWL measurement and constitute what is herein referred to as "classical methods." These are summarized in Table 1.

Table 1. Time and Frequency Domain HRV Metrics*.

Domain	Metric	Units	Definition
Time	SDNN	Milliseconds (ms)	Standard deviation of all NN intervals
	RMSDD	Milliseconds (ms)	Root mean square of differences between adjacent intervals
	pNN50	Percentage or proportion	NN50 count divided by the total number of NN intervals
	HRV TI	No unit	HRV triangular index
	TINN	Milliseconds (ms)	Triangular interpolation of the NN histogram
Frequency	LF	Normalized units (nu)	Power in the LF range (0.04-0.15 Hz)
	HF	Normalized units (nu)	Power in the HF range (0.15-0.4 Hz)
	LF/HF	No unit	Ratio LF / HF

*Note: metrics presented in this table are limited to those included in this analysis.

More contemporary metrics are based on the nonlinear elements of the ECG signal. Applications of these metrics have become commonplace in clinical HRV analysis to diagnose disease. However, nonlinear metrics have not been fully explored in applications involving normal healthy individuals, such as in MWL research (H. Young & Benton, 2015).

This section reviews common time and frequency domain measures of HRV. A short summary of nonlinear methods is then provided. The discussion of nonlinear methods, however, focuses on one method that is the subject of the research question in this thesis.

Time Domain

The mean value of HR or mean RR interval (the two are analytically equivalent) are the simplest metrics derived from the time series of RR intervals. These metrics can either be analyzed via comparison of means or relative change from a baseline resting state. Mean HR is widely used but often criticized for its sensitivity to confounding factors such as respiration, physical exertion, and stimulants such as caffeine and nicotine (Roscoe, 1992).

Lee et. al. measured incremental HR (difference in HR between resting and working states) in a Boeing 747 flight simulation study and concluded both HR and HRV to be consistent with NASA-TLX indices, concluding both variables to be sufficient predictors of MWL (Lee & Liu, 2003). A similar study by Zhoe et. al. more recently produced similar findings comparing HR and HRV to NASA-TLX (Zhou, He, Wang, & Fu, 2014). Some question the consistency of HR itself, however, due to its sensitivity to multiple factors such as physical exertion (Scerbo et al., 2001). This presents limitations in aviation experiments which contain components of physical task demand. However, HR measurement has proven reliable researching less physically demanding aviation-related tasks, such as Air Traffic Control (ATC) and Air Battle Management (ABM) (Strang, Best, & Funke, 2014; Vogt, Hagemann, & Kastner, 2006). Sensitivity of HR along with HRV was evaluated in a another simulator study of pilots completing instrument landing system (ILS) approaches (Mansikka, Simola, Virtanen, Harris, & Oksama, 2016). The researchers concluded that both HR and HRV comparably sensitive to changes in MWL from baseline to high-MWL conditions.

Time domain HRV metrics are derived from basic statistical analysis of the NN intervals in time series (Task Force, 1996). Common metrics include the mean RR interval (or mean NN), the standard deviation of NN intervals (SDNN), the square root of the squared successive NN intervals (RMSSD), the number of successive differences in NN intervals differing by greater than 50 ms (NN50), and the percentage of the NN50 relative to all of the NN intervals in a given timeframe (pNN50). Geometric indices may also be derived from the times series of NN intervals. Two commonly used geometric indices include the HRV triangular index (HRV TI) and the triangular interpolation of the NN histogram (TINN).

SDNN is a means to compute the overall variability of the RR series and is considered an appropriate measure of both short and long-term variability (Task Force, 1996). Lower SDNN values indicate lower variability and higher MWL. SDNN is calculated with the below equation.

The mean RR interval length is depicted by $\overline{RR}$ and the RR interval for each sample (j) in the recording is depicted by RR_j. The total number of RR intervals in sample is depicted by N.

$$SDNN = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(RR_j - \overline{RR})^2}$$

RMSSD is another common measure considered an appropriate measure of short term variability. Likewise, lower RMSSD values indicate lower variability and higher MWL. It is calculated by the below equation where $RR_{j+1} - \mathrm{RR_j}$ denotes the difference between the length of j^{th} RR interval and the successive interval.

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}(RR_{j+1} - \mathrm{RR_j})^2}$$

PNN50 is simply a measure of the percentage of successive NN intervals differing by greater than 50 ms over a given recording period. A 50 ms calculation is the most common interval used in this calculation, but other intervals have been used in different applications (e.g. 20 ms). Like the previous two metrics, lower PNN50 indicates lower variability. The calculation is as follows, where NN50 is the total number of successive differences differing by greater than 50 ms and N is the total number of RR intervals in the time series:

$$pNN50 = \frac{\mathrm{NN50}}{\mathrm{N-1}}\,\mathrm{x}\,100\%$$

The time series of RR intervals can also be analyzed geometrically. There are two common geometric methods. The first, the HRV Triangular Index (HRV TI), measures the integral of the density distribution of NN intervals (i.e. the total number of NN intervals) divided by the maximum of the distribution. Height of the histogram, naturally, is affected by size of the bins when generating the histogram and therefore must be consistently selected when comparing time intervals.

$$HRV\ TI = \frac{Total\ number\ of\ NN\ intervals}{Number\ of\ NN\ intervals\ in\ modal\ bin}$$

The second method is the triangular interpolation of the NN histogram (TINN) which is similar to the HRV TI. To calculate TINN, the baseline width of the NN interval distribution is measured as the base of a triangle, which is then used to interpolate (i.e. approximate) the distribution. The unit for the TINN is ms.

Both statistical and geometric methods for HRV in the time domain are analytically simple and well-validated measures. The primary drawback is their sensitivity to overall recording length. In most cases, they are not well suited for short duration recordings (i.e. must have a larger number of NN intervals) (Task Force, 1996). Further, they are not well suited for comparing recordings of unequal length (Etzel, 2006). Geometric methods are resistant to sample quality issues but are sensitive to differences in parameter selection in calculation (e.g. bin width of the NN histogram) (Task Force, 1996).

Frequency Domain

Frequency analysis is conducted by decomposing the ECG into sinusoidal waves to ascertain its frequency components (Schumacher, 2004). Common techniques include Fast Fourier Transforms (FFT), periodogram, and autoregressive (AR) modeling (Schumacher, 2004). Newer methods allow for accurate time varying analysis using moving window techniques, such as the Kalman smoothing method proposed by Tarvainen, Georgiadis, Ranta-aho, and Karjalainen (2006). Spectral decomposition allows for distinction between relative sympathetic and parasympathetic modulations of heart rhythm by measuring the amplitudes of the ECG interval signal at different frequencies (Lean & Shan, 2012).

As discussed in the previous section, the LF and HF components are of most interest for MWL applications. While both bands appear to show modest sensitivity, researchers have produced mixed conclusions as to which band is more diagnostic (Scerbo et al., 2001). The HF component (0.15-0.4 Hz), or RSA, is dominated primarily by parasympathetic modulations and is associated with mechanical and reflex respiratory activity components and is usually interpreted as reflecting oscillations caused mainly by changes in vagal tone of heart rhythm. The LF component (0.04-0.15 Hz) reflects both sympathetic and parasympathetic modulations (Berntson et al., 1997). The LF/HF ratio is useful as it can be interpreted as a reflection of sympathovagal balance or sympathetic modulation (Tarvainen et al., 2006). Spectral power in the respective bands may be presented as a percentage of total power or normalized units.

Generally speaking, frequency domain methods are preferred over time domain methods for short-duration recordings (Task Force, 1996). However, it is suggested that that the recording length should last at least ten times the length of the lower band of the frequency band being analyzed. This equates to approximately 1 minute for the HF component and 2 minutes for the LF component (Kuo & Chen, 1998).

Nonlinear Domain

Nonlinear analysis of HRV is based on the notion that the cardiovascular system, particularly with respect to heart rhythm generation, is complex, dynamic, nonlinear, and nonstationary (Schumacher, 2004). Therefore, analysis purely in the linear domain may not capture all the underlying patterns exhibited in the ECG signal. The introduction of nonlinear dynamics (NLD) into HRV analysis dates back to the 1980's (Goldberger & West, 1987). Since then, a considerable amount of work has been done to explore the nonlinear properties of heart rhythms (Voss et al., 2009). Common methods include Poincaré plots, approximate entropy (ApEn), sample entropy (SampEn), correlation dimension, detrended fluctuation analysis (DFA), and recurrence plot (RP) analysis (Tarvainen et al., 2014). To date, applications of these methods have become commonplace in clinical applications of HRV analysis (Francesco et al., 2012), but have been less prominent psychophysiological research (H. Young & Benton, 2015).

However, a small number of studies have employed nonlinear HRV analysis in the domain of psychophysiological research. Sammer (1998) demonstrated that nonlinear properties of HRV may be useful in discriminating mental and physical load. More recently, Bornas et al.

(2013) conducted a study to evaluate attentional control (AC) in healthy students using correlation dimension (CD), fractal-like properties, and sample entropy. The authors found positive correlations with AC and several of these properties, concluding they may be useful in psychophysiological applications.

Schneider et al. (2017) conducted a study to assess MWL of emergency room physicians using 18 different linear and nonlinear measures of HRV. Their results indicated permutation entropy outperformed all linear HRV metrics in terms of receiver operating characteristics curve (AUC), leading them to conclude nonlinear metrics provide sensitive and valid measures of MWL. H. Young and Benton (2015) conducted a study relating linear nonlinear indices to cognition and mood, producing similar conclusions. In the aviation domain, one study compared linear HRV metrics to Fuzzy Approximate Entropy (fApEn), noting correlations between the two domains (Strang et al., 2014). Sauvet et al. (2009) utilized linear HRV and Poincaré plots to analyze pilot MWL in a multileg cross-country flight, also producing comparable results. These initial studies show promising validity of nonlinear methods; however, further research is needed to fully evaluate their utility.

The focus of the analytical portion of this thesis is on a novel nonlinear method developed by the University of Iowa OPL. This method is based on transition probability variances (TPV) of ergodic transition matrices (ETM) derived from the ECG signal. The method is based on work by J. J. Engler (2011). It has been used in multiple previous studies and has shown promise in detecting MWL changes with high temporal sensitivity.

To estimate MWL based on TPV, there are several steps. The raw ECG signal is first transformed into multidimensional embedded phase space. This type of transformation is based on the embedding theorem first conceived by Takens (1981) and there is precedent for its application to the ECG signal (Richter & Schreiber, 1998). Specifically, the scalar time series is transformed into a vector using a time-delay parameter (τ) and embedding dimension (n). A mutual information function is used to calculate the time delay parameter and the embedding dimension is calculated using a false nearest neighbor technique. An example of ECG and its embedding phase space are shown in Figure 3. In this example, the time delay (τ) is 12 seconds. This embedding phase space is extracted from the Cognitive Assessment Toolset (CATS) architecture (discussed in Chapter 3).

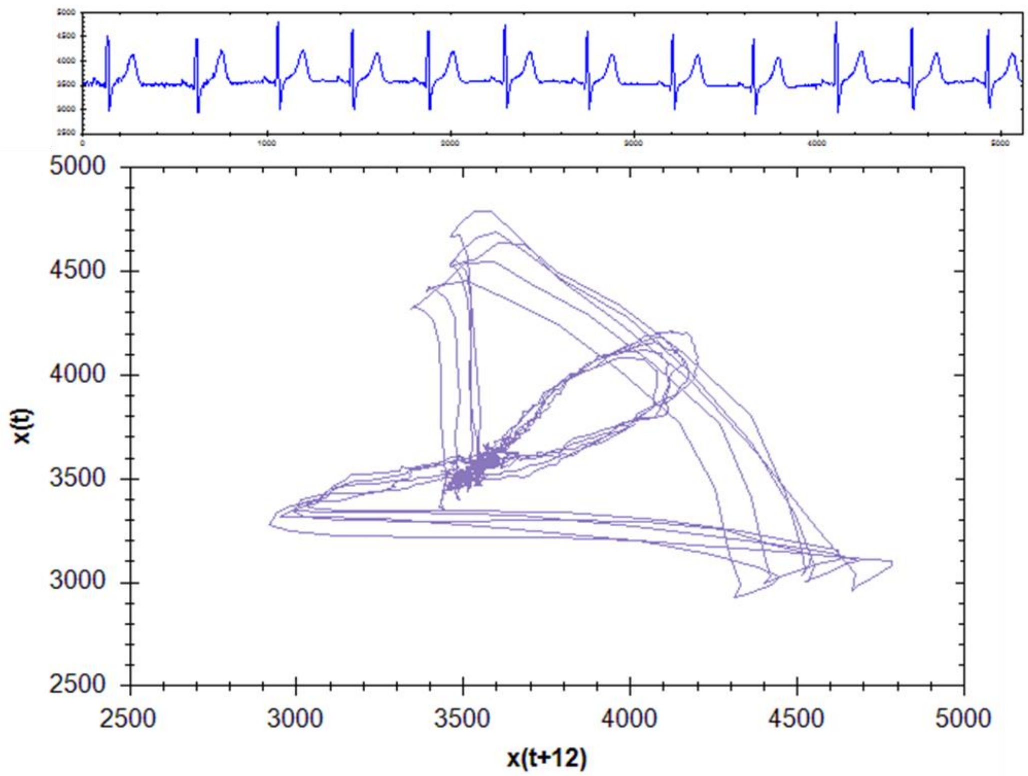


Figure 3. Embedding Phase Space for ECG Signal (extracted from CATS).

Complex chaotic systems can be represented in more simplified manner with a technique called coarse graining applied to partition the embedding phase space, creating discrete states. Thus, the *n*-dimensional hypercube contains a finite number of bins, or cells. A simplified coarse graining of the embedding phase shown in Figure 3 is shown in Figure 4. From this, a 2-dimesional matrix of transitional probabilities is created, which is referred to as the ETM. Variance within the ETM can be summarized with a single metric, the transition probability variance (TPV). This is typically calculated as an average value over a buffer of samples (e.g. 2

seconds, or 1024 samples at 512 Hz). The inverse of this TPV is thought to vary proportionally with changes in MWL.

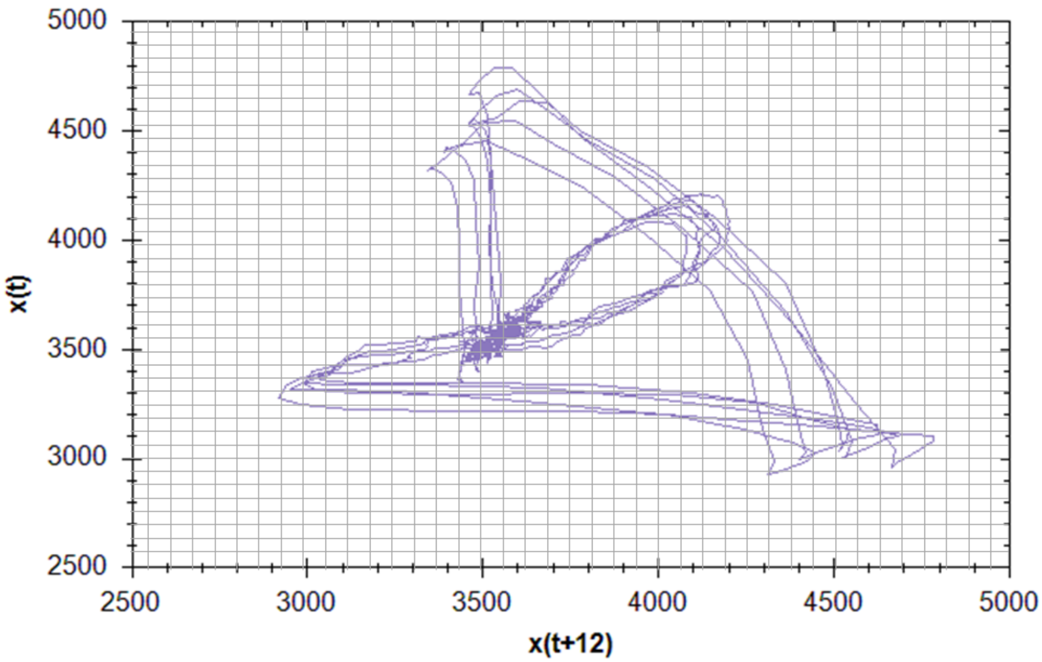


Figure 4. Coarse Grained (40 x 40) Representation of the Phase Space in Figure 3.

In the OPL implementation of this method, which is used in the HMD study described in Chapter 3, TPV is used to calculate a relative MWL value (hereafter referred to as TPV relative workload). This is accomplished by calculating baseline values during a resting state and maximum values recorded during the recording of interest. Between these bounds, an arbitrary scale is created (e.g. 0-10) to indicate the relative change in MWL.

A significant apparent advantage of the TPV method over classical time and frequency domain methods is its ability to measure MWL with very high temporal sensitivity. As values can be generated on the order of a few seconds, this method can detect momentary, sharp fluctuations in MWL that the literature would indicate classical time and frequency domain methods cannot. This is important in certain applications such as the research study presented in

Chapter 3. Despite its promise, the TPV method has never been compared empirically to classical time and frequency domain methods on the same dataset. This research aims to address this question.

