2013

Simulation-based Cognitive Workload Modeling And Evaluation Of Adaptive Automation Invoking And Revoking Strategies

Christina Rusnock
University of Central Florida

SIMULATION-BASED COGNITIVE WORKLOAD MODELING AND EVALUATION OF ADAPTIVE AUTOMATION INVOKING AND REVOKING STRATEGIES

by

CHRISTINA F. RUSNOCK
B.A. Economics-Government, Claremont McKenna College, 2004
M.S. Research and Development Management, Air Force Institute of Technology, 2008
M.S. Industrial Engineering, University of Central Florida, 2011

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2013

Major Professor: Christopher D. Geiger

ABSTRACT

In human-computer systems, such as supervisory control systems, large volumes of incoming and complex information can degrade overall system performance. Strategically integrating automation to offload tasks from the operator has been shown to increase not only human performance but also operator efficiency and safety. However, increased automation allows for increased task complexity, which can lead to high cognitive workload and degradation of situational awareness. Adaptive automation is one potential solution to resolve these issues, while maintaining the benefits of traditional automation. Adaptive automation occurs dynamically, with the quantity of automated tasks changing in real-time to meet performance or workload goals. While numerous studies evaluate the relative performance of manual and adaptive systems, little attention has focused on the implications of selecting particular invoking or revoking strategies for adaptive automation. Thus, evaluations of adaptive systems tend to focus on the relative performance among multiple systems rather than the relative performance within a system.

This study takes an intra-system approach specifically evaluating the relationship between cognitive workload and situational awareness that occurs when selecting a particular invoking-revoking strategy for an adaptive system. The case scenario is a human supervisory control situation that involves a system operator who receives and interprets intelligence outputs from multiple unmanned assets, and then identifies and reports potential threats and changes in the environment. In order to investigate this relationship between workload and situational awareness, discrete event simulation (DES) is used. DES is a standard technique in the analysis

of systems, and the advantage of using DES to explore this relationship is that it can represent a human-computer system as the state of the system evolves over time. Furthermore, and most importantly, a well-designed DES model can represent the human operators, the tasks to be performed, and the cognitive demands placed on the operators. In addition to evaluating the cognitive workload to situational awareness tradeoff, this research demonstrates that DES can quite effectively model and predict human cognitive workload, specifically for system evaluation.

This research finds that the predicted workload of the DES models highly correlates with well-established subjective measures and is more predictive of cognitive workload than numerous physiological measures. This research then uses the validated DES models to explore and predict the cognitive workload impacts of adaptive automation through various invoking and revoking strategies. The study provides insights into the workload-situational awareness tradeoffs that occur when selecting particular invoking and revoking strategies. First, in order to establish an appropriate target workload range, it is necessary to account for both performance goals and the portion of the workload-performance curve for the task in question. Second, establishing an invoking threshold may require a tradeoff between workload and situational awareness, which is influenced by the task's location on the workload-situational awareness continuum. Finally, this study finds that revoking strategies differ in their ability to achieve workload and situational awareness goals. For the case scenario examined, revoking strategies based on duration are best suited to improve workload, while revoking strategies based on revoking thresholds are better for maintaining situational awareness.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair, Dr. Christopher Geiger, whose enthusiasm for, and commitment to, this endeavor have served as a constant inspiration. He has been a mentor, friend, and partner on this academic journey, providing me with both positive encouragement and refreshingly honest feedback. He has spent countless hours involved in the details of this effort, while also helping me to maintain focus on the big picture. He had the good wisdom to set high standards early on, and has been a model example of academic excellence. Without his mentoring and guidance, this dissertation would not have been possible.

I would also like to thank my committee members, Dr. Waldemar Karwowski, Dr. Petros Xanthopoulos, and Dr. Lauren Reinerman-Jones for their support, encouragement, and feedback throughout the research project.

I would also like to express my appreciation to Dr. Stephanie Lackey and Dr. Lauren Reinerman-Jones of the Applied Cognition and Training in Immersive Virtual Environments Laboratory in the Institute for Simulation & Training at the University of Central Florida (UCF IST ACTIVE Lab) and to the U.S. Army Research, Development and Engineering Command (RDECOM) for providing the human-performance data (W91CRB08D00150068) for this research project.

I would also like to express my appreciation to the Systems Engineering and Management Department of the Air Force Institute of Technology (AFIT), especially Dr. Adedeji Badiru, for the sponsorship of my doctoral education.

Last, but certainly not least, I would like to thank my husband, Noah Rusnock, and my two kitties, Pixel and Orion, for their unwavering love and emotional support, and for giving me a reason to undergo this and all of life's adventures.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1. Background

In January 2012, United States Secretary of Defense Leon Panetta reveals the Pentagon's plan for reducing the United States military's budget by almost $500 billion within 10 years. Several programs are identified to be delayed or eliminated, including maritime vessels, submarines, fighter aircrafts, and ground combat vehicles. The military personnel budget is also considered for reduction, with approximately 100,000 troops slated to be cut over the next five years. The Pentagon plans to increase its inventory of unmanned vehicles in order to balance the reduction in troops and traditional weapons systems (AccuVal Associates, 2012; Fox News, 2012; Martinez, 2012). Unmanned vehicles (UVs) are used frequently in tactical- and theater-level missions including electronic attack missions, neutralization of enemy air defense, combat search and rescue and intelligence, surveillance, and reconnaissance (ISR) missions. Utilizing these complex automatic control systems not only saves lives by limiting warfighter exposure to threatening, harmful and deadly situations, but UVs can also act as force multipliers and augment the capabilities of the troops and traditional weapons systems, thanks to increased automation. Automation in UVs allows the unmanned systems to perform functions autonomously with reliability and precision, enabling a single UV human operator to accomplish more tasks simultaneously since it reduces human task load requirements. These advantages have permitted increased system complexity and allowed for an increased number of tasks to be allocated to the human operator. Therefore, the need for humans for supervisory control of UV systems has replaced the need for humans to directly control manual systems.

With the planned decreases in the size of the U.S. military forces and the simultaneous increases in the UV fleet, the Department of Defense is investigating the ability for a single human operator to remotely control multiple unmanned vehicles simultaneously (Dixon, Wickens, & Chang, 2004; McGrogan, Schneider, Wirthlin, Coloumbi, & Miller, 2011). This increased complexity, coupled with automation, can ultimately lead to high operator cognitive workload, while degrading the operator's skills and situational awareness (Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006). Other issues with integrating the interactions of human cognitive capability and automated computer decision-making include the appropriate levels of trust in automation, over-reliance by the human, reliability of the automation, and human complacency (Bailey & Scerbo, 2007; Lee & See, 2004; Sheridan & Parasuraman, 2006). Adaptive automation is one method that has been used to address these and other issues, while maintaining the benefits of traditional automation.