CHAPTER 3 – SUMMARY OF HMD STUDY

A study was conducted to compare three off-boresight (OBS) helmet-mounted display (HMD) symbology sets (Schnell, Reichlen, & Reuter, 2017). This chapter provides a summary of the background, objectives, methods, key results, and conclusions. The primary results from this study have been previously reported. The workload data collected served as the basis for the analysis presented in Chapter 4 – Chapter 6 of this thesis.

Background

This study was motivated by the fact that throughout history the technological landscape of a fighter airframe changes dramatically during its life cycle to continually provide both tactical and safety of flight improvements. In some cases, new technologies, intended primarily to bring tactical benefits, introduce new and unforeseen risks. One such risk in fighter-type aircraft is spatial disorientation (SDO). SDO has been a costly, deadly problem across all domains of aviation, factoring in as many as one third of serious mishaps and having a nearly 100% fatality rate (Gibb, Ercoline, & Scharff, 2011). Further, it has affected the fighter community at a disproportionately high rate (Lyons, Ercoline, O'Toole, & Grayson, 2006).

This study contributed to the development and of Helmet-Mounted Display (HMD) symbologies displaying ownship status information in the OBS regime. The OBS regime is defined as the portion of the pilot's visual field of regard which is outside of the traditional Heads-Up Display (HUD), or in the case of modern HMD systems, a virtually rendered HUD (vHUD) (see Figure 5). Attitude information is especially important for SDO prevention. In the absence of reliable visual references in the external environment, attitude symbologies constitute the single means by which a pilot may orient in relation to the Earth's surface.

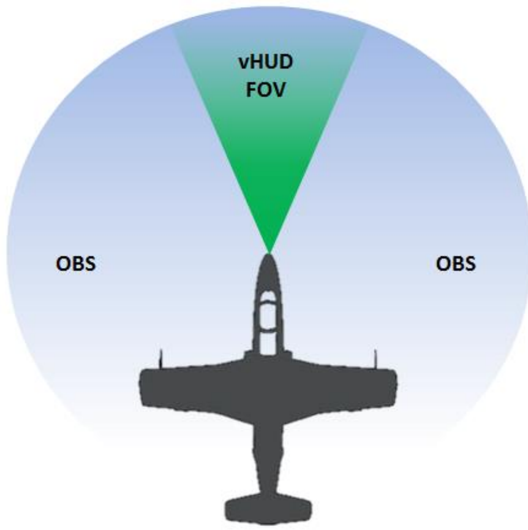


Figure 5. Off-Boresight vs. vHUD FOV.

Previous research has shown that displaying information in the OBS regime changes pilots' scanning behaviors; pilots tend to look OBS further and for longer durations (Geiselman, 1999). Risks associated with these behavior changes in relation to SDO have not been well characterized for modern HMD systems. Further, this study was partially motivated by the fact that certain HMD systems currently in use do not contain attitude symbology in the OBS regime. The objective was to characterize this risk as a function of symbology type during a live-flight operational task that elicited OBS visual requirements. The scenarios developed were intended to represent operationally demanding tasks anticipated to present SDO risk in 5th generation fighter aircraft.

Methods

Participants

In the original design of experiment, a total of 10 subject pilots were sought. However, due to unforeseen circumstances, 11 subjects participated in total. This was because one subject could only complete sortie, so an additional subject was recruited to complete two additional sorties. All subjects were current and qualified military pilots from a variety of airframes, to include the A-10, B-1, C-130, F-16, FA-18, F-22, and F-35. Preference was given to pilots with air-to-ground (A/G) tactical experience to facilitate execution of the experimental scenarios. Additionally, subjects had mixed previous experience with operational HMD systems, such as the Joint Helmet Mounted Cueing System (JHMCS), Scorpion HMD, and F-35 HMD. The original design intended to balance the subject pool by previous HMD experience, having 5 subjects with previous HMD experience and 5 without. Due to recruitment and timeline constraints, this balance was not achievable. Of the 11 subjects, 3 had no previous HMD experience and 8 had experience with at least one of the HMD variants listed above.

Apparatus

The experimental aircraft was an Aero Vodochody L-29 Delphin operated by the OPL. It is a single engine, 2-seat jet trainer aircraft capable of high performance maneuvering representative of fighter aircraft. The cockpit is highly instrumented with state of the art avionics, simulated weapons, and simulated radar systems. Additionally, the aircraft is outfit with human performance assessment equipment to allow collection of flight control inputs, physiologic data for MWL assessment, head-tracking data, and 4-channel video and audio recording. The L-29 is an aircraft-in-the-loop (AIL) simulator in addition to a live-flight experimental aircraft. This capability allowed subjects to familiarize with the aircraft, HMD, and cockpit avionics in the experimental environment on the ground prior to the live-flight scenarios.

During live flight, the Safety Pilot (SP) flew in the front cockpit, serving as the pilot-in-command (PIC), and was ultimately responsible for safe flight operations. The SP controlled the aircraft during all non-experimental phases of flight (engine start, taxi, takeoff, navigation to test area, and landing) and handled all coordination with air traffic control (ATC). The experimental pilot (i.e. subject) flew in the rear cockpit and took control of the aircraft during the experimental scenarios. The aircraft was data-linked to a ground station at the OPL enabling communication with the experimental payload controller (EPC). The EPC coordinated execution of the experiment and managed data collection from the ground-based station.

Subjects were equipped with a fifth-generation fighter representative HMD, integrated into the L-29, with magnetic head tracked graphics processor (Figure 6). To simulate a nighttime environment, a canopy view limiting device (CLVD) and forward view limiting device (FVLD) were installed in the rear cockpit to block the entrance of ambient light. The simulated environment was projected onto the HMD Organic Light Emitting Diode (OLED) imagers in monochrome green to mimic the visual experience of using a night vision device such as the F-35 Distributed Aperture System (DAS). The HMD combiner presented imagery as a biocular, overlapped image at 1280 x 1024 pixels on a 30 x 40 degrees field of view.

Figure 6. Representative 5th Generation HMD Used in Study.

The L-29 rear cockpit contains a 15-inch head-down display (HDD) with which the subject pilot could control the simulated weapons systems by selecting the type and quantity of ordinance. The HDD also provided additional situation awareness information including an electronic attitude-direction indicator (ADI) and horizontal situation display (HSD). The standard HDD configuration for this study, with the weapons selection page on the left and HSD on the right, is shown in Figure 7. In addition to touch screen control, the pilot could also manipulate these displays via the hands-on throttle and stick (HOTAS) controls integrated into the rear cockpit.

Figure 7. Head-Down Display (HDD)

Real-world flying operations were conducted in local Iowa airspace between approximately 10,000-17,000 ft Mean Sea Level (MSL). Using the upfront control panel in the rear cockpit, the subject pilot would lateral proxy the simulated environment to a mountainous location in Afghanistan. Depending on real-world airspace constraints, the subject would typically vertical proxy approximately 5,000 down, meaning simulated altitude would be lower than real-world altitude, to facilitate the experimental scenarios.

Figure 8 and Figure 9 depict the physiological data collection set-up. ECG was collected using the Nexus-4 physiological monitoring system connected to the L-29 onboard computer via Bluetooth (Figure 8). ECG electrodes were placed in a three-lead configuration worn under a standard Nomex flight suit and cotton t-shirt. ECG data was sampled at 512 Hz. The Nexus monitor was stowed in the subject pilot's flight suit pocket during flight (Figure 9).

Figure 8. ECG Sensor System and Lead Placement.

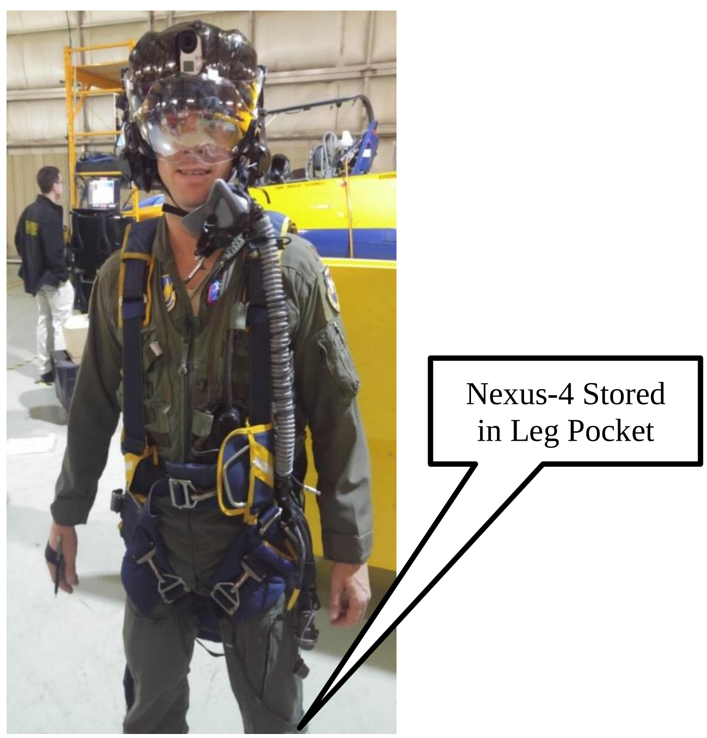


Figure 9. Subject Helmet and Sensor Configuration.

Data was collecting using the OPL-developed Cognitive Assessment Toolset (CATS). The general flow of information is shown in Figure 10. Aircraft state and physiologic data were collected and stored in a time-stamped relational database. In this study, the CATS architecture facilitated data collection, synchronization, and analysis.

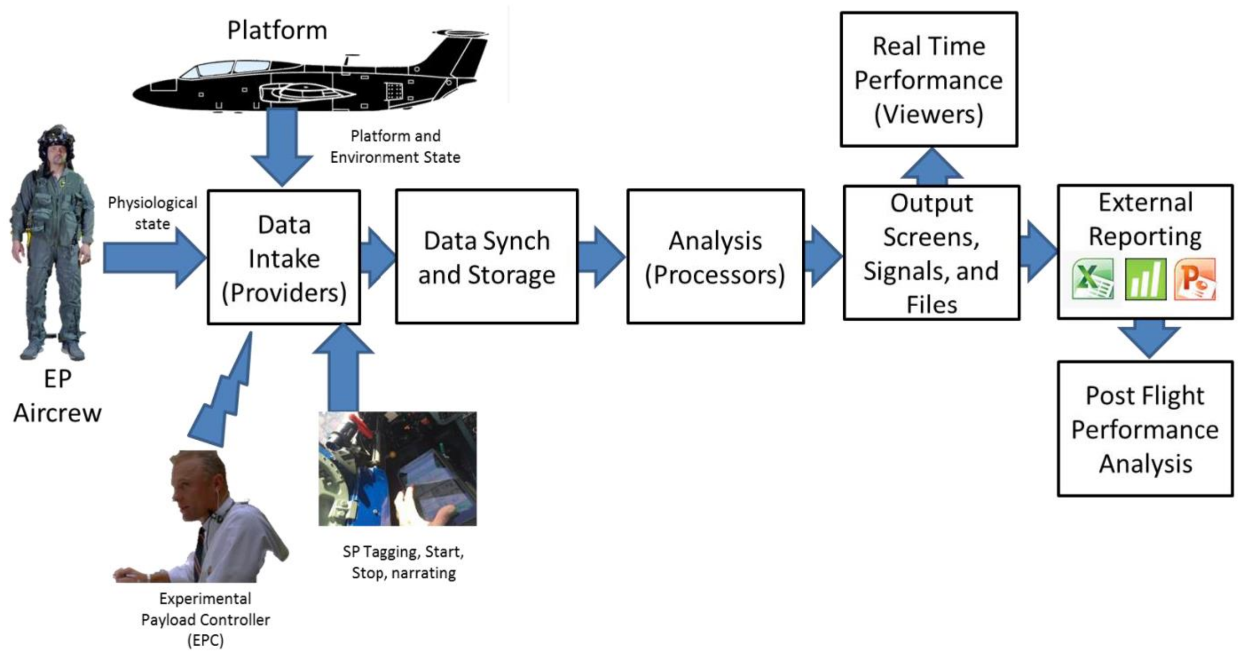


Figure 10. High Level Data Flow Diagram of CATS System on L-29.

<u>Test Symbologies</u>

This study examined three candidate OBS symbology sets, which served as the independent variables in the experiment:

1) Current Display Format (CDF)
2) Distributed Flight Path Reference (DFR)
3) Non-Distributed Flight Path Reference (NDFR)

These symbologies appeared in the HMD OBS regime, i.e. while the subject looked more than 15 degrees laterally and/or 25 degrees vertically from the aircraft center line (boresight cross). The principle difference between the three conditions was the display ownship information, specifically aircraft attitude, altitude, airspeed, and heading. The DFR and NDFR conditions contained an additional attitude symbol referred to as an Arc Segmented Attitude Reference (ASAR). The ASAR was comprised of a fixed aircraft symbol and a dynamic earth reference symbol that would rotate in response to roll and grow or shrink in response to pitch changes. Two tick marks on the ends of the aircraft symbol depicted the nearest horizon. During straight and level flight, the ASAR would be a semicircle below the aircraft symbol with the ends touching the tick marks. For a full description of the symbology, see Geiselman (1999). The ASAR was displayed in a fixed position in the upper right corner of the pilot's OBS field of view in both the DFR and NDFR conditions. On the aircraft centerline axis, a virtual Heads-Up Display (vHUD) was displayed (Figure 11) in all three conditions.

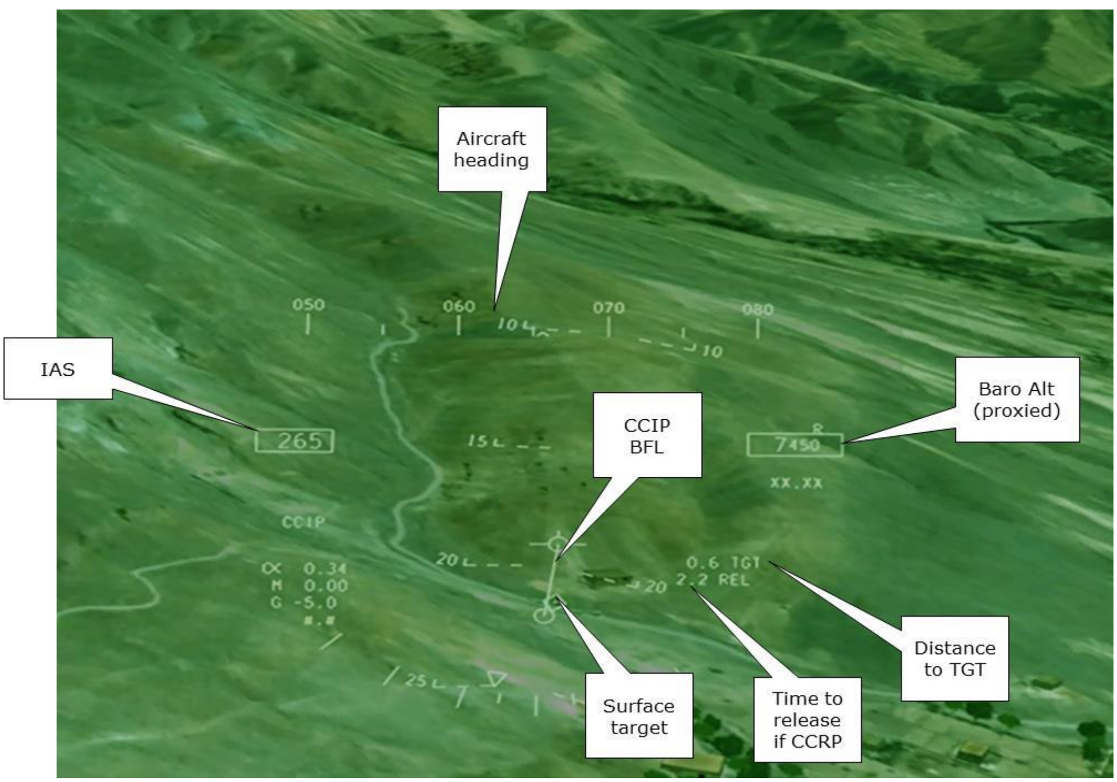


Figure 11. Monochrome Green Depiction of Virtual Heads-Up Display (vHUD).

The CDF, shown in Figure 12, was essentially the control condition. This symbology set is representative of what is currently in-use in certain fifth generation fighter HMD systems. Head-heading, airspeed, and altitude are displayed along with an aiming reticle which represents the center of the pilot's field of view. Of note, there is no attitude reference.