Unlike traditional automation, adaptive automation occurs dynamically, with the level of automation changing in real-time to meet performance or workload goals, triggered by real-time information regarding task performance, events, or operator states (Kaber, Perry, Segall, McClernon, & Prinzel, 2006). By automating tasks during periods of high cognitive workload, adaptive automation can increase performance and relieve the operator's cognitive workload while allowing the operator to maintain situational awareness. Furthermore, during times of low workload, adaptive systems enable the operator to take on more tasks; thus, these systems can potentially prevent operator complacency. Ideally, these systems would respond to the perceived or experienced cognitive workload of the individual user, and not simply the task load experienced by the system (for clarity, "workload" is used to describe the load experienced by the human, while task load is used to distinguish objective differences in quantities of tasks).

UV operations are particularly well-suited for adaptive automation because their workload fluctuates between short periods of high activity and long periods of low activity (Parasuraman, Cosenzo, & De Visser, 2009).

1.2. Current Applications for Adaptive Automation

Military, law enforcement, and national security organizations are interested in using automation to enhance human performance and safety. By keeping the human in the loop through dynamic task allocation, adaptive automation provides a unique opportunity to improve current system performance, as well as to expand into new endeavors including disaster recovery, maritime surveillance, and traffic management. Adaptive automation is particularly well-suited for these types of tasks because they entail uneven workload, critical safety aspects, and require human judgment.

For emergency response, especially those scenarios involving chemical, biological, radiological, nuclear, and high-yield explosive (CBRNE) events, it is often difficult and dangerous for humans to perform rescue, containment, decontamination and clean-up missions. Thus, robotic and unmanned vehicles are appropriate for disaster response and recovery by mitigating risk to personnel, assets and civilians keeping them out of harm's way. Recently, remotely-controlled robots have been used to collect environmental samples to monitor radiation levels and clear radioactive debris as part of the Fukushima recovery-effort (Greenemeier, 2011).

Unmanned vehicles are increasingly being used in maritime missions including border patrol, port security, submarine tracking, antipiracy, and protection of off-shore oil drilling platforms. Unmanned maritime missions use airborne and seaborne systems, and are utilized internationally from Israel to India. The U.S. military's inventory of underwater unmanned vehicles alone totals to 450 vehicles, and the U.S. Navy is currently replacing its manned P-3

Orion reconnaissance aircraft with unmanned aerials vehicles such as the MQ-4C Broad Area Maritime Surveillance UAV (Eshel, 2011; Martin, 2012). These unmanned surveillance missions would be appropriate in which to integrate adaptive automation technology, which would allow the human operator to control multiple unmanned vehicles simultaneously in order to cover a broader surveillance area, with the human intervening as necessary to make decisions upon potential target identification.

1.3. Research Gap

Most current research in adaptive automation focuses on proof of concept; thus, most studies seek to demonstrate the performance benefits of adaptive automation over user-initiated automation, static automation, or manual systems (e.g., Arciszewski, de Greef, & van Delft, 2009; Clamann, Wright, & Kaber, 2002; Cosenzo, Chen, Reinerman-Jones, Barnes, & Nicholson, 2010; Dorneich et al., 2006; Haarmann, Boucsein, & Schaefer, 2009; Kaber, Wright, Prinzel, & Clamann, 2005; Kaber, Wright, & Sheik-Nainer, 2006; Parasuraman, Cosenzo, & De Visser, 2009; Taylor, Reinerman-Jones, Cosenzo, & Nicholson, 2011; Wilson & Russell, 2007). Other studies explore design and implementation aspects such as the types of tasks to be automated, the levels of automation that should occur, or interface design. While there are numerous methods for invoking adaptive automation, little research exists regarding their relative effectiveness. However, numerous theoretical arguments favor the use of adaptive automation based on physiological measures, which use real-time, operator-specific cues, rather than other methods that provide cues based on time-lagged information or aggregate data. Based on the current literature, little attention has been paid to the dynamic revocation of automation. Revoking methods tend to use the same criteria as the invoking method or a method of convenience, with little consideration of the potential implications. If automation revocation occurs too soon, the human operator likely experiences increased stress and workload along with performance degradation. If automation revocation occurs too late, there could be negative impacts to situational awareness as well as boredom-induced performance degradation. The revoking method should be chosen carefully and account for a number of potential issues including the effects of rapid changes in automation level, the effects of task interruption, the differences between perceived workload and actual workload, and the differences between physiological measures and workload. Using the same threshold to invoke and revoke adaptive automation could lead to rapid hand-offs of tasks to and from the human, ultimately leading to increased operator confusion, stress and distrust. Unexpected hand-offs, either from task interruption or from disconnects between measured and perceived workload can also cause the operator to distrust, and thus reject, the system. To date, no study has focused on the relative effectiveness and impact of automation revoking methods.

1.4. Research Objectives

This research investigation seeks to address this gap in the current adaptive automation research by evaluating the effectiveness of automation revoking strategies. Specifically, this research seeks to capture the relative impacts of various automation revoking strategies on cognitive workload. This research also seeks to utilize a unique approach to explore automation revocation strategies – computer-based modeling and simulation. While modeling and simulation has been used extensively in other areas, this is the first study to model human behavior in order to simulate the performance of an adaptive system and automation revoking strategies. Due to the increase in the use of remotely-controlled unmanned vehicles and the numerous opportunities that UVs provide for adaptive automation, the relevant practical scenario focuses on intelligence, surveillance, and reconnaissance tasks using unmanned vehicles.

However, the findings of this research investigation are not only relevant to unmanned vehicles and ISR tasks, but they also have broader applicability to other adaptive systems.

1.5. Research Questions

The main research question to be addressed in this investigation is: Can simulation-based modeling of cognitive workload be used for evaluating adaptive automation invoking and revoking strategies? This research question is supported by five sub-questions:

Sub-Question 1: Can simulation modeling predict cognitive workload as well as established measures of cognitive workload? This question seeks to determine whether simulation can be used to compute a valid measurement of cognitive workload. The validity of the cognitive workload score should be tested against other known measures of workload, such as self-reported subjective workload scores and physiological measures.

Sub-Question 2: Can computer simulation modeling be used to evaluate system designs based on predicted cognitive workload? For this question, several system designs should be evaluated based on the simulation model's predicted cognitive workload scores in order to determine the feasibility of using the simulation's outputs for system evaluation.

Sub-Question 3: How can simulation modeling be used to determine the target level or range of cognitive workload scores for adaptive automation? The response to this question demonstrates how to identify a target workload level by comparing the cognitive workload scores produced by the computer simulation model with performance measures in order to establish the range of cognitive workload scores that correspond with peak performance. The answer to this question also requires an understanding of the relationship between workload and performance for this task (increasing, decreasing, parabolic).