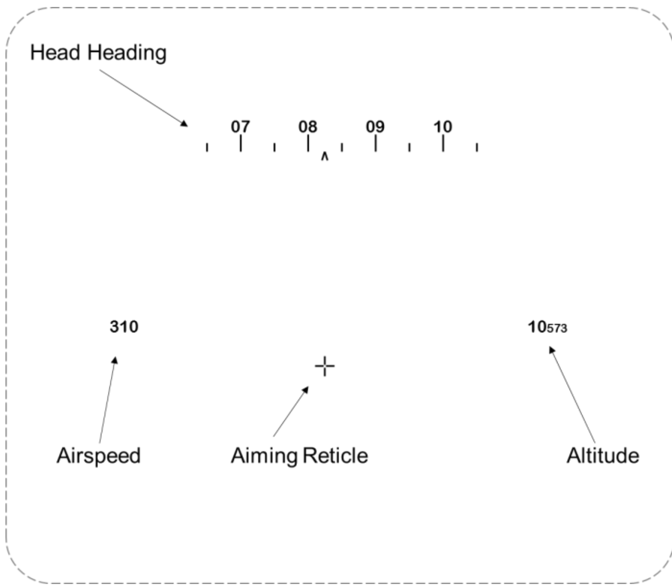


Figure 12. Current Display Format (CDF) Symbology.

Figure 13 shows the DFR. In this condition, the ASAR and fixed aircraft symbol are displayed in the upper right corner of the display FOV. Airspeed and altitude are "distributed" around the center of the FOV. Aircraft heading is not included.

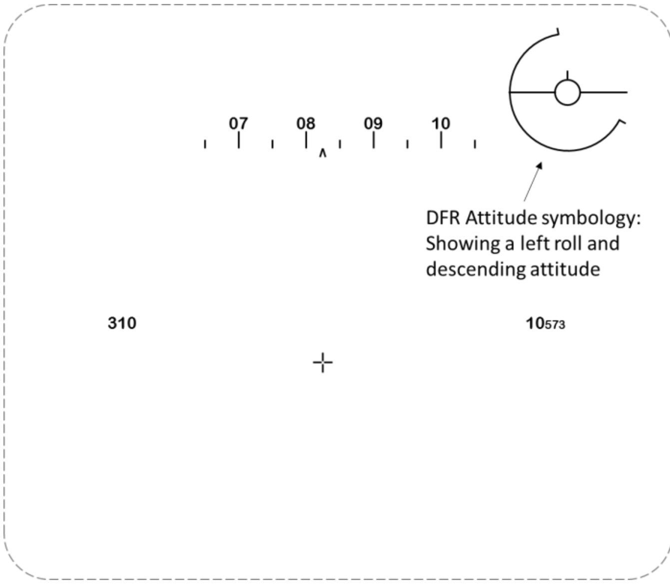


Figure 13. Distributed Flight Path Reference (DFR) Symbology.

The NDFR is depicted in Figure 14. Unlike the DFR, airspeed and altitude are superimposed on the ASAR along with aircraft heading.

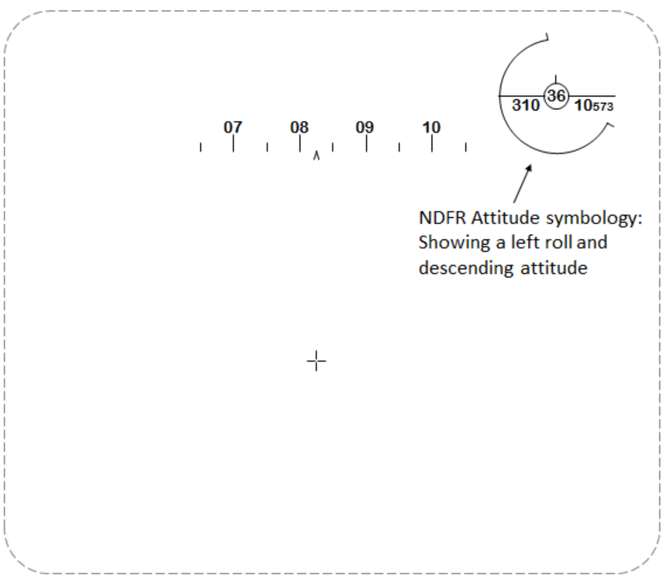


Figure 14. Non-Distributed Flight Path Reference (NDFR) Symbology.

Experimental Scenarios

There were a total of three scenarios which occurred on three separate flights. All scenarios were designed with the basic intent of simulating operationally realistic mission tasks, specifically those which elicited OBS head movements. The three scenarios were set in a simulated nighttime environment and accordingly labeled N1, N2, and N3. The selected mission was Close Air Support (CAS) in which the aircraft was under the operational control of a ground-based Joint Terminal Air Controller (JTAC) who directed weapons delivery to ground targets. The scenarios were designed to align with Joint Publication (JPUB) 3-09.3, *Close Air Support,* and were constructed with the assistance of fighter subject matter experts (SME). The EPC served as the JTAC for all scenarios and utilized a script developed by fighter CAS SMEs. Recent operational experience in fighter CAS was considered in the selection of subject pilots to alleviate the need to provide additional training.

All scenarios were Type II control, bomb on target (BOT) attacks. This differs from a bomb on coordinate (BOC) type attack in which a weapon would be delivered to a preprogrammed location with the aid of GPS. In a BOT attack, the JTAC "talks on" the pilot visually to the target by issuing a series of instructions to reference objects or features in the target area (e.g. buildings, roads, rivers, etc.) while the aircraft loiters within visual range of the target area. The pilot is eventually funneled to the target and confirms he/she has visually acquired it with the JTAC. Both the JTAC and pilot are often aided with common printed maps or products (i.e. "placemats") based on gathered intelligence of the target area. In this experiment, the pilot and JTAC both had available the map shown in Figure 15. Note that the callout balloons for numbered building, building group, and echo point were not included in the actual experimental map. This map was referred to as "Placemat Alpha 01" during the experimental scenarios.

BOT attacks were specifically chosen over BOC attacks to necessitate OBS head movements since BOC attacks would primarily require interaction with displays and sensors up front in the cockpit. This required the pilot to perform the search task and maintain control of the aircraft while looking down and to the side of the aircraft. In the mission brief, the subject pilots were instructed to execute “left-hand” patterns during the talk-on phase. Thus, all talk-ons occurred with the pilot searching for the target off the left side of the aircraft. This standardization proved to be useful in the interpretation of flight technical data.

The N1, N2, and N3 scenarios were designed to have progressively higher complexity. In the N1 scenario, there was only one target per symbology. In addition, the JTAC issued a “floor” altitude restriction of 9,000 ft Mean Sea Level (MSL). In the N2 scenario, the pilot was given an altitude block of 9,000-11,000 ft MSL and had to perform an additional show of force (SOF) in addition to dropping simulated ordinance for each symbology. The N3 scenario had same restrictions as N2 but introduced clouds, decreasing the presence of visual cues.

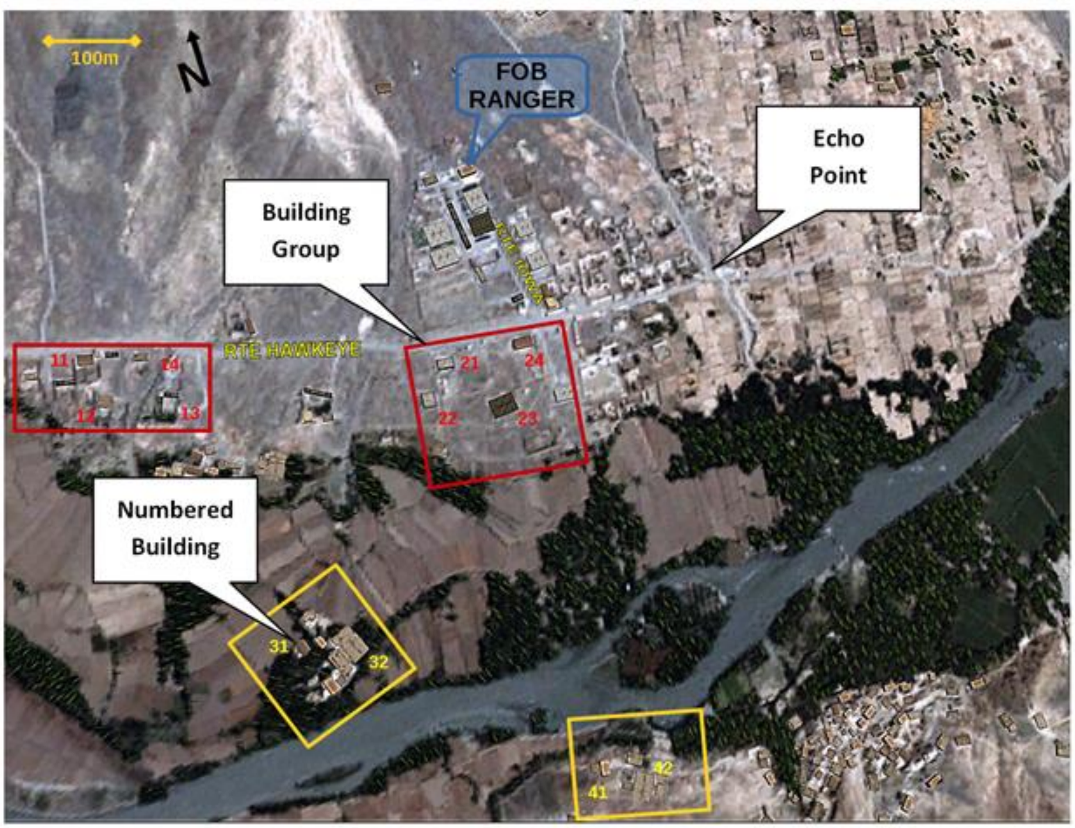


Figure 15. Local Map (Placemat Alpha 01) Used by JTAC and Subject.

Data Collection Procedures

Upon arrival at the OPL, subjects were given a familiarization brief that had been emailed to them prior for review. The briefing included an overview of the study and test symbologies, local area flying procedures, cockpit and equipment familiarization, and emergency egress procedures. Subjects were equipped with ECG sensors and then given a familiarization simulator session in the L-29 (on ground in sim mode). As this was for familiarization only, no data was collected in simulator mode.

Analysis Methodology

During the execution of the experiment and initial phases of analysis, it became clear that there were significant effects of sortie (N1, N2, N3) on flight technical and MWL performance metrics that were not consistent with the intentions of the experiment. The design of experiment (DOE) intended for the first sortie (N1) to be the easiest and for the third sortie (N3) to be the most difficult. However, the opposite effect was observed. This was likely due to learning effects, such as progressively greater familiarity with the flight characteristics of the L-29, onboard weapons systems, and the simulated operational environment (i.e. target area). It is possible the overall experience level of the subject pilots contributed to this learning effect occurring more quickly than expected. As a result, effects on dependent variables were essentially washed out on the second and third sorties. For these reasons, the analysis focused solely on the first sortie (N1). Given the fact that CAS missions must often be executed in unfamiliar environments in real-world combat, the conclusions gleaned from analysis of this sortie have important practical implications.

Within the N1 sortie, the analysis was further narrowed to the talk-on portion of each attack, despite the fact that the situation report (SITREP) and 9-Line (attack order) exchanges as well as weapons delivery phases were initially examined. The rationale for this was two-fold. First, SITREPs and 9-Lines were excluded because only one was issued per subject on the N1 sortie. This resulted in an imbalance and insufficient data for comparisons between symbologies. Second, initial analysis revealed that weapons delivery was primarily accomplished while the pilot was looking through vHUD and not OBS. Therefore, weapons delivery metrics provided little insight into performance effects of the OBS symbologies.

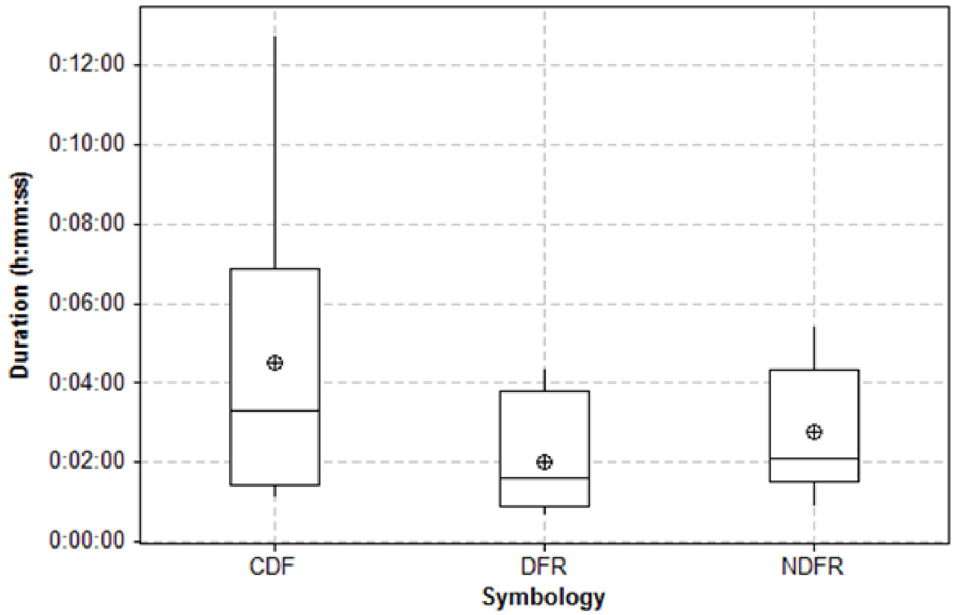
Due to technical issues during the execution of the experiment, a small amount of data was lost. For the N1 sortie specifically, aircraft state data (flight technical) was lost for Subject 2 and ECG data was lost for Subject 6. Because of this, despite having 11 subjects fly the N1 sortie, for results relating to flight technical performance and ECG-based workload, there were 10 subjects.

Results and Discussion

This section provides an overview of the most significant findings from the study. Refer to the full report for additional metrics not summarized below. These were selected for their relation to MWL analysis in this thesis. The following results are presented and discussed: task duration, flight technical (pitch and roll rates), and workload. For the full analysis refer to Schnell, Reichlen, et al. (2017).

Task duration

Time required to complete the talk-on phase was compared between symbologies. As Figure 16 shows, the NDFR and DFR facilitated a shorter duration talk-on than the CDF. GLM ANOVA showed a significant effect of symbology on talk-on duration ($F_{2,20}$=4.64, p=0.022). A post-hoc Tukey t-test indicated a significant difference (shorter time) for the DFR compared to the CDF (t=-2.965, p=0.0201) but not the NDFR (t=-2.091, p=0.1170) average talk-on durations. This result was of great practical significance. It demonstrated the tactical advantage of the symbology by showing a difference in how efficiently the pilot was able to acquire the target as a function of symbology, and therefore execute the attack in a shorter period of time. This was an important result with respect to the MWL analysis, as described below. Despite the DFR condition eliciting the highest MWL, it facilitated the most efficient execution of the task.




```
Source    DF  Seq SS   Adj SS  Adj MS    F    P
SubjectId  9 14.4423 14.4423 1.6047 2.54 0.044
Symbology  2  8.5698  8.5698 4.2849 6.79 0.006
Error     18 11.3639 11.3639 0.6313
Total     29 34.3760
Symbology = CDF subtracted from:
          Difference     SE of       Adjusted
Symbology   of Means  Difference T-Value  P-Value
DFR         -1.277    0.3553   -3.594   0.0056
NDFR        -0.888    0.3553   -2.500   0.0555
```

Figure 16. Mean Talk-On Duration for N1 Sortie by Symbology.

Flight Technical

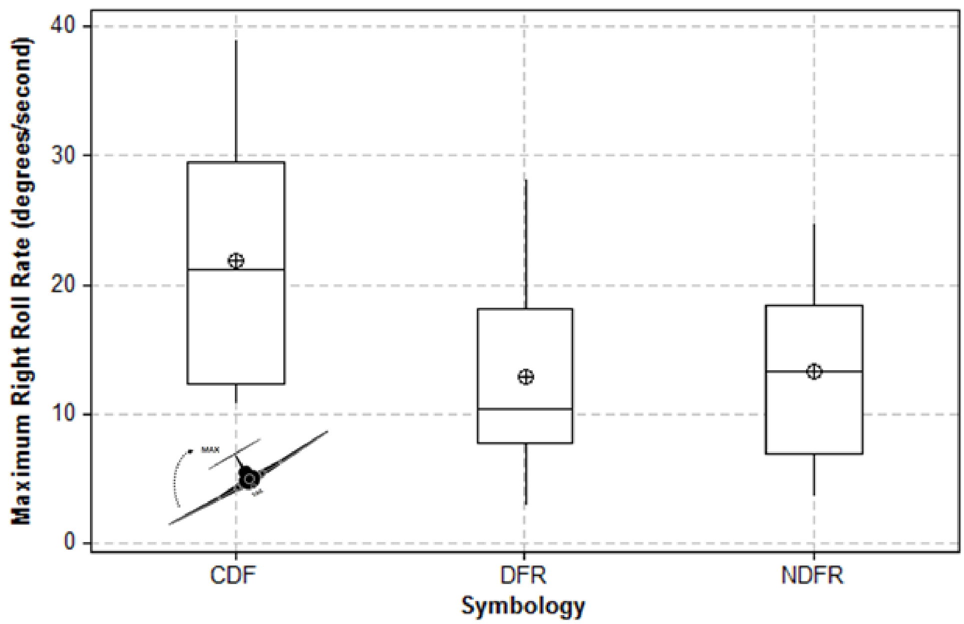
Flight state data from the L-29 was harvested to analyze flight technical performance. In this applied experimental setup, strict flight technical parameters were not dictated. The pilots were simply asked to comply with JTAC instructions and execute tasks as if they were in a real combat environment. There were altitude blocks assigned by the JTAC, but deviations were rarely observed. This fact made analysis of flight technical performance more complicated since as simple flight technical error could not be calculated. Given the overall aim of this study was SDO prevention, the analysis sought to derive precursors of unusual attitudes. As such, the exploration of aircraft state parameters focused on the pilot's ability to maintain stability of the aircraft throughout the talk-on. Parameters explored included aircraft attitude measures (bank angle, pitch angle, flight path angle) and their derivatives (roll rate, pitch rate).