Sub-Question 4: How can simulation modeling be used to determine a preferred invoking threshold for adaptive automation? The response to this question demonstrates how to identify the most appropriate threshold for invoking ("turning on") automation. Based on theoretical arguments in the literature, this research investigation assumes that the best method for invoking automation is to use a workload trigger (i.e., physiological measures/cognitive workload) and not task performance or critical events. Given that workload score is the preferred strategy for triggering adaptive automation, this study evaluates various thresholds to establish the preferred invoking threshold.

Sub-Question 5: How can simulation modeling be used to determine a preferred revoking strategy for adaptive automation? The response to this question demonstrates how to identify the preferred strategy for dynamically revoking ("turning off") the automation. Using the same criterion for revoking that is used for invoking will most likely not yield the best solution, since this criterion is likely to produce excessive hand-offs between the human operator and the unmanned system.

1.6. Expected Contributions of This Research Investigation

The research investigation makes three significant contributions to the body of knowledge. First, this research is the first study to evaluate the impact and relative effectiveness of adaptive automation invoking and revoking strategies. Second, this study is the first to use computer simulation modeling of cognitive workload for adaptive automation purposes. Finally, this research is the first to demonstrate that computer modeling and simulation of cognitive workload is as effective at predicting cognitive workload as well-established subjective measures, and is more predictive than most physiological measures.

The outcomes of this study include a demonstration of the utility of using computer simulation models for evaluating system designs; a methodology for selecting adaptive automation invoking thresholds, automation duration, and revoking thresholds; guidelines for performing workload and situational awareness tradeoffs; and a recommendation of a preferred revoking strategy for the particular system evaluated.

1.7. Organization of the Remainder of This Document

The remainder of this document is organized as follows. CHAPTER 2 presents a literature review, outlining relevant theories of cognitive workload, methods for measuring workload, methods of invoking adaptive automation, and the use of computer simulation to model cognitive workload. CHAPTER 3 provides a detailed discussion of the research methodology, and CHAPTER 4 discusses the specific practical human-supervisory control situation as well as the baseline simulation model of the situation. CHAPTER 5 provides the validation of the baseline model and demonstrates the use of discrete event simulation for system design evaluations. CHAPTER 6 expands on the baseline model by incorporating resource channel interference and details a workload-performance analysis in order to identify the target range for this workload score. CHAPTER 7 and CHAPTER 8, then, incorporate the adaptive automation aspects of this study and present the analysis and results for adaptive automation invoking thresholds and revoking strategies experiments, respectively. Finally, CHAPTER 9 concludes the study and discusses the future work of this research investigation.

CHAPTER 2
REVIEW OF PREVIOUS RELATED LITERATURE

2.1. Introduction

The primary goal of adaptive automation is to reduce the cognitive workload of the operator, in order to ensure peak performance and maintain system stability. This chapter begins by discussing how automation is classified in order to understand the degree or level to which a task can be automated and then the chapter reviews theories and empirical research regarding the types of tasks that can be automated, and the relationship between task type and level of automation. Following this is a discussion of the various methods for invoking adaptive automation, as well as an examination of the effectiveness of adaptive automation in reducing cognitive workload, improving task performance, and enhancing situational awareness.

The chapter then turns to an examination of cognitive workload—since this is the factor that adaptive automation is trying to impact. This discussion provides a brief overview of relevant cognitive workload theories and cognitive workload measurement tools. The chapter concludes with a discussion of human performance modeling, modeling cognitive workload using simulation, and potential future areas of research.

2.2. Classification of Automation

Automation is the use of mechanical or electronic devices to conduct tasks that were previously accomplished by humans. Automation has become ubiquitous in modern society, and is not only present in military systems, but also throughout a wide range of business and industrial endeavors, and even personal and home use (Sheridan & Parasuraman, 2006).

Whether it is irrigation in agriculture, robotic assemblies in manufacturing, spreadsheets in corporate finance, or a dishwasher at home, it is difficult to imagine an area of life that has not been improved through the use of automation.

As discussed in CHAPTER 1, automation enables increased performance, efficiency, and safety. However, it can also have negative impacts on cognitive workload, situational awareness, operator skillsets, over-reliance, and complacency. Adaptive automation provides a potential solution to these problems by adjusting the level of automation based on the operator or system states. Adaptive automation is also known as adaptive aiding, dynamic task allocation, dynamic function allocation, adaptive function allocation, and knowledge-based systems (Arciszewski et al., 2009; Clamann, Wright, & Kaber, 2002; De Visser et al., 2008; Inagaki, 2003a).

Adjusting the level of automation is more complex than simply having a task be fully automated or fully manual. Sheridan and Verplank (1978) outline ten levels of automation (LOAs). The lowest level, LOA 1, is when the human performs all tasks and no automation is present. In the highest level, LOA 10, all tasks are automated without human involvement. The eight other LOAs provide varying degrees of automation, encompassing a range of scenarios, from the system making recommendations for the human to execute to the system performing tasks with human approval or supervision. Table 1 summarizes the ten LOAs.

Table 1: Levels of Automation (adapted from Sheridan & Verplank, 1978).

	Determines Alternatives	Suggests Alternative	Selects Alternative	Executes Alternative	Informs of Action
Level 1	Human	Human	Human	Human	N/A
Level 2	Computer	Human	Human	Human	N/A
Level 3	Computer	Computer	Human	Human	N/A
Level 4	Computer	Computer	Computer, Human may or may not approve	Human	N/A
Level 5	Computer	Computer	Computer	Computer, if Human approves	N/A
Level 6	Computer	Computer	Computer	Computer, unless Human vetoes	N/A
Level 7	Computer	Computer	Computer	Computer	Always
Level 8	Computer	Computer	Computer	Computer	If Human requests
Level 9	Computer	Computer	Computer	Computer	If Computer decides to inform human
Level 10	Computer	Computer	Computer	Computer	N/A

While the LOAs of Sheridan and Verplank's (1978) are defined for traditional automation, these levels are equally relevant for adaptive automation. Designers of adaptive systems must decide which level(s) of automation exist for a given task. In some cases, the level of adaptive automation may be binary with the system switching back and forth between LOA 1 and another LOA. In other cases, the system may traverse along the full gamut of these LOAs. Not only is the best LOA likely to be different from operator to operator, but it can also be expected that, for a single operator, the best LOA varies depending on the specific context (e.g. task load, fatigue, environmental stressors, etc.). Arciszewski et al. (2009) simplify the LOAs of Sheridan and Verplank's into five levels: Manual (LOA 1), Advice (LOA 2-4), Consent (LOA 5), Veto (LOA 7), and System (LOA 10).

Parasuraman, Sheridan, and Wickens (2000) further extend Sheridan and Verplank's LOAs by adding a second dimension based on the four-stage model of information processing: sensory processing, perception/working memory, decision-making, and response selection. They propose that tasks can be categorized into the four corresponding stages: information acquisition, information analysis, decision-making, and action implementation. Thus, automated tasks can be described both in terms of the stage of information processing and the LOA (see Figure 1 for an example of how two systems can vary along these two dimensions) (Parasuraman, Sheridan, & Wickens, 2000).Thus, an automated system might have an LOA of 7 for information acquisition tasks and an LOA of 3 for information analysis tasks.

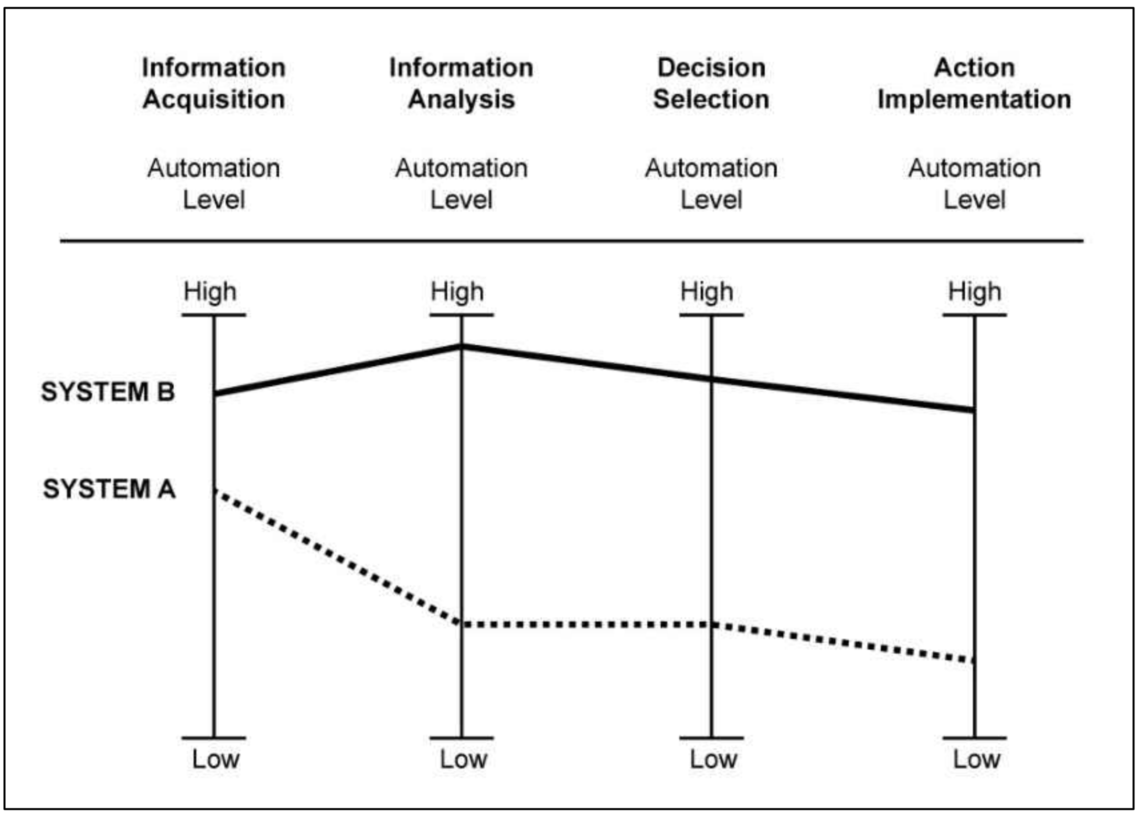

Figure 1: Levels of Automation and Stages of Information Processing (recreated from Parasuraman et al., 2000)

Each of these categories involves a unique set of activities and cognitive resources. The information acquisition and information analysis stages are highly-coupled; both involve using posterior regions of the brain. For information acquisition, the human engages in sensing the environment, processing data, and utilizing selective attention. A machine performing information acquisition conducts activities such as identification and classification of data, and presents that information to the user. For information analysis, the human manipulates and integrates information, allowing them to make judgments and predictions. An example of a machine engaged in information analysis includes highlighting pieces of information or providing forecasted information (Kaber et al., 2006; Parasuraman & Wickens, 2008).

The decision-making and action implementation stages are associated with activity in the frontal cortex. For humans, decision-making involves selection between two alternatives. A machine can perform this function by making a recommendation between two or more alternatives. Action implementation consists of executing the selection from the decision phase. For the human, this is often a psychomotor function, unlike the other stages which are more cognitive. For the machine, action implementation involves executing an action, which may have been decided upon by either the human or the machine (Kaber et al., 2006; Parasuraman & Wickens, 2008). Action implementation can be associated with early forms of automation, which sought to alleviate the human's physical workload.

2.3. Determining Which Tasks Should Be Automated

One of the first tasks in designing an adaptive system (or even a traditionally automated system) is to determine which tasks should be automated. Using the four stages of information processing, Clamann, Wright, and Kaber (2002), Kaber et al. (2005), and Kaber et al. (2006) found increased performance and situational awareness when using adaptive automation for information acquisition and action implementation tasks. These findings support the use of adaptive automation for lower-level information processing and psychomotor tasks. Conversely, using adaptive automation for information analysis and decision-making results in increased workload. The Clamann, Wright, & Kaber (2002) and Kaber et al. (2005) studies found that automating the information acquisition task relieves time pressure, allowing for additional time on secondary tasks; whereas automating the information analysis decreases time spent on secondary tasks (Clamann, Wright, & Kaber, 2002; Kaber et al., 2005). In the Kaber et al. (2006) study, the increase in workload for information analysis and decision making tasks is attributed to the visual implementation of the automation, which creates additional visual aids on the screens (Kaber et al., 2006). Thus, automation of higher level information-processing tasks poses the added design challenge of conveying that information without increasing visual clutter.

Parasuraman et al. (2000) provide a framework for selecting the appropriate level and type of automation based on a number of factors including mental workload, situational awareness, complacency, skill degradation, automation reliability, and costs. This framework is consistent with the work of Kaber et al. (2006), encouraging increased levels of automation for information acquisition and action implementation tasks and lower levels of automation for informational analysis and decision making tasks (Parasuraman et al., 2000).