Roll and Pitch Rates

Roll rate was used to infer the intensity with which the subject pilots made lateral side-stick inputs. The focus remained on the talk-on phase of the N1 sorties. The subjects were instructed to orbit the target area in a "left-hand" orbit, which generally required a sustained left bank to visually search the target area. The experiment did not prescribe a precise bank angle. The subjects simply had to control the aircraft as they would in the real-world equivalent task.

Maximum roll rates observed in the right and left directions during the talk-on period for each symbology were analyzed. The data was tested with a Kolmogorov-Smirnov (Right direction: KS=0.126,p>0.15; Left Direction: KS=0.119,p>0.15) and found to satisfy the normality assumption. A GLM AVOVA found a statistically significant effect of symbology on maximum right roll rate ($F_{2,18}$=6.91, p=0.006). A post-hoc Tukey t-test indicated a statistically significant difference between the CDF and the DFR (t=-3.292, p=0.0108) and the NDFR (t=-3.142, p=0.0148). These results are shown in Figure 17. Because the subject pilots' goal in the talk-on was to maintain aircraft control while in a left-hand orbit, right stick inputs were interpreted as "corrective", with the intent to decrease bank angle. A higher rate indicated a more aggressive correction. It was suspected that more aggressive corrections in the right direction observed in the CDF condition were in response to unintentionally steep bank angles that developed in the absence of attitude information while looking OBS.

Left roll rates also showed a nearly significant effect of symbology ($F_{2,18}$=3.27, p=0.061) as shown in Figure 18. These inputs steepened or increased bank angle while orbiting around the target area. This finding is consistent with the right roll input finding above, in that lower attitude awareness would necessitate more corrective control inputs to compensate for unintentional shallow bank angles. Together, these two metrics suggest that EPs were able to maintain more consistent control of bank angle with the DFR and NDFR symbologies when compared to the CDF.

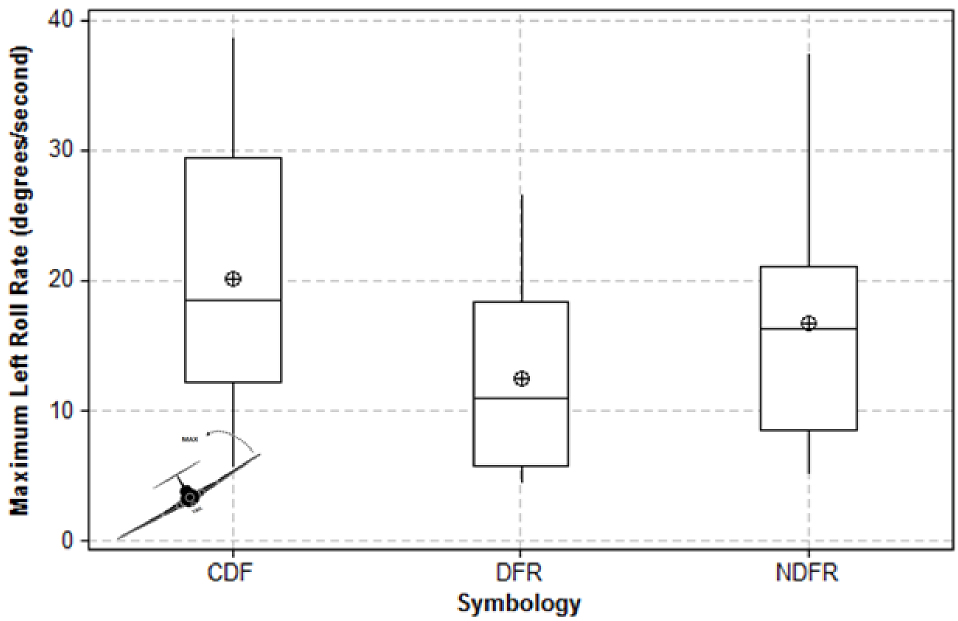


Source	DF	Seq SS	Adj SS	Adj MS	F	P
SubjectId	9	1176.17	1176.17	130.69	3.47	0.012
Symbology	2	520.55	520.55	260.28	6.91	0.006
Error	18	677.97	677.97	37.66		
Total	29	2374.69				

Symbology = CDF subtracted from:

Symbology	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
DFR	-9.036	2.745	-3.292	0.0108
NDFR	-8.622	2.745	-3.142	0.0148

Figure 17. Average Maximum Right Roll Rate by Symbology.

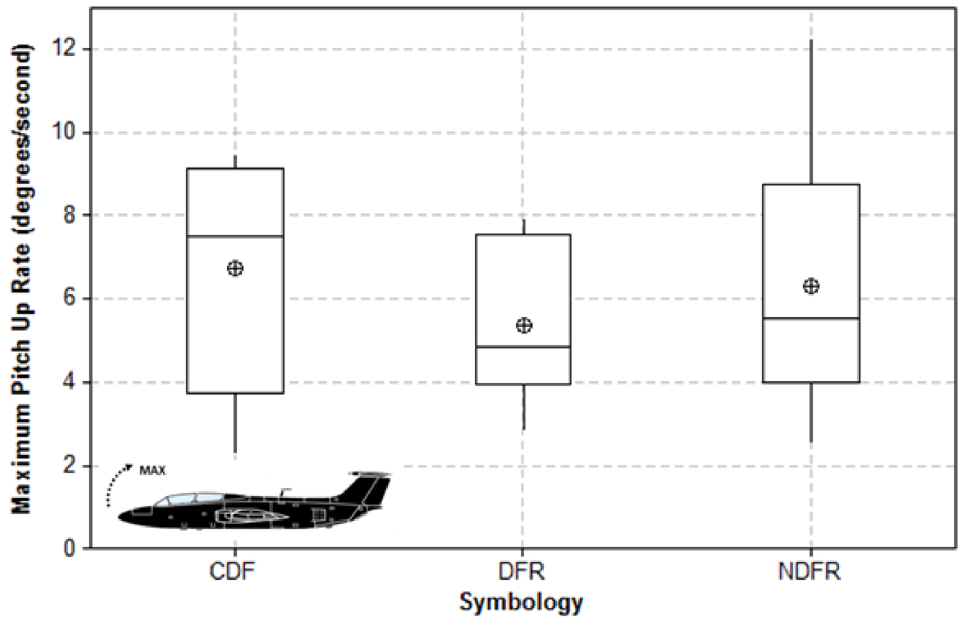



```
Source      DF  Seq SS   Adj SS Adj MS   F    P
SubjectId    9 1535.12 1535.12 170.57 3.83 0.007
Symbology    2  291.25  291.25 145.62 3.27 0.061
Error       18  800.61  800.61  44.48
Total       29 2626.98
Symbology = CDF subtracted from:
      Difference     SE of      Adjusted
Symbology  of Means Difference T-Value  P-Value
DFR        -7.619     2.983  -2.555   0.0498
NDFR       -3.424     2.983  -1.148   0.4981
```

Figure 18. Average Maximum Left Roll Rate by Symbology.

Examination of pitch rate revealed similar results. Higher values were interpreted as corrective inputs in response to undesired pitch attitudes and/or undesired climb/descent rates, as it was assumed the subject pilots were trying to maintain level flight. A Kolmogorov-Smirnov test (K=0.104, p>0.015) of this data indicated the normality assumption could be made. GLM ANOVA indicated a statistically significant effect of symbology for maximum observed pitch down rates ($F_{2,18}$=6.79, p=0.006) but not significant for pitch up rates ($F_{2,18}$=1.15, p=0.34). A Post-hoc Tukey t-test showed a statistically significant difference in maximum pitch–down rate between the CDF and DFR (t=-3.594, p=0.0056) and nearly significant difference between the CDF and NDFR (t=-2.50, p=0.055). As with roll, it appears that pitch control is better with the

DFR when compared to the CDF or NDFR test symbologies. These results are shown in Figure 19 and Figure 20.

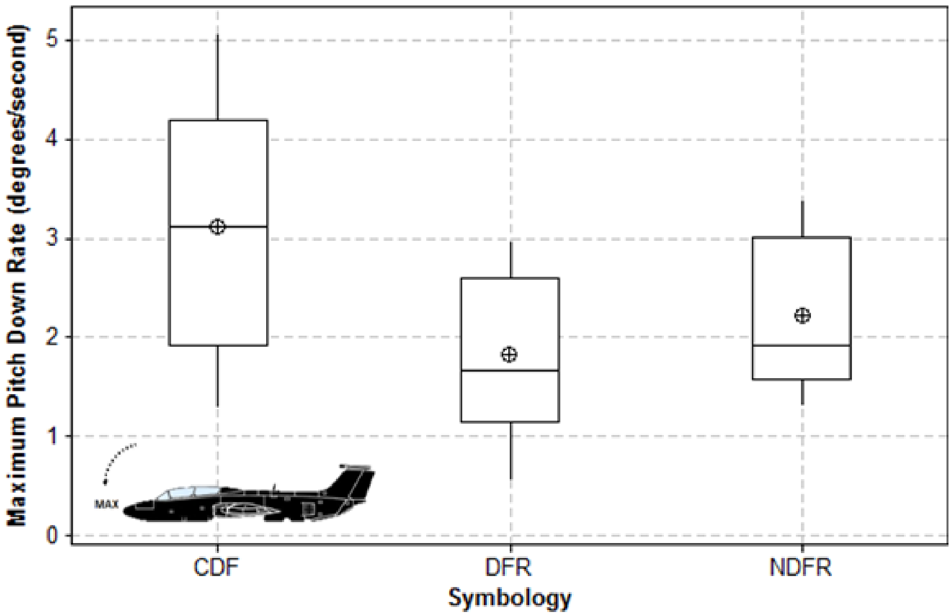


Source	DF	Seq SS	Adj SS	Adj MS	F	P
SubjectId	9	101.256	101.256	11.251	2.72	0.034
Symbology	2	9.490	9.490	4.745	1.15	0.340
Error	18	74.541	74.541	4.141		
Total	29	185.287				

Symbology = CDF subtracted from:

Symbology	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
DFR	-1.348	0.9101	-1.481	0.3232
NDFR	-0.426	0.9101	-0.468	0.8870

Figure 19. Average Maximum Pitch Up Rate by Symbology.




```
Source    DF  Seq SS  Adj SS Adj MS   F    P
SubjectId  9 14.4423 14.4423 1.6047 2.54 0.044
Symbology  2  8.5698  8.5698 4.2849 6.79 0.006
Error     18 11.3639 11.3639 0.6313
Total     29 34.3760
Symbology = CDF subtracted from:
      Difference    SE of      Adjusted
Symbology  of Means Difference T-Value  P-Value
DFR        -1.277    0.3553  -3.594   0.0056
NDFR        -0.888    0.3553  -2.500   0.0555
```

Figure 20. Average Maximum Pitch Down Rate by Symbology.

Workload and Situation Awareness

MWL and situation awareness were assessed subjectively using the Bedford Workload Scale and Situation Awareness Rating Technique (SART) respectively. Subject pilots completed these questionnaires verbally following each symbology condition. Mean values for each symbology condition were compared. A Kolmogorov-Smirnov test indicated that the normality assumption could be made for the Bedford (KS=0.03, p>0.15) and the SART data (KS=0.021, p>0.15). Subjective workload showed no significant effect of symbology. The GLM ANOVA on the SART ratings indicated a statistically significant ($F_{2,78}$=5.99, p=0.004) effect of symbology. Post-hoc pairwise comparisons showed significantly higher SART scores for the

NDFR condition (t=3.43, p=0.0027) and nearly significant for the DFR condition (t=2.03, p=0.11) compared to the CDF. These are summarized in Table 2.

Table 2. Workload and Situation Awareness Subjective Measures.

	Bedford	SART
CDF	4.75	7.0
DFR	4.0	7.75
NDFR	4.0	8.0*

* denotes statistically significant difference to CDF

ECG was recorded for each subject during all sorties. The portions from the talk-on phase were analyzed and used to calculate TPV-based relative workload. In this initial data analysis, mean values of N1 talk-on phases were compared. The below empirical CDF (Figure 21) shows a general trend of higher MWL in the DFR and NDFR conditions. However, this did not reach a level of significance. These results are elaborated in the remainder of this thesis.

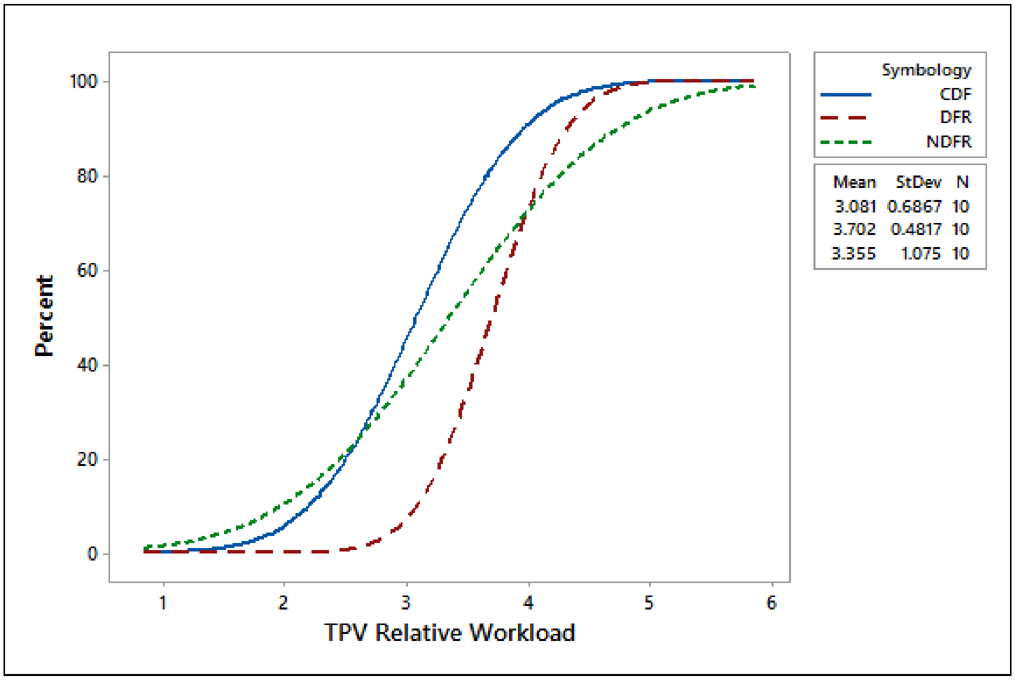


Figure 21. TPV-based Relative Workload by Symbology.

Conclusions

The results of this study effectively differentiated performance impacts of the three symbology sets as they related to SDO prevention. Analysis of task duration indicated that the DFR and NDFR facilitated a faster and more efficient talk-on than the CDF. This finding is of tactical significance as time is of the essence in real-world CAS scenarios. The flight technical data, namely roll and pitch rates, generally show that the DFR enabled the most stable control of the aircraft during OBS activities. Both the DFR and NDFR outperformed the CDF in this regard. These findings support the conclusion that the ASAR may be beneficial in preventing unusual attitudes from developing and leading to SDO in tactically demanding OBS visual tasks when compared to the CDF.

CHAPTER 4 – ANALYSIS METHODOLOGY

This chapter details the methodological approach for the HRV analysis. The objective was two-fold. First, it sought to explore the utility of the TPV method relative to classical HRV methods in a live-flight tactical setting. Second, it sought to provide additional insights on MWL characteristics of the candidate HMD symbologies as discussed in the previous chapter. The approach utilized mixed-methods and was organized in three levels described below.

Three Level Approach

This comparative analysis used mixed methods (quantitative and qualitative) and was approached on three levels of granularity. At the first and highest level, all HRV methods considered were used to calculate and compare mean values for the talk-on portion of the N1 sortie for all subjects. This level of analysis compared sensitivity of the HRV metrics in consideration. At the second level, three subjects of interest were chosen because analysis of video and subjective feedback indicated considerable variance in MWL between the three symbology conditions. At this level, empirical cumulative distribution functions (CDF) of the time series of LF power were compared to the TPV method to examine whether the distribution of values over time could elucidate MWL characteristics between symbology conditions. The third level of analysis drilled further down to significant events. In this analysis, four special case talk-ons containing significant events determined to drive momentary changes in MWL were selected. For these events, individual talk-on portions selected were plotted to show HRV metrics' response momentary MWL variations (on the order of ~30 seconds) when plotted in time series. The first level of analysis was purely quantitative while the second and third levels focused on qualitative insights. Each level sought to both compare HRV metrics and supplement the analysis of the HMD study.

The analysis used two basic criteria for assessment: sensitivity and diagnosticity. Sensitivity of measures was assessed by the degree to which the measure can distinguish levels of MWL between the candidate symbologies. Diagnosticity was assessed by the degree to which the method could distinguish individual task elements driving changes in MWL.