Steinhauser, Pavlas, and Hancock (2009) recommend using adaptive aiding for those tasks that are most at risk of failure, which ensures that automation provides a safety net for critical tasks. On the other hand, Arciszewski et al. (2009) advocate using automation for less critical and repetitive tasks, while reserving higher-priority and more engaging tasks for the human. This allows the human to be an active participant for those tasks that can have significant consequences or require human judgment. Steinhauser et al. (2009) recommend that the system maintain control of the automation, with the operator having the opportunity to intervene in order to provide the operator with a sense of control. Arciszewski et al. (2009) take a different approach, proposing that the user play an active role in defining acceptable levels of automation to encourage user trust and buy-in. They propose treating the automated machine as a "virtual team member," with the user establishing "working agreements" that outline the acceptable lower and upper bounds for levels of automation based on task type. For example, benign categorization tasks might have permission to range from an LOA 3 to LOA 7, whereas a weapons firing task might range from LOA 1 to LOA 4 (Steinhauser, Pavlas, & Hancock, 2009).

2.4. Invoking Automation

By definition, adaptive systems do not maintain a constant level of automation. Changes in the level of automation can be triggered by either the user or by the system. Automation that is user-initiated is often referred to as "adaptable," whereas "adaptive" automation can be user- or system-initiated (Scerbo, 2001). Adaptable systems operate under the philosophy of Management by Consent (or Explicit Control), since it is the operator that selects, or at least approves, when and which automated actions occur. Adaptive systems, on the other hand, operate using Management by Exception (or Implicit Control), where the machine chooses and executes a specific course of action, with the operator intervening only when the operator disagrees (Liu, Wasson, & Vincenzi, 2009).

2.4.1. User-Initiated Automation

Adaptable, or user-initiated, systems are usually easier to implement and provide a number of additional advantages including providing the operator with an increased sense of control, increased user acceptance, reducing operator stress and mistrust, increased operator situational awareness, increased flexibility, and use of deliberate human judgment. Of course, there are also several disadvantages, when compared to adaptive systems, such as requiring greater operator involvement, higher task times, and greater operator workload (Liu et al., 2009).

One primary concern with adaptable automation is whether the user will choose to initiate automated modes, even when the user's workload is high. Tattersall and Fairclough (2003) report that, despite high workload, 25% of operators never initiate automated functions. This is primarily due to the operator's desire to maintain control of the task and the operator's lack of trust in automation. Another factor preventing an operator's use of automation is that the operator does not know when to change into an automated mode; that is, the operators do not have a sense of when workload, fatigue, or performance would indicate that they should switch to an automated mode (Tattersall & Fairclough, 2003).

In their study on adaptable systems, Bailey et al. (2006) also find that operators prefer manual control to automated control, with the average operator choosing the automated mode only 7% of the time. Those who choose to use automated control experience higher subjective workload and poorer task performance than those who remain mostly in manual mode. Bailey et al. (2006) attribute this increased workload and declining task performance to the additional task burden that operators experience while having to monitor their own performance and initiate the automation. This explanation is corroborated by the performance of the yoked control groups and adaptive automation groups, which had better task performance and lower subjective workload (Bailey et al., 2006).

A potential solution to operator hesitation to using automated modes is to design adaptable automation where the machine plays a more active role in the initiation process. Thus, in addition to the more common free choice automation, where users can change the LOA at any time, the machine could also be designed for prompted choice or forced choice automation. In prompted choice, the system periodically reminds users that they have the option to change the LOA. In forced choice, operators are periodically required to chance the LOA (either higher or lower) (Sauer, Kao, Wastell, & Nickel, 2011).

Sauer et al. (2011) find that operators use more automation under forced choice than prompted choice, suggesting that alternatives to free choice can help operators overcome uncertainty about when to use automation. Contrary to Tattersall and Fairclough's findings, Sauer et al. (2011) find that operators consistently choose to use more automation under high-stress conditions. Furthermore, the operators maintained the higher level of automation even after the stressful event had passed. This indicates that under clear workload pressure, operators tend to initiate adaptable automation systems, and that once they are using the automation, the operators readily overcome a need for control and mistrust of automation. Performance is slightly better for forced choice automation than the other two types, which the researchers attribute to the higher LOAs used by the forced choice group. Sauer et al. (2011) also suggest that forced choice reduces complacency, since the operator must think about their current level of automation and make a conscious decision to decrease or increase it. While Sauer et al. (2011) use 5-min intervals to trigger the prompted choice and forced choice notifications, further study is needed using prompted choice and forced choice based on context-specific criteria such as critical events or task load (Sauer et al., 2011).

Liu et al. (2009) also investigate differences in levels of automation for adaptable systems to determine if specific levels impact performance or workload. Liu et al. (2009) compare Management by Consent and Management by Exception tasks. In the Management by Consent task, the human manually accepts or rejects the system's selections (equivalent of Level 5 from

Table 1) before the system processes the items. In the Management by Exception task, the system automatically processes any of its selections that are not vetoed within 15 seconds (Level 6 from Table 1). Surprisingly, Liu et al. (2009) find no statistical difference in task performance or workload between the groups. While the Management by Exception group had longer processing times, Liu et al. (2009) attributes this to the 15 second delay in processing when the human chooses not to veto the computer's selection.

2.4.2. System-Initiated Automation

The majority of current research and development in adaptive automation focuses on system-initiated automation. There are five primary mechanisms for the system to trigger a change in automation: critical events, operator performance, performance models, physiological measures, or a combination of these methods (Parasuraman, Mouloua, & Molloy, 1996).

Critical event triggers are based on specific events, such as when the number of simultaneous tasks reaches a threshold or when a particular event occurs. A task analysis can be used to identify critical events; these events will be task steps that require the operator's immediate attention or decision-inputs when they arise (De Visser et al., 2008). Critical event triggers are typically based on the assumption that the occurrence of a particular event leads to high workload or a decrease in performance if automation is not present. While relatively easy to implement, this method fails to account for differences in individual operators or actual workload/performance (Inagaki, 2003b).

Operator performance triggers, such as an error rate, respond directly to operator performance. By being directly tied to the operator, these triggers account for operator differences. However, degradation in operator performance is associated with both high and low workload, and declining performance is also associated with the presence of automation (Young

& Stanton, 2002). Furthermore, performance measures are a lagging indicator; performance declines typically occur after the period of time that would have been best suited for automation (Inagaki, 2003b). This is especially likely to be the case for very short periods of intense activity.

Performance models make use of mathematical algorithms that predict operator performance or workload. Automation is thus triggered by predicted performance or workload, rather than actual performance or workload. These models may incorporate past operator or system performance, system states, critical events, and user actions (Bailey et al., 2006). Unlike other measures, performance models have the ability to predict when automation is going to be needed (Inagaki, 2003b). This provides the opportunity to invoke automation before there are significant impacts to task performance or operator workload. However, like critical event triggers, performance models fail to account for operator differences, unless they are based solely on that operator's performance.