The rationale for conducting this analysis in conjunction with the previously discussed HMD study was that is representative of many of the challenges associated with MWL measurement in tactically relevant settings. There are several elements which make this dataset unique amongst settings in which physiologic measures are applied. First and foremost, the study was designed to represent a real-world CAS mission, which makes findings bear operational relevance. Second, because this study was conducted in live flight, the sensor ensemble needed to be equipped in such a way that it was compatible with a real-world flight equipment ensemble. Third, the durations of the time periods of interest for comparison were variable in length, which is often considered problematic for HRV analysis. Fourth, the initial analysis indicated subtle MWL differences between test symbologies, making this dataset useful for rigorously evaluating sensitivity. For these reasons, this analysis provides useful insights for evaluating design alternatives in a tactically relevant setting.

Experimental Data

To maintain consistency with initial analysis of the HMD study, data analyzed herein was extracted from only the talk-on portion of the N1 sortie. Both raw ECG signals and calculated TPV-based workload values from the previous analysis were utilized. Raw ECG signals were processed and used to calculate time and frequency domain HRV metrics. Additionally, instantaneous roll and pitch rates collected during the previous analysis were utilized. All raw data was contained in comma separated value (CSV) format.

Software

HRV analysis was accomplished using the Kubios HRV version 3.0.2 software package (Tarvainen et al., 2014). This included preprocessing of the data (loading, filtering ECG signals) and implementation of the algorithms to compute time and frequency domain HRV indices.

OPL developed algorithms and the CATS architecture were used to calculate TPV-based workload. Plots depicting flight technical data and workload were generated in LabView 2015 Student Edition (Elliott, Vijayakumar, Zink, & Hansen, 2007). Statistical tests, boxplots, and empirical CDFs were generated Minitab 2014 and 2017.

Calculations

For the first level of analysis, mean values for the duration of the N1 talk-on portions were calculated using the formulas described in Chapter 2 in the Kubios software. Spectral powers were estimated using Welch's periodogram method (Welch, 1967) for FFT, in which the RR series is divided into overlapping segments (windows). The window width and overlap for this portion of the analysis were 300 seconds and 50% respectively. TPV Relative Workload was calculated using CATS. This analysis used a time delay of 12 seconds and embedding dimension of 4 to generate the ETMs. Additionally, a buffer size of 1024 samples and overlap of 50% were used to generate an average workload value at exactly 1 second intervals. Results from analysis of each individual subject were compiled to preform statistical analysis.

For the second and third level of analysis, a window average method was used to calculate time-varying values for time and frequency domain metrics. This window was calculated using a 30 second window with 29 second overlap to establish a time series of values at 1 second intervals. The rationale for this was for ease of comparison with the TPV dataset, as described above, which generates a unique value at 1 second intervals. For the second level of analysis, these values were used to generate empirical cumulative distribution functions (CDF) for the talk-on periods of each symbology for the selected subjects.

For the third level of analysis, the time series data for each metric were imported into the Labview software to generate time series charts of the talk-ons for comparison with flight technical data for the selected events. Flight technical data that was included (roll rate and pitch rate) was collected continuously at a sampling rate of 50 Hz. In the raw data extracted from CATS, these values were calculated continuously for each sample. As a result, the raw data series contained considerable noisy fluctuations. For this reason, a moving window average of 1 second (50 samples) was used to smooth the time series for plotting.

Criteria for Comparison

This analysis focused primarily on two of the criterion for assessing MWL measures described previously in Chapter 2: sensitivity and diagnosticity. Sensitivity was assessed primarily in the first two levels of analysis. This was simply evaluated as the degree to which the HRV metrics in question could differentiate MWL between the three symbology conditions. This was analyzed quantitatively in the first level of analysis by means of statistical tests. The second and third level of analysis employed more qualitative approach through visualization of the data. Diagnosticity was evaluated as the ability of the HRV metrics to identify MWL driving elements of the talk-on task. This criterion was evaluated solely in the third level of the analysis. This approach was somewhat unconventional relative to other studies which have compared MWL measurement techniques in that mental resource types and task elements were not controlled for. Rather, this analysis treated the temporal sensitivity of the time series plots for the included metrics as a diagnostic quality. Through review of video and audio, specific elements/events of the talk-on (e.g. visual search, unusual events) identified as MWL-driving were correlated with elements of the HRV metrics' time series. The responsiveness of each metric to elements was interpreted as their level of diagnosticity.

CHAPTER 5 – RESULTS AND DISCUSSION

This chapter is divided into three subsections to align with the three-level analysis approach described in the previous chapter. First, sensitivity analysis on all ECG metrics of interest is reported. In this sense, sensitivity refers to the degree with which each metric could detect MWL level differences between the three symbologies based on comparison of mean values for the entire talk-on. The second and third section show results of time varied calculations of each metric. In the second section, three subjects of interest who, based on analysis of video and subjective feedback, exhibited substantial differences in MWL between symbologies. These data are presented with empirical CDFs to compare the distribution of values. The focus of this section is also sensitivity but narrowed to a within-subject scope. The third section presents a qualitative case analysis of select subjects in which significant MWL driving events during the talk-on phase were identified. The third section focuses on diagnosticity with respect to specific MWL-driving events in the talk-on phase.

Sensitivity – Talk-On for All Subjects

Mean values for HRV metrics were calculated for the talk-on phase to examine which metrics could effectively detect MWL differences between symbologies. Descriptive statistics (mean and standard deviation) are shown in Table 3. These data show several key results. First, for most of the time domain metrics, the differences in means between the DFR and the CDF and NDFR exhibit the same general trend observed in the initial MWL analysis discussed in Chapter 3 (TPV Workload also in Table 3), which indicated the highest MWL occurred in the DFR condition. In this case, these metrics show lowest HRV in the DFR condition. The only exception to this trend was mean HR, which showed almost no difference between symbology conditions.

The results from the frequency domain, however, appear to diverge from the time domain metrics and TPV workload. The condition eliciting the highest MWL would be expected to show the highest value in the LF power spectrum, lowest value in the HF power spectrum, and highest LF/HF ratio. The opposite trend was observed. However, the results of the statistical tests discussed below (Table 4) show the weakest effect of symbology on these metrics, which may partially explain this divergence. This may also have been impacted by the fact that respiration effects could not be accounted for since this data was not available.

Table 3. Descriptive Statistics for Talk-On HRV Metrics (N=10).

		CDF		**DFR**		**NDFR**	
	Metric	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Time Domain	Mean HR (bpm)	81.32	19.33	80.97	18.71	80.55	23.13
	SDNN (ms)	72.58	28.02	58.37	27.26	73.40	40.30
	RMSSD (ms)	75.00	39.50	60.60	38.10	76.10	49.40
	pNN50 (%)	22.86	13.61	18.35	12.35	23.41	14.58
	HRVTI	11.01	3.500	9.940	3.42	11.57	2.689
	TINN	417.3	204.3	309.2	175.5	389.1	237.4
Frequency Domain	LF Power (nu)	55.05	23.44	51.15	16.02	55.63	21.75
	HF Power (nu)	44.73	23.23	48.68	16.02	44.02	21.42
	LF/HF Ratio	1.947	1.716	1.301	0.920	2.174	2.407
Nonlinear Domain	TPV Workload	3.081	0.687	3.702	0.482	3.355	1.075

The statistical summary for the global sensitivity analysis is shown in Table 4. Kolmogorov-Smirnov tests were applied to all the data to test the normality assumption. If the normality assumption could be made, General Linear Model Analysis of Variance (GLM ANOVA) was performed. In the case of violations of the normality assumption and no suitable transformation could be found, Kruskal-Wallis nonparametric rank tests were applied. As Table 4 shows, the sample distributions for half of the metrics analyzed did not follow a normal distribution.

Table 4. Statistical Tests for Talk-On HRV Metrics (N=10).

	Metric	Normality Test	Significance Test	Test Statistic	p value
Time Domain	Mean HR (bpm)	KS=0.223; p<0.010	Kruskal-Wallis	H=0.19	0.910
	SDNN (ms)	KS=0.152; p=0.076	Kruskal-Wallis	H=2.11	0.348
	RMSSD (ms)	KS=0.152; p=0.076	Kruskal-Wallis	H=1.23	0.542
	pNN50 (%)	KS=0.162, p=0.046	Kruskal-Wallis	H=1.14	0.566
	HRVTI	KS=0.124, p>0.150	ANOVA	$F_{2,18}$=1.27	0.304
	TINN	KS=0.130, p>0.150	ANOVA	$F_{2,18}$=3.32	0.059
Frequency Domain	LF Power (nu)	KS=0.129, p>0.150	ANOVA	$F_{2,18}$=0.26	0.775
	HF Power (nu)	KS=0.132, p>0.150	ANOVA	$F_{2,18}$=0.28	0.759
	LF/HF Ratio	KS=0.187; p<0.010	Kruskal-Wallis	H=0.34	0.842
Nonlinear Domain	TPV Workload	KS=0.133, p>0.150	ANOVA	$F_{2,18}$=2.82	0.086

In short, test symbology did not produce statistically significant effect for any of the metrics tested at an alpha level of 0.05. This could possibly be explained by the small sample size (N=10) and the wide sample variances within each metric. Additionally, there was no difference observed in subjective (Bedford) ratings from the previous analysis. Other than subjective reports, there was no other way to evaluate "ground truth" MWL levels between symbologies. This indicates the MWL difference between symbology conditions was subtle (perhaps barely existent). Interestingly, the one metric that came the closest to significance was the TINN (p=0.059), followed closely by TPV workload (p=0.086). The fact that TINN showed a stronger effect than other metrics was surprising given that the literature indicates it is the least well-suited for the short duration recordings evaluated in this study. The nearly significant effect of TPV workload is consistent with the hypothesis. Boxplots of two time domain (SDNN) and two frequency domain (LF Power and HF power) metrics are shown in Figure 22 and Figure 23 respectively for visualization. Figure 24 shows mean TPV workload for comparison.

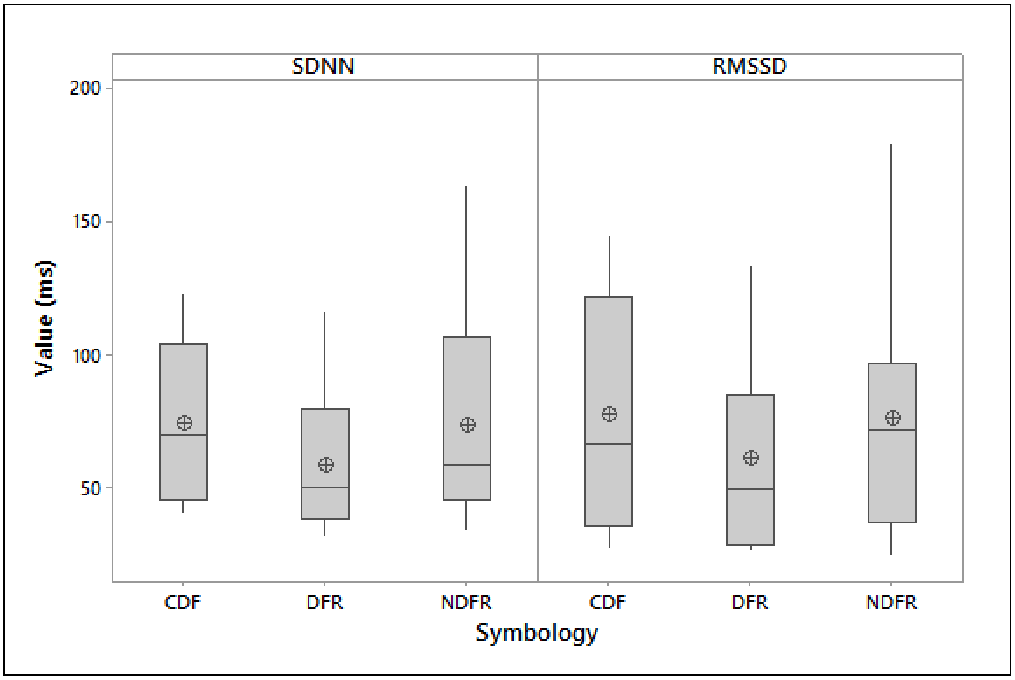


Figure 22. Mean SDNN and RMSSD by Symbology.

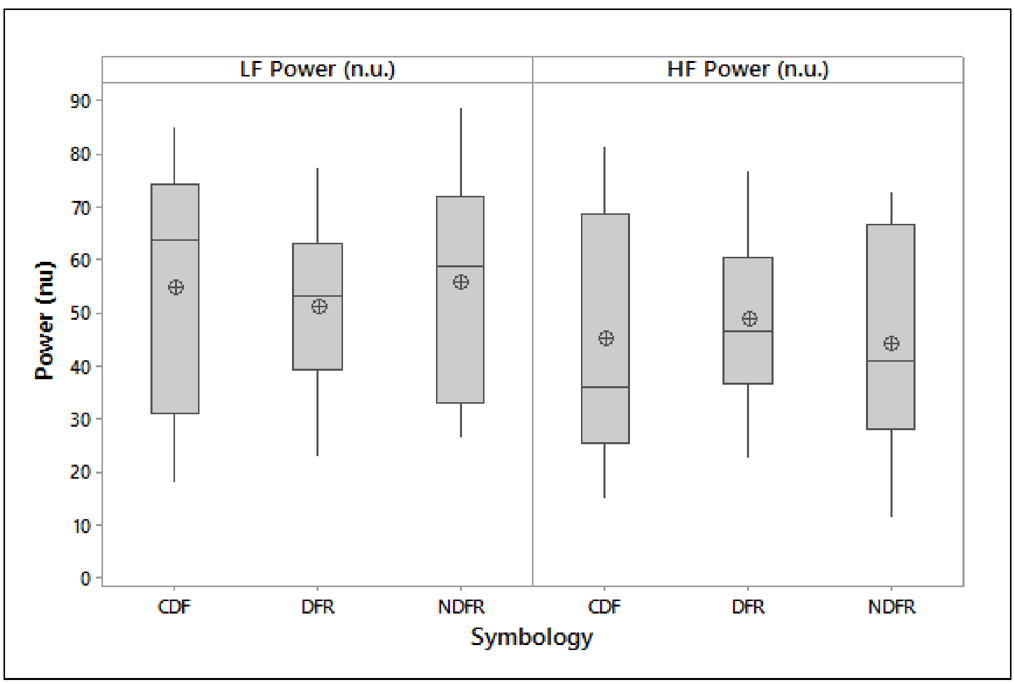


Figure 23. Mean LF Power and HF Power by Symbology.

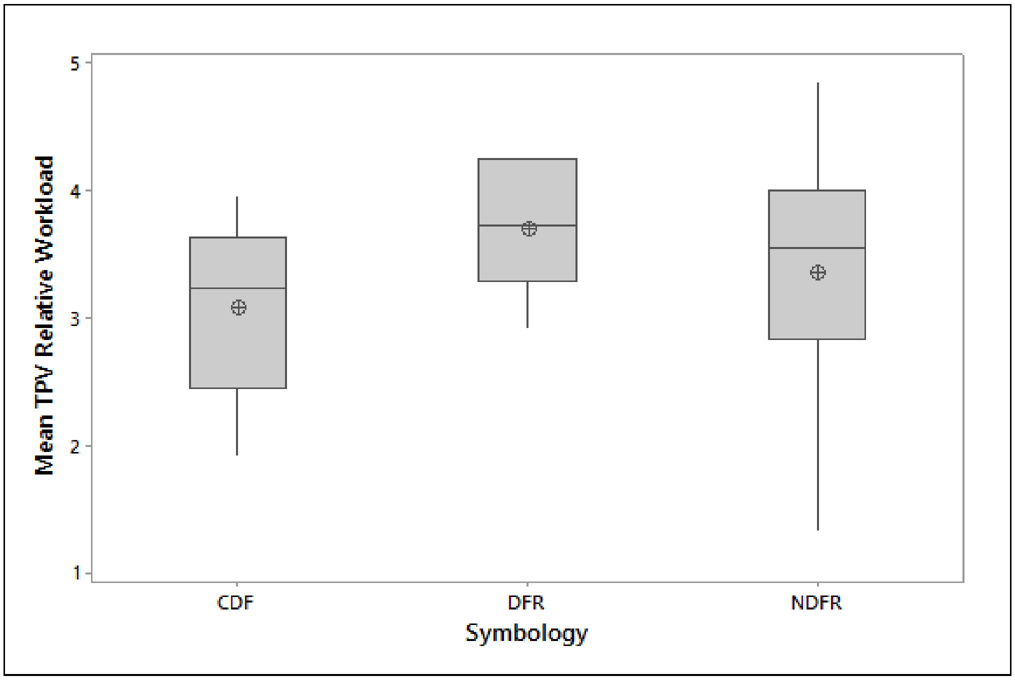


Figure 24. Mean TPV Relative Workload by Symbology.