Physiological assessments aim to provide real-time measurements of operator workload through the measurement of brain, cardiovascular, or eye activity. Adaptive automated systems using physiological measures maintain a target range of physiological arousal to achieve peak operator performance. These measures are responsive to the operator and the context, and thus are more precise than critical event or operator performance triggers (Bailey et al., 2006). However, physiological trigger thresholds are typically based on group-derived data, which is not as effective as individually derived thresholds (Wilson & Russell, 2007). A potential issue with physiological measures is the potential that the body may delay the physiological response. A delayed reaction could result in the adaptive automation triggering at a point later than desired. Another potential issue is that an increase in physiological arousal during more engaging activities may not coincide with a decrease in performance. In this instance, the adaptive automation may trigger unnecessarily. Historically, the strongest argument against the use of physiological measurers is their cost and size. While these systems have been expensive and cumbersome in the past, recent technological advances have made them more cost-accessible and relatively non-intrusive (Dorneich, Mathan, Ververs, & Whitlow, 2007; Gevins & Smith, 2003). Table 2 summarizes a selection of existing adaptive automation studies that consider the triggering method, performance metrics, and revoking method.

Table 2: Triggering Methods, Metrics, and Revoking Methods

Source	Critical Events/ Task Load	Operator Perfor-mance	Perfor-mance Models	Physio-logical Measures	SA	Perfor-mance	Work-load	Handoff – Return to Manual
Freeman et al., 2000	-	-	-	X	-	X	-	EEG Engagement Index
Clamann et al., 2002	-	X	-	-	-	X	-	Performance Based
Wilson & Russell, 2004	-	-	-	X	-	X	X	ANN – physiological state
Kaber et al. 2005	-	X	-	-	-	X	-	Performance Based
Bailey et al., 2006;	-	-	-	X	X	X	X	EEG Engagement Index - Workload
Dorneich et al., 2006;	-	-	-	X	X	X	X	User-initiated
Kaber et al. 2006	-	x	-	-	X	X	X	Performance based
Wilson & Russell, 2007	-	-	-	x	-	X	X	Remains on until completion of task
Parasuraman et al., 2009	-	x	-	-	X	X	X	Remains on until completion of task
Cosenzo et al., 2010	X	-	-	-	X	X	X	Task Difficulty
De Visser & Parasuraman, 2011	X	-	-	-	X	X	X	Task Load
Taylor et al., 2011	X	-	-	-	-	X	X	Task Load

2.4.3. Recovery and Hand-Off

In addition to determining how to invoke adaptive automation, it is also necessary to determine how or when the automated task returns to manual mode. To be effective, adaptive automation needs to sustain an optimal workload level, by automating tasks throughout the period of high-potential workload while returning tasks to the human once the workload returns to a manageable level (Arciszewski et al., 2009). If hand-off occurs too soon, the user likely experiences increased stress and workload along with performance degradation. If hand-off occurs too late, there could be negative impacts to situational awareness as well as boredom-induced performance degradation.

Studies have used a variety of approaches to hand off workload back to the user. Most rely upon the same methodology that is used to invoke the adaptive automation; thus, the systems return the tasks to the operator when a performance or physiological threshold is restored (Bailey et al., 2006; Clamann, Wright, & Kaber, 2002; Freeman, Mikulka, Scerbo, Prinzel, & Clouatre, 2000; Haarmann et al., 2009; Kaber et al., 2005; Kaber et al., 2006) or when the task load returns to a manageable level (Cosenzo et al., 2010; De Visser & Parasuraman, 2011; Taylor et al., 2011). However, several studies have chosen to keep the automation active once it is triggered (Parasuraman et al., 2009; Wilson & Russell, 2007) or require that deactivation be user-initiated (Dorneich et al., 2006). There does not appear to be any recent studies evaluating the relative effectiveness of these various hand-off mechanisms, nor the impact that these mechanisms may have on performance or situational awareness.

Systems that use task load to initiate hand-off of tasks from the machine to the human make the assumption that the human's perceived workload is reduced as quickly as the actual workload. Similarly, systems that use physiological measures assume that the human's physiological statistics recover at the same rate as their perceived workload (or even their actual workload). Rottger, Bali, & Manzey (2009) find discrepancies between perceived and measured workload, where subjective workload decreased as the level of automation increased, while objective workload, measured by heart rate and heart rate variability, did not change (Rottger, Bali, & Manzey, 2009). Hjortskov et al. (2004) also find conflicts in physiological measures, with heart rate readings recovering during 8-minute rest periods even though blood pressure readings do not recover (Hjortskov et al., 2004). If there is a disconnect between the adaptive system's measurements and perceived workload, the tasks may be handed back at inappropriate times. Inconsistencies with automation response and operator expectations can impact the system's perceived reliability. Reduced reliability and automation unpredictability can lead to a decrease in trust and user acceptance, as well as increase in operator stress (Dehais, Causse, Vachon, & Tremblay, 2012; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Funk & Miller, 2001).

An additional concern with using workload (either task load or physiological measures) as the basis for hand-off back to the human is the potential for the adaptive automation to execute detrimentally rapid changes in switching the automation on and off. One potential solution to this is to increase the length of time between measurements for comparison with the threshold. While increasing this time period is likely to result in fewer hand-offs, it also decreases the fidelity of the system (Freeman, Mikulka, Prinzel, & Scerbo, 1999). Another option is to use a threshold range with the upper and lower bounds serving as separate triggers. This method was successfully implemented in the adaptive automation study by Clamann, Wright, & Kaber (2002), which triggered adaptive automation to activate when performance was below one standard deviation of a pre-determined performance mean and triggered the adaptive automation to deactivate when performance was above one standard deviation of the mean.

Other potential solutions could include requiring a minimum period of time for automation or requiring a minimum period of time beyond the hand-off threshold before returning to manual mode. Ideally, the hand-off to the human should not result in the human's workload threshold being breached, thus triggering another round of adaptive automation.

2.5. Effectiveness of Adaptive Automation

In addition to understanding the types of tasks to be automated and the method of invoking adaptive automation, adaptive system designers also need to be aware of the expected effectiveness of adaptive systems. This section discusses the relationship between workload and performance, as well as current study findings of the impact of adaptive automation on workload, performance, and situational awareness.

2.5.1. Workload and Performance

Studies examining the impact of adaptive automation on cognitive workload have consistently found that adaptive automation results in lower workload than static automation, random automation, or manual modes (De Visser & Parasuraman, 2011; Dorneich et al., 2006; Parasuraman et al., 2009; Wilson & Russell, 2004). While most studies also find improvements in performance from adaptive automation over static automation, random automation, or manual modes (Cosenzo et al., 2010; Dorneich et al., 2006; Haarmann et al., 2009; Kaber et al., 2005; Parasuraman et al., 2009; Wilson & Russell, 2007), a few have found no statistical difference in performance between adaptive and non-adaptive systems (Arciszewski et al., 2009; Kaber et al., 2006; Taylor et al., 2011). Comparing performance during periods of manual control, Clamann, Wright, & Kaber (2002) find that operators perform better during manual periods of the adaptive automation than under completely manual control. Kaber et al. (2005) also find that action implementation adaptive automation performance under manual periods was better than during the completely manual mode. These finding suggests that adaptive automation not only improves performance while the automation is activated, but also while the automation is deactivated.

The degree of effectiveness for adaptive automation also depends on the nature of the task and the level of automation. Kaber et al. (2005), Kaber et al. (2006) and Clamann et al. (2002) find performance improvements for adaptive automation in information acquisition and action implementation tasks, but not for information analysis tasks. Furthermore, Kaber et al. (2006) find that using adaptive automation for information analysis and decision making tasks actually increases workload. Task complexity also factors into the effectiveness of adaptive automation, with adaptive automation providing performance improvements when task difficulty is high, but no difference between manual control or random automation when task difficulty is low (Wilson & Russell, 2004).

As for levels of automation, Jou, Yenn, and Yang (2011) find that as the level of automation increases, performance increases. However, these results contradict those of Rottger et al. (2009), which did not find any statistical difference in performance between levels of automation. Interestingly, Rottger et al, find that subjective workload is lower for automation that uses action implementation (LOA 6) than for automation using information analysis and decision selection (LOA 4); although, objective workload, as measured by heart rate and heart rate variability, is not statistically different for these different levels and types of automation (Rottger et al., 2009).

In terms of triggering the adaptive automation, there appears to be little difference between the various invoking methods. Most studies using physiological invoking methods have successfully demonstrated improvements in task performance (Dorneich et al., 2006; Haarmann et al., 2009; Wilson & Russell, 2007) and workload (Bailey et al., 2006; Dorneich et al., 2006). Despite criticism that critical event triggers fail to account for differences in individual operators or actual workload, studies using these methods have also been successful in demonstrating improvements in task performance (Cosenzo et al., 2010; De Visser & Parasuraman, 2011; Parasuraman et al., 2009) and workload (De Visser & Parasuraman, 2011; Parasuraman et al., 2009). However, these gains are most likely not as high as those based on real-time metrics from the individual operator. Using a combination of physiological measures to trigger adaptive aiding for a UAV task, Wilson and Russell (2007) find that adaptive-aiding based on the individual improves performance by 50%, whereas adaptive-aiding based on the mean of the sample group produces an improvement of 35%.

Furthermore, in their study on imperfect automation, De Visser and Parasuraman (2011) show that even under extreme degradation in reliability (as low as 30%), operators perform better with automation than without. This suggests that adaptive automation based on imperfect information or assumptions, including those systems that use critical events or models to invoke the automation, can still be expected to provide performance gains. On the other hand, Dixon, Wickens, and Chang (2004), Dixon and Wickens (2006), and Dixon, Wickens, and McCarley (2007) reveal that during times of high workload, imperfect automation with high reliability (80% reliable) results in better performance than no automation, but that low reliability (60-70% reliability) results in equivalent or worse performance than with no automation. Furthermore, the performance impacts to primary versus secondary tasks (and compliance versus reliance tasks) depend on whether the reliability issues are manifested as false-alarms or misses (failures to issue alerts). False alarms also have a greater impact on performance degradation and result in increased task time, as operators spend additional time verifying the automated alerts for accuracy (Dixon, Wickens, & Chang, 2004; Dixon & Wickens, 2006; Dixon, Wickens, & McCarley, 2007).

2.5.2. Situational Awareness

While adaptive automation benefits for workload and performance are promising, the benefits for situational awareness (awareness of one's environment) are less clear. While most studies find that adaptive automation provides situational awareness improvements over manual control, they also fail to show a significant difference between adaptive and static automation (Bailey et al., 2006; Cosenzo et al., 2010; De Visser & Parasuraman, 2011; Parasuraman et al., 2009).

The effect of adaptive automation on situational awareness may depend on the level of situational awareness. Endsley (1995) provides a three-level model of situational awareness consisting of perception, comprehension, and projection. In Level 1, the human perceives their current environment. In Level 2, situational awareness extends beyond perception to comprehensions of the elements in the human's surroundings. In Level 3, the human uses their perception and comprehension to project potential future occurrences. For primary tasks, using Parasuraman et al.'s (2000) four categories of information processing, Kaber et al. (2006) find that perception-based situational awareness improves with adaptive automation of information acquisition tasks, but that comprehension-based situational awareness is best when there is no automation (manual control).

Even though adaptive automation is expected to reduce cognitive workload and increase situational awareness, these measures do not always align. As Endsley (1993) discusses, workload and situational awareness can either converge or diverge. It is thus possible to perform at any point along the cognitive workload-situational awareness continuum shown in Figure 2.

In the ideal case, situational awareness is high and workload is low. However, in cases of overload, situational awareness can be low while workload is high. Furthermore, individuals can be challenged to withstand high amounts of cognitive workload while maintaining high situational awareness. Likewise, both situational awareness and workload can both be low, such as when performing long, tedious vigilance tasks.

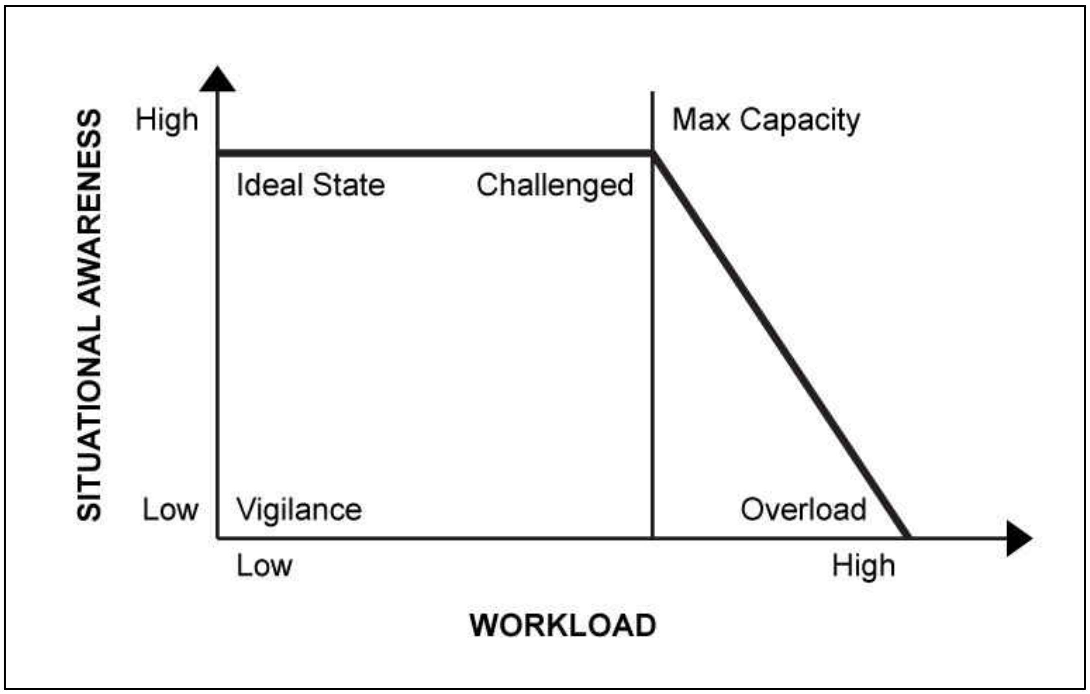

Figure 2: Workload – Situational Awareness Continuum (recreated from Endsley, 1993)