Time Varying Sensitivity - Within Subject Comparisons

Three subjects were compared with exhibiting noticeable MWL differences between symbologies after review of video, audio, and subjective feedback. TPV workload and LF power, calculated in time series, were used to generate the empirical CDFs shown in Figure 25-Figure 30. These CDFs elucidate the distribution of these variables over time as they relate to the symbology conditions for the duration of the talk-on. A general trend that can be observed, which again validates the findings from the original MWL analysis, is that the CDF consistently elicited overall lower MWL than the NDFR and DFR.

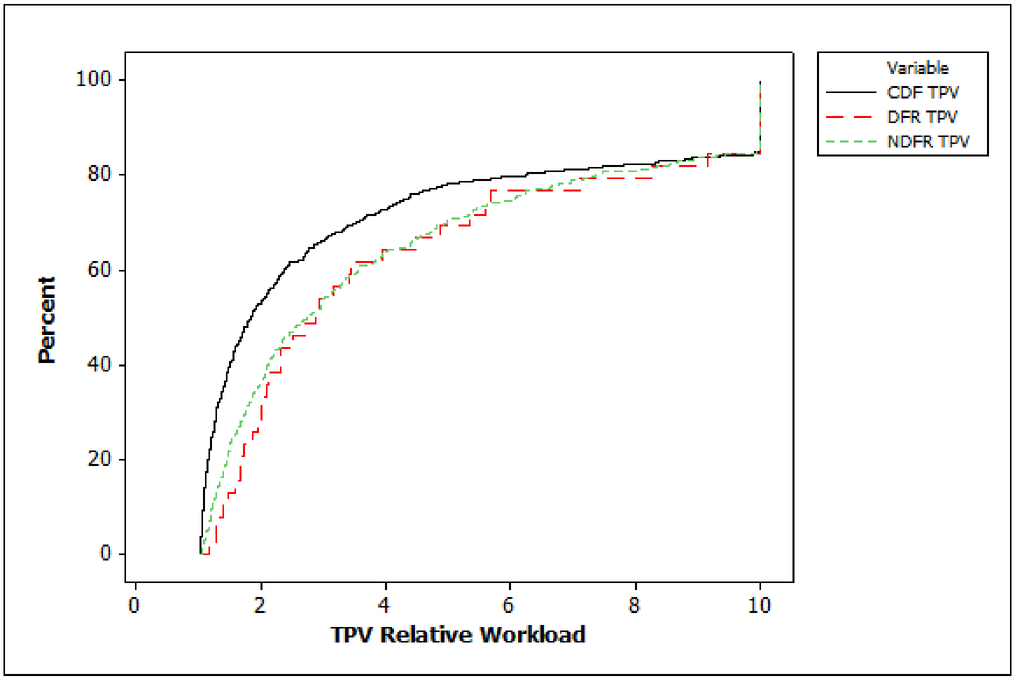


Figure 25. CDF of Subject 1 TPV Relative Workload for N1 Talk-On.

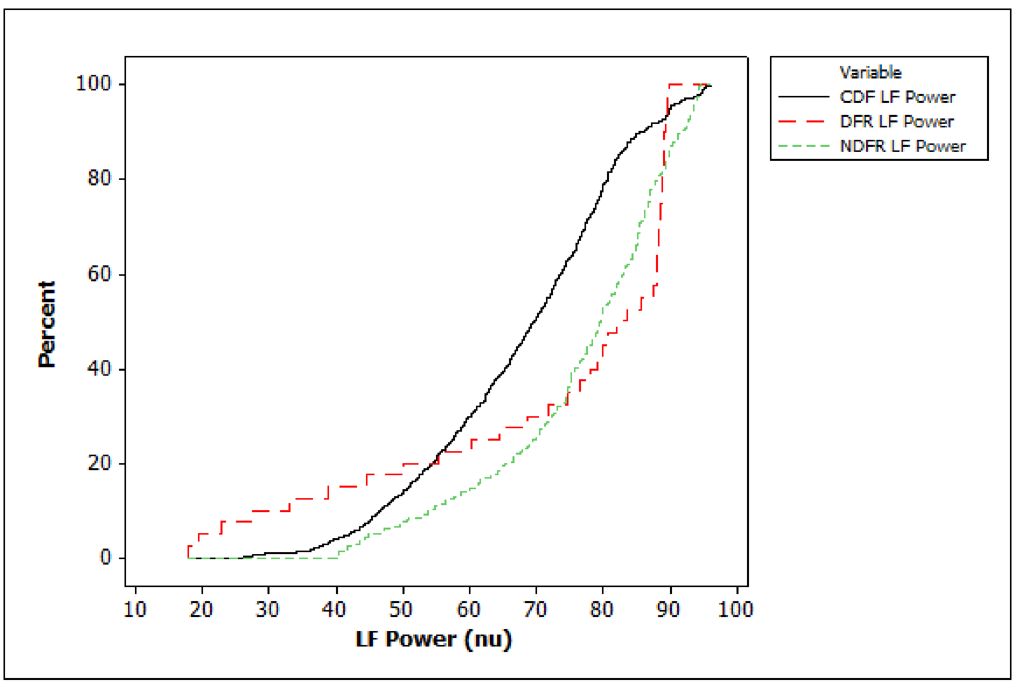


Figure 26. CDF of Subject 1 LF Power for N1 Talk-On.

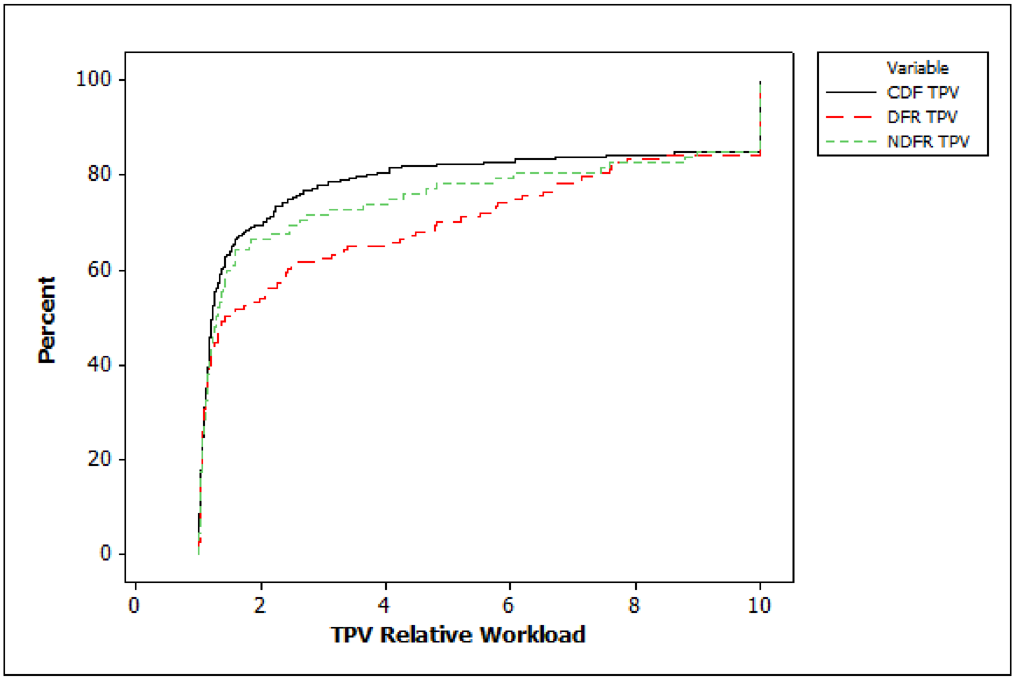


Figure 27. CDF of Subject 5 TPV Relative Workload for N1 Talk-On.

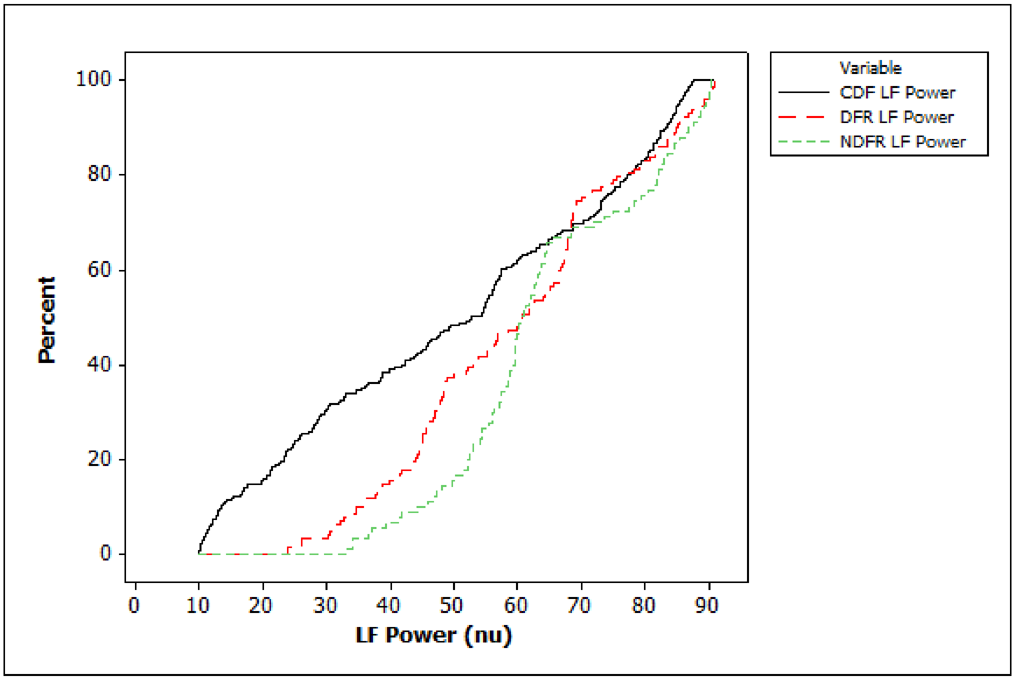


Figure 28. CDF of Subject 5 LF Power for N1 Talk-On.

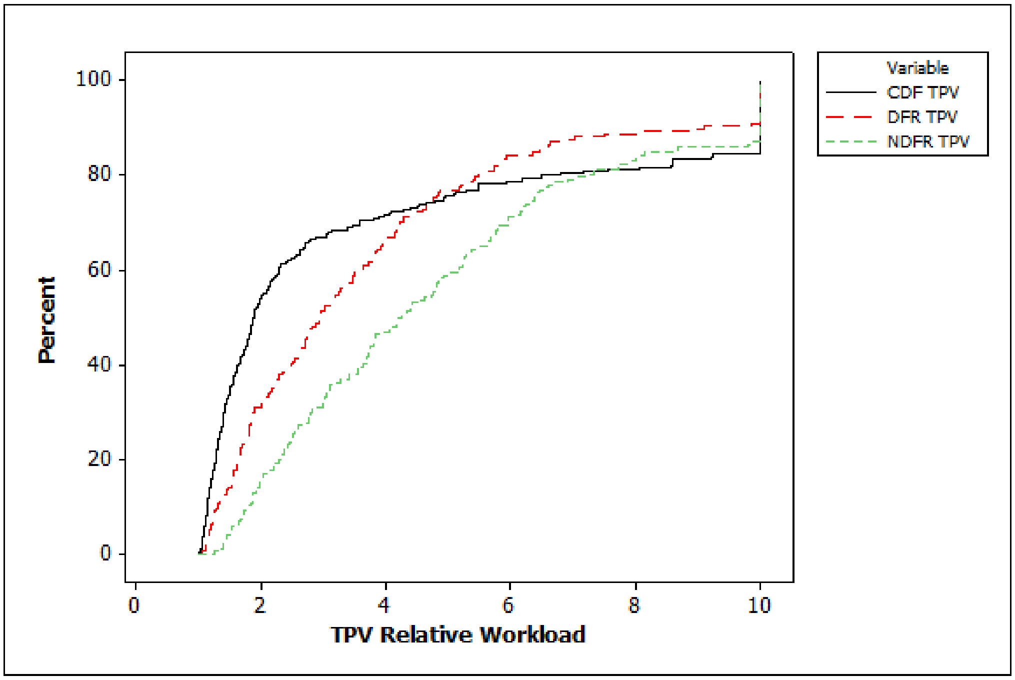


Figure 29. CDF of Subject 11 TPV Relative Workload for N1 Talk-On.

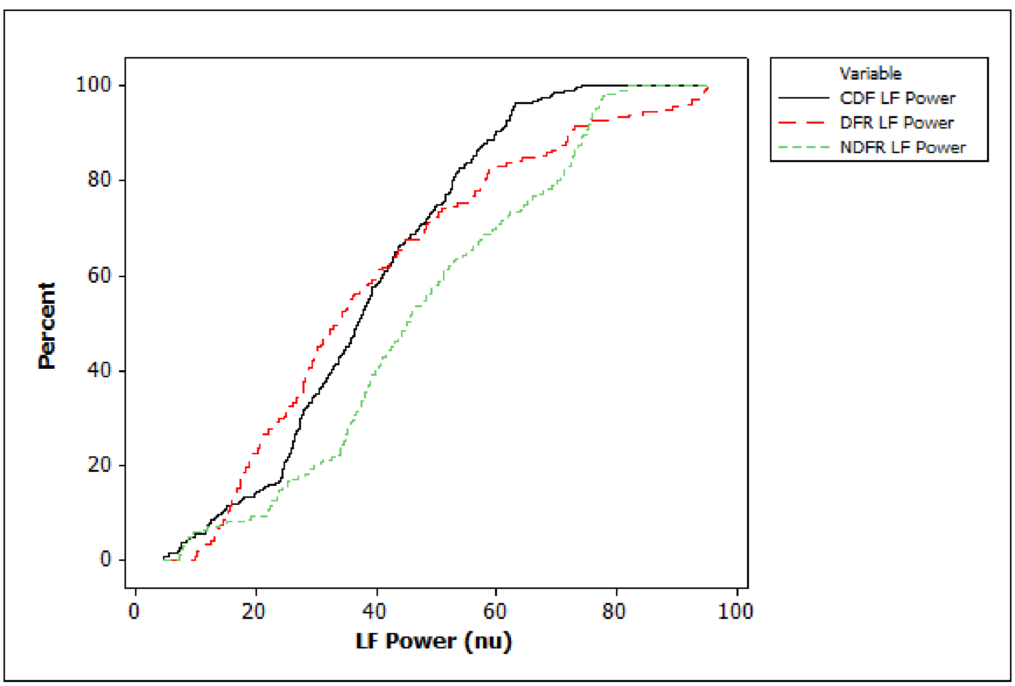


Figure 30. CDF of Subject 11 LF Power for N1 Talk-On.

Diagnosticity - Event Identification

Figure 31-Figure 34 show four different special case talk-on phases that were selected after careful analysis of video, audio, and workload data for all subjects during the N1 sortie. These represent the most prominent findings from this qualitative analysis relating to the performance, specifically with respect diagnosticity, of the HRV metrics in question. Each chart shows the change in the respective metrics plotted in time series for the entire duration of the talk-on. The top row shows roll and pitch rate to visualize the relationship of HRV to aircraft control. These rates are shown as absolute values, meaning they indicate only the magnitude of the rate but not the direction. The second row depicts HR in beats per minute (bpm). The third row shows the one selected time domain metric, SDNN, in milliseconds (ms). It must be noted that despite performing well in the first level of analysis, the TINN metric was not suitable for calculating short-duration, windowed values in time series, and was therefore excluded from the second and third levels of analysis. In the fourth row, LF power in normalized units (nu) is shown to represent the frequency domain. The last row shows the TPV-based workload. In each graph, shaded areas marked with letters depict periods of interest, which are discussed in the narrative below. In most of the highlighted events, it was expected that each metric would indicate an increase in MWL. The talk-on from Subject 3 (Figure 33) was the one exception to this, in which the selected segments demonstrate diagnosticity at low MWL levels.

Figure 31 shows the talk-on for Subject 1 in the CDF condition. There were five segments of interest noted in the video review, which were as follows:

A) The pilot is looking intensely at Placemat Alpha 01 while simultaneously talking to the SP and maneuvering the aircraft to establish a clear visual angle to the target area. The pilot's visual scan is alternating between the placemat and the outside visual scene to orient by correlate terrain features. In this segment and

(B), there is heavy distracting communication from other non-tactical air traffic over the radio.

B) After the pilot has visual oriented to the target area, he is communicating over the radio and searching to establish visual contact with the reference point, FOB Ranger. This visual search continues for approximately 1 minute while maintaining line-of-sight purely in the OBS regime. Of note, the pilot must make several corrective stick inputs to keep the aircraft in the appropriate position to facilitate the visual search.

C) After acquiring visual contact with FOB Ranger, the pilot is distracted by a failure of the upfront lower multi-function display (MFD). He attempts to troubleshoot this issue while in coordination with the SP and EPC. While troubleshooting, the pilot abandons the visual search and loses contact with FOB Ranger. Small stick corrections are observed as the pilot works to maintain level flight.

D) During this period, the pilot again attempts to establish visual contact with FOB Ranger, while again aggressively maneuvering the aircraft for ideal visual angle on the target area.

E) The pilot has made visual contact with the target and is confirming the shape and orientation of the building back to the JTAC.

Overall, this was one of the only selected portions to demonstrate appreciable changes in mean HR, as it shows a slight elevation in segments A, B, and D. Lower relative variability is observed in the initial portion in SDNN, which correspond to elevations in TPV workload. This segment shows generally high LF power, but no consistent diagnostic trend. Elevated TPV workload is observed in all 5 highlight segments. Also in each highlighted segment, effortful and somewhat erratic control of the aircraft is observed and also reflected in the roll and pitch rate plots. More erratic aircraft control was suspected to be a trend in the CDF condition from the initial data analysis and is further validated here.

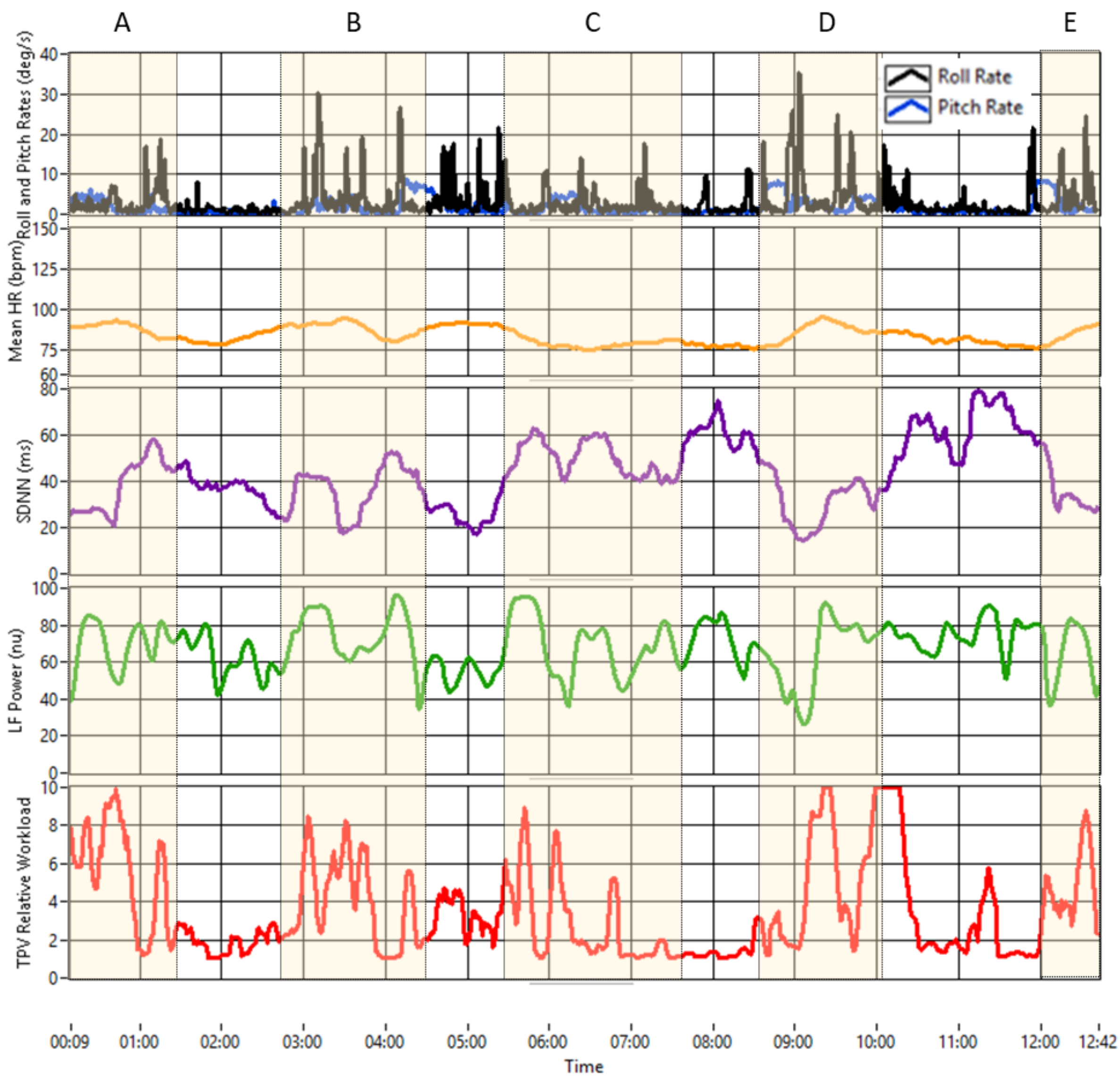


Figure 31. Case 1: Subject 1, CDF Talk-On.

Figure 32 shows the NDFR talk-on for Subject 1. The significant events in the shaded areas are as follows:

A) The pilot is communicating with the JTAC and describing the shape and orientation of a group of buildings which he believes contains the target. The pilot's visual angle is in the OBS regime. There is apparent uncertainty in the verbal exchange with the JTAC.

B) After a moment of confusion, the pilot and JTAC realize they are describing different building group. After reexamining the placemat, the pilot realizes he is searching on the wrong side of the reference road and makes an aggressive correction to reposition the aircraft.

C) During this approximately 1:30 period, the pilot is alternating his scan between the placemat and the outside visual scene to find the building group of interest, with dialogue indicating apparent confusion and high effort level.

Trends observed in this talk-on bear some similarity to the previous. The visual search task, with respect to both the target area and the placemat, elicited spikes in MWL. These again are most apparent in TPV workload. SDNN shows modest relative reductions in segments A and C, while LF power was again generally elevated but did not display a consistent trend. This talk-on represents two other important trends uncovered in analysis. First, a state of confusion, as explained in all three segments, appeared to be a significant workload-driving element. Secondly, it is apparent that the visual search task was associated with more stable aircraft control in this condition (NDFR) than the CDF shown previously. A similar effect was observed in the DFR condition.

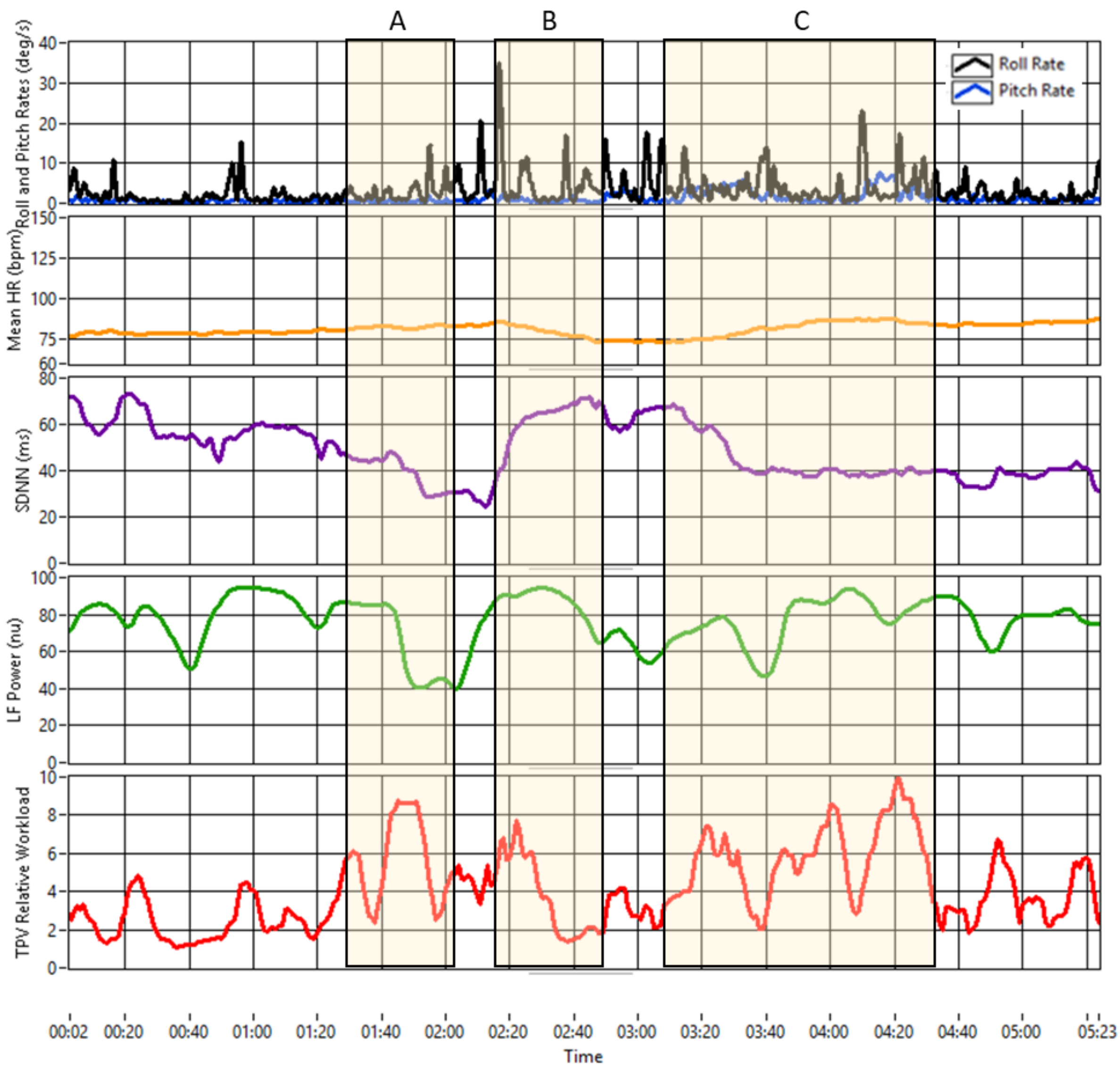


Figure 32. Case 2: Subject 1, NDFR Talk-On.

Figure 33 shows the CDF talk-on for Subject 3. Just two notable items segments in this talk-on phase are shown, which are as follows:

A) The pilot is executing a visual search for the target area. The pitch and roll rate traces indicate erratic aircraft control with very little effort exerted.

B) The pilot verifies a positive ID of the target and communicates this confirmation to the JTAC.

Subject 3 was a unique case in that this subject was noted as putting forth exceptionally low effort during the experiment. This is reflected in the relatively flat plots for each variable. A small increase in TPV workload is shown in segment B, but not reflected in the other variables. Segment A shows the visual search task. Contrary to the previous two examples, low workload is reflected in the TPV, yet somewhat erratic aircraft control is reflected in the roll rate plot. This example demonstrates how the combination of physiologic workload and task performance, in conjunction, can be diagnostic. In this example, the combination of low MWL and poor task performance validate the observation for the previous two examples that the visual search was a MWL-driving element of the talk-on.

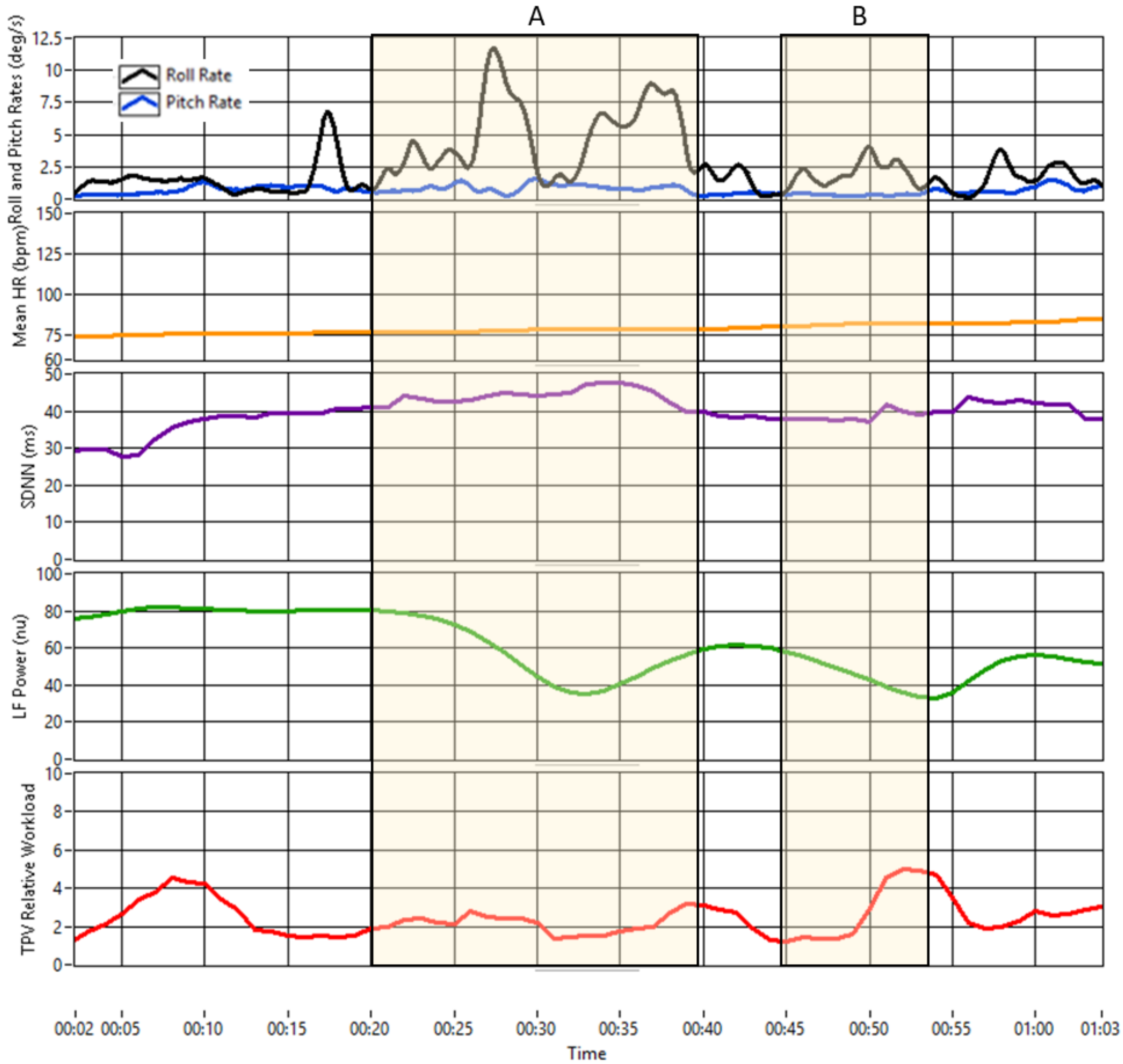


Figure 33. Case 3: Subject 3, CDF Talk-On.

The last selected case was the CDF talk-on for Subject 5 shown in Figure 34. This case differs from the previous three in that recording includes an unsuccessful attack attempt. The previous recordings terminated at the end of the talk-on (prior to attack execution). This was included to highlight several interesting findings. The significant events in the shaded areas are as follows:

A) The pilot is searching intently for FOB Ranger while making heading adjustments to gain visual line-of-sight to the target area. The pilots visual angle to target is large, requiring an effortful gaze far OBS. Meanwhile, he is adjusting aircraft heading with small stick inputs to reverse course to a south facing heading.

B) This segment shows an event unique to this case. The pilot is rolling onto a short final to execute a weapon delivery on the target. The final roll-in occurs at 2:30.

C) For an undetermined reason, the pilot in unable to release the weapon as planned. This may have been due to inattention with respect to the attack parameters.

D) In the segment (around 4:40), the aircraft nearly impacts a ridgeline while approaching the target area. Just prior to the end of this segment, the pilot reacquires visual contact with the target and communicates this to the JTAC.

Contrary to the other examples, the visual search in segment A only elicited a small elevation in MWL. SDNN was relatively constant and low throughout. A spike in TPV workload is apparent in segment B as the pilot attempts to establish the aircraft within the proper attack parameters. This elevation is also apparent in increased LF power. Segment C shows unusual behavior in the SDNN, LF power and TPV plots. High SDNN and low LF power values indicate very low MWL, which is contrary to corresponding TPV value. The difference could be explained temporal lag in the SDNN and LF power plots. The pilot's failure to release the ordinance on time may have been a result of inattention, reflected in lower MWL just prior to this segment. Finally, as expected, the near collision with the ridgeline is segment D corresponds with heightened TPV workload, lower SDNN, and increased LF power.

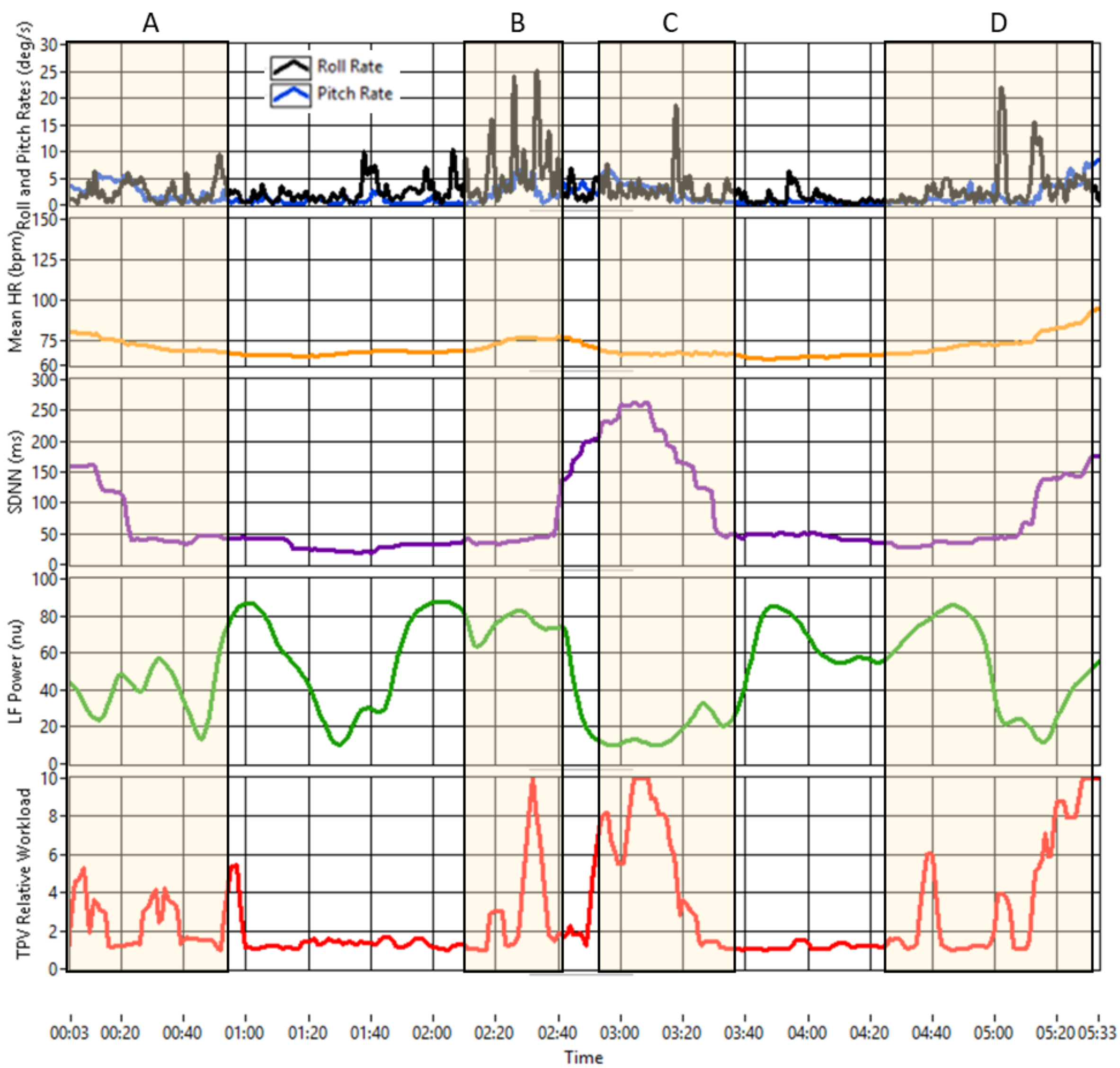


Figure 34. Case 4: Subject 5, CDF Talk-On.

CHAPTER 6 – CONCLUSIONS

Limitations

There are several limitations of this research that must be disclosed. Likely the most notable is the unconventional design for the analysis. The HMD study was not deliberately designed to compare MWL assessment methods. Had it been, different controls could have been implemented to better address the evaluation criteria, particularly regarding diagnosticity. However, the approach was deemed sufficient due to its applicability to applied research settings in which experimental control is often difficult to fully achieve. It should also be noted that all HRV metrics were analyzed from the same dataset and are likely related. This has been noted by others conducting studies comparing various HRV metrics (Heine et al., 2017; Verwey & Veltman, 1996).

This analysis did not account for respiration effects in the calculation of HRV metrics as this data was not available. This factor likely impacted the calculation of frequency domain measures most of all and may explain the lack of significant effect for these metrics in the first level of analysis. The precise effect of respiration on the TPV workload is unknown and warrants further investigation. Also, in the third level of analysis, some of the visual searches displaying heightened MWL levels were accompanied by talking (i.e. comm with the JTAC).

The study examined short duration and variable recording lengths for comparison. As the literature indicates, comparisons on unequal duration recordings is unconventional, and even discouraged in clinical HRV analysis (Task Force, 1996). However, the design of the experiment did not allow for control of the length of ECG recordings as the talk-on phases played out as required to meet the objectives of the scenario.

ECG in general is subject to motion artifact from non-cardiac muscle activity. This has not been fully explored and its impacts not fully characterized with respect to the TPV method. Further research is needed to account for motion artifact when using ECG to assess MWL in tasks with a significant physical component. However, this is currently an ongoing effort at the OPL.

Effects of increased G forces were also not accounted for. To the best of the author's knowledge, this effect has not been explored with specific focus on changes in nonlinear properties and complexity of the ECG signal. It likely had little impact in this analysis as G forces were minimal during the CAS scenario. However, this warrants research to strengthen the validity of the TPV and potentially other nonlinear HRV metrics in live-flight tactical settings with more dynamic maneuvering (e.g. Air-to-Air scenarios).

Summary of Findings

This research compared the TPV workload method to nine different classical time and frequency domain HRV analytical methods. The goal was to evaluate the relative utility of these methods in a live-flight, tactically relevant task. Sensitivity and diagnosticity were the primary criteria for the evaluation. Additionally, this research sought to provide additional insights into the differences in MWL characteristics of the candidate symbologies driving SDO-risk in the HMD study.

Review of the literature indicated a need to further explore the nonlinear domain in MWL applications of HRV analysis. Although classical time and frequency domain methods are well validated, they have limitations that may be overcome by incorporating the nonlinear domain. While this research only explored one nonlinear method in detail, the findings warrant further investigation of the nonlinear elements of the ECG signal. Further, the literature indicates ECG continues to be one of the more suitable physiologic signals to assess operator MWL in live-flight settings.

The three-level analysis, overall, indicated comparable if not superior performance of the TPV method relative to most other methods analyzed with respect to sensitivity and diagnosticity. With the exception on the TINN method, the TPV method showed higher ability to detect workload differences between the symbologies than the other methods. The general trend of the DFR condition showing the highest workload validated the findings of the initial analysis in the HMD study. The empirical CDFs in the within-subjects comparisons demonstrated the TPV method could effectively visualize the time-varying distributions of workload for qualitative comparisons between task conditions. However, in this respect performance of the TPV method and the spectrogram FFT method (LF power) produced comparable results. Nonetheless, this level of analysis also validated the original findings. This type of analysis revealed important information not apparent when comparing mean values for the duration of the task.

The third level of analysis generated several interesting findings. First, it validated the expectation that the TPV method would exhibit a high level of temporal sensitivity and ability to detect momentary fluctuations in workload. The time domain metrics, especially mean HR, did not show an appreciable level of diagnosticity with respect to the highlighted workload driving events. LF power showed modest temporal sensitivity at certain times but not at others.

Secondly, it revealed the TPV method could be diagnostic to workload-driving elements in the talk-on task. This enriched the MWL analysis of the HMD study. It became apparent that the visual search task and unusual events (e.g. display failures) induced higher MWL. It further revealed insights into the relationship between task performance, which in this case referred to aircraft control, and MWL. This helped support conclusions in the initial analysis regarding flight technical-related precursors to SDO.

Applications

The TPV workload method exhibits strong temporal sensitivity, such that it shows potential for detecting MWL fluctuations in near real time. In the context of this study, it proved useful for assessing MWL in SDO-risk assessment. SDO is a rare but deadly event and no bona fide incidents were observed in this study. However, TPV seemed to correlate with flight technical data indicating impending SDO events.

A second potential application of this method is real-time operator state assessment and adaptive automation such as described in Scerbo et al. (2001) and Kraft et al. (2017). Model-based workload assessment is beyond the scope of this thesis. However, the findings herein may prove useful for efforts to incorporate psychophysiological measures into existing models.

Conclusion

Overall, this research met the objective of establishing the relative utility of the TPV method live-flight tactically relevant settings. One should be cautious to overextend these conclusions given the applied nature of this study. Nonetheless, these findings contribute to knowledge of nonlinear HRV analysis in MWL research. Further efforts in this area are warranted.

REFERENCES

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage, 59*(1), 36-47.

Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., . . . Stone, P. H. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology, 34*(6), 623-648.

Bornas, X., Llabrés, J., Morillas-Romero, A., Aguayo-Siquier, B., Balle, M., & Tortella-Feliu, M. (2013). Complexity of everyday life heart rate dynamics and attentional control in healthy students. *Nonlinear Dynamics Psychol Life Sci, 17*, 345-360.

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention, 42*(3), 898-903.

Cain, B. (2007). *A review of the mental workload literature*. Defence Research and Development Toronto (Canada).

Casner, S. M., & Gore, B. F. (2010). Measuring and evaluating workload: A primer. *NASA Technical Memorandum, 216395*, 2010.

De Waard, D. (1996). *The measurement of drivers' mental workload*: Groningen University, Traffic Research Center Netherlands.

Elliott, C., Vijayakumar, V., Zink, W., & Hansen, R. (2007). National Instruments LabVIEW: A Programming Environment for Laboratory Automation and Measurement. *JALA: Journal of the Association for Laboratory Automation, 12*(1), 17-24.

Engler, J., & Schnell, T. (2012). Classifying Workload using Discrete Deterministic Nonlinear Models Across Subject Populations and for Extended Time. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*.

Engler, J. J. (2011). *Characterization of normality of chaotic systems including prediction and detection of anomalies*: The University of Iowa.

Etzel, J. A. (2006). *Algorithms and procedures to analyze physiological signals in psychophysiological research*. Iowa State University.

Fahrenberg, J., & Wientjes, C. (1999). Recording methods in applied environments. Engineering psychophysiology: issues and applications. RW Backs and W. Boucsein: Mahwah, NJ., Lawrence Erlbaum Associates, Inc.

Francesco, B., Maria Grazia, B., Emanuele, G., Valentina, F., Sara, C., Chiara, F., . . . Francesco, F. (2012). Linear and nonlinear heart rate variability indexes in clinical practice. *Computational and Mathematical Methods in Medicine, 2012*.

Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook*: CRC Press.

Geiselman, E. E. (1999). *Development of a non-distributed flight reference symbology for helmet-mounted display use during off-boresight viewing*.

Gibb, R., Ercoline, B., & Scharff, L. (2011). Spatial disorientation: decades of pilot fatalities. *Aviation, space, and environmental medicine, 82*(7), 717-724.

Glass, L. (2009). Introduction to controversial topics in nonlinear science: Is the normal heart rate chaotic? : AIP.

Goldberger, A. L., & West, B. J. (1987). Applications of nonlinear dynamics to clinical cardiology. *Annals of the New York Academy of Sciences, 504*(1), 195-213.

Govindan, R., Narayanan, K., & Gopinathan, M. (1998). On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 8*(2), 495-502.

Guyton, A. C. (2006). *Textbook of medical physiology* (Eleventh edition.. ed.). Philadelphia, PN: Philadelphia : Elsevier Saunders.

Hancock, P. (1989). The effect of performance failure and task demand on the perception of mental workload. *Applied ergonomics, 20*(3), 197-205.

Hart, S. G. (2006). *NASA-task load index (NASA-TLX); 20 years later.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.

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*(Supplement C), 31-43.

Henelius, A., Hirvonen, K., Holm, A., Korpela, J., & Muller, K. (2009). *Mental workload classification using heart rate metrics.* Paper presented at the Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology, 45*(1), 73-93.

Hsu, B.-W., Wang, M.-J. J., Chen, C.-Y., & Chen, F. (2015). Effective indices for monitoring mental workload while performing multiple tasks. *Perceptual and motor skills, 121*(1), 94-117.

Kraft, A. E., Russo, J., Krein, M., Russell, B., Casebeer, W., & Ziegler, M. (2017). *A systematic approach to developing near real-time performance predictions based on physiological measures.* Paper presented at the 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA).

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, 279-328.

Kuo, C.-D., & Chen, G.-Y. (1998). Heart rate variability standards. *Circulation, 98*(15), 1587b-1590.

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, Y.-H., & Liu, B.-S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, space, and environmental medicine, 74*(10), 1078-1084.

Longo, L. (2015). A defeasible reasoning framework for human mental workload representation and assessment. *Behaviour & Information Technology, 34*(8), 758-786.

Lyons, T. J., Ercoline, W., O'Toole, K., & Grayson, K. (2006). Aircraft and related factors in crashes involving spatial disorientation: 15 years of US Air Force data. *Aviation, space, and environmental medicine, 77*(7), 720-723.

Mansikka, H., Simola, P., Virtanen, K., Harris, D., & Oksama, L. (2016). Fighter pilots' heart rate, heart rate variation and performance during instrument approaches. *Ergonomics, 59*(10), 1344-1352.

Meshkati, N., Hancock, P., Rahimi, M., & M. Dawes, S. (1995). *Techniques in mental workload assessment*.

Owis, M. I., Abou-Zied, A. H., Youssef, A.-B., & Kadah, Y. M. (2002). Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification. *IEEE transactions on Biomedical Engineering, 49*(7), 733-736.

Reid, G. B., Potter, S. S., & Bressler, J. (1989). Subjective workload assessment technique (SWAT): A user's guide. *Wright Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory*.

Richter, M., & Schreiber, T. (1998). Phase space embedding of electrocardiograms. *Physical Review E, 58*(5), 6392.

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

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use*.

Sammer, G. (1998). Heart period variability and respiratory changes associated with physical and mental load: non-linear analysis. *Ergonomics, 41*(5), 746-755.

Sauvet, F., Jouanin, J. C., Langrume, C., Van Beers, P., Papelier, Y., & Dussault, C. (2009). Heart rate variability in novice pilots during and after a multi-leg cross-country flight. *Aviation, space, and environmental medicine, 80*(10), 862-869.

Scerbo, M. W., Freeman, F. G., Parasuraman, R., Di Nocero, F., & Prinzel, L. J. I. (2001). The efficacy of psychophysiological measures for implementing adaptive technology.

Schneider, F., Martin, J., Hapfelmeier, A., Jordan, D., Schneider, G., & Schulz, C. M. (2017). The validity of linear and non-linear heart rate metrics as workload indicators of emergency physicians. *PLoS One, 12*(11), e0188635.

Schnell, T., Reichlen, C. P., & Reuter, C. M. (2017). *Spatial Disorientation Threat Characterization for F-35 Representative Helmet-Mounted Display Use in the Flight*.

Schnell, T., Reuter, C., & Cover, M. (2017). *Visual dominance in pilots during recovery from upset.* Paper presented at the Digital Avionics Systems Conference (DASC), 2017 IEEE/AIAA 36th.

Schumacher, A. (2004). Linear and nonlinear approaches to the analysis of RR interval variability. *Biological research for nursing, 5*(3), 211-221.

Strang, A. J., Best, C., & Funke, G. J. (2014). *Heart Rate Correlates of Mental Workload in a Large-Scale Air-Combat Simulation Training Exercise.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Takens, F. (1981). Detecting strange attractors in turbulence *Dynamical systems and turbulence, Warwick 1980* (pp. 366-381): Springer.

Tarvainen, M. P., Georgiadis, S. D., Ranta-aho, P. O., & Karjalainen, P. A. (2006). Time-varying analysis of heart rate variability signals with a Kalman smoother algorithm. *Physiological measurement*, *27*(3), 225.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine, 113*(1), 210-220.

Task Force of the European Society of Cardiology (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *circulation*, *93*, 1043-1065.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of experimental psychology: Applied*, *2*(3), 270.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of human factors and ergonomics*, 243-273.

Vogt, J., Hagemann, T., & Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. *Journal of psychophysiology, 20*(4), 297-314.

Voss, A., Schulz, S., Schroeder, R., Baumert, M., & Caminal, P. (2009). Methods derived from nonlinear dynamics for analysing heart rate variability. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 367*(1887), 277-296.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics, 15*(2), 70-73.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman (Ed.), *Varieties of attention* (pp. 63-102). New York: Academic Press.

Wickens, C. D. (2002a). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177.

Wickens, C. D. (2002b). Situation awareness and workload in aviation. *Current directions in psychological science, 11*(4), 128-133.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors, 50*(3), 449-455.

Wickens, C. D. (2017). *Mental Workload: Assessment, Prediction and Consequences.* Paper presented at the International Symposium on Human Mental Workload: Models and Applications.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human factors, 35*(2), 263-281.

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological psychology, 34*(2-3), 163-178.

Wilson, G. F. (2002a). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18.

Wilson, G. F. (2002b). Psychophysiological test methods and procedures. *Handbook of human factors testing and evaluation,* 127-156.

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, H., & Benton, D. (2015). We should be using nonlinear indices when relating heart-rate dynamics to cognition and mood. *Scientific Reports, 5,* 16619. doi:10.1038/srep16619

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.

Zhang, Y., Zheng, H., Duan, Y., Meng, L., & Zhang, L. (2015). *An integrated approach to subjective measuring commercial aviation pilot workload.* Paper presented at the 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA).

Zhidong, Z., Yi, L., & Qing, L. (2011). Adaptive noise removal of ECG signal based on ensemble empirical mode decomposition *Adaptive filtering applications*: InTech.

Zhou, Y., He, J., Wang, Z., & Fu, S. (2014). *Study on the correlation between physiological and subjective workload measurements.* Paper presented at the 2014 International Conference on Information Science, Electronics and Electrical Engineering (ISEEE).