Adaptive automations effectiveness for situational awareness may also depend on the task being automated or the level of automation. Kaber and Endsley (2004) find that situational awareness is highest during intermediate levels of automation (Level 4, Table 1), while situational awareness is worse for lower and higher levels (Level 3 and Level 9). The researchers explain that this is most likely due to a balance between keeping the human involved in performing the task while freeing up cognitive resources (Kaber & Endsley, 2004). Kaber, Wright, and Sheik-Nainar (2006) find that using bi-modal auditory and visual cues has a greater positive impact on situational awareness than using just one of these modes. However, Dorneich et al. (2006) find that while subjects are able to maintain a high level of situational awareness using adaptive technologies, situational awareness of lower priority information suffers.

While current assessments of situational awareness appear inconclusive, this may be due to the studies implementation of the adaptive automation and the situational awareness task. For the most part, the adaptive automation takes over the primary task while providing little to no support for the situational awareness task. Furthermore, the situational awareness task typically requires the subjects to recall gauges or other information that are not used in executing the primary task. It can be inferred that the operators relegated the situational awareness tasks to a lower-priority secondary or even tertiary task. Thus, the design of the studies likely leads to performance improvements of the primary task, while adaptive automation does little to aid operators in their situational awareness tasks.

2.6. Theories of Cognitive Workload

In order to design an effective adaptive system, that will reduce cognitive workload and enhance situational awareness and performance, it is necessary to understand how the mind processes information and experiences cognitive workload. Figure 3 provides a basic model of information processing; in this model, information processing begins with stimuli from the environment. These stimuli are perceived through sensory receptors, which are then perceived by the human brain. Depending on the nature of the information, cognitive processing may occur, integrating working memory and/or long-term memory. Next, the human decides on an appropriate response, and then executes the response, ultimately proving feedback to the environment (Wickens & Hollands, 2000).

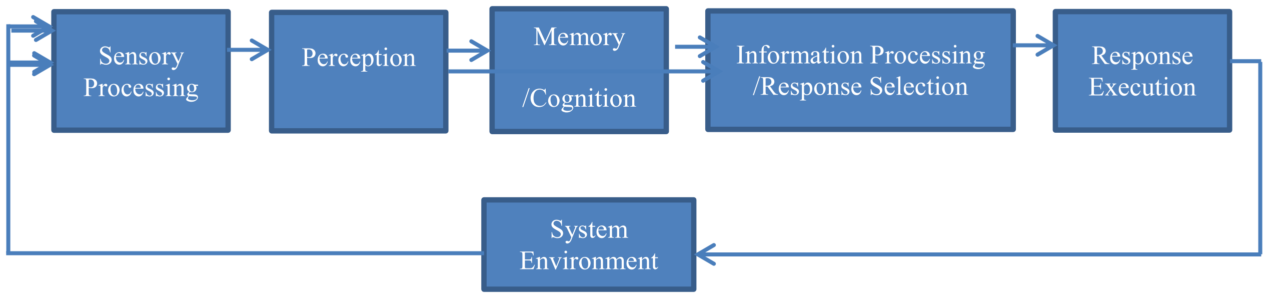

Figure 3: Model of Human Information Processing (adapted from Wickens and Holland, 2000)

Although this model provides an explanation of the flow or sequencing or information processing, it does not reveal how multiple stimuli are processed. Theories of how the mind handles multiple stimuli include Structural Bottleneck Theories (Kahneman, 1973), Capacity Theory (Kahneman, 1973; Norman & Bobrow, 1975), and Multiple Resource Theories. This research leverages off of multiple resource theory, since this is currently the predominant view. Multiple resource theories presume that man has limited mental resources. However, unlike capacity theory, multiple resource theories do not combine the total mental resources into a single pool. Instead, man's cognitive resources are divided into multiple channels.

Wickens' multiple resource theory provides a framework for differentiating types of workload along different channels. According to Wickens, an operator is able to more easily manage two tasks that utilize different channels versus two tasks that utilize the same channels. This theory describes channels that occur for the three main phases of information processing: sensory perception, information processing, and response. The multiple resource theory separates sensory perception channels into auditory or visual (and the visual channel is further divided into focal/detail and ambient/peripheral), processing channels into verbal or spatial, and response channels into vocal or manual (Wickens, 1984; Wickens & Hollands, 2000).

McCracken and Aldrich (1984) describe workload as being composed of three variables: sensory (perceiving visual or auditory stimuli), cognitive (mental processing), and psychomotor (behavioral responses). This model consists of four workload components: Visual, Auditory, Cognitive, and Psychomotor (VACP), each of which can be evaluated based on task complexity. Bierbaum, Szabo, and Aldrich (1989) adapt McCracken and Aldrich's original scale, by using pairwise comparisons made by human factors experts to convert the original ordinal scale into an interval scale. Table 3 provides this adapted scale. For concurrent tasks, values of the same component are summed, with a value of "8" representing overload for that component. (Note that the highest value for any specific task is "7", thus a single task will not produce an overload condition). Bierbaum et al. (1989) do not provide a mechanism for determining a composite workload score. Both Keller (2002) and Laux and Plott (2007) suggest that the workload component scores can be summed to obtain an overall workload score, which would provide a maximum score of 28 for a single task (Keller, 2002; Laux & Plott, 2007). However, there does not appear to be any empirical evidence or theoretical explanation as to why these scores would be additive, nor to whether these components should be given equal weight when creating a combined overall score.

The VACP model has several limitations including its assumption that all workload channels are completely independent without any interference with each other and it workload scoring fails to account for experience or personal stressors, but rather assume that all operators will experience the same level of workload for the same task. Although VACP has several limitations, it has been found to be comparable to other workload scores such as the Workload Index (WINDEX) and Time-Line Analysis and Prediction (TLAP), as well as fairly predictive of workload conditions (Sarno & Wickens, 1992).

Table 3: Revised VACP Values and Descriptions (recreated from Bierbaum, Szabo, & Aldrich, 1989)

Value	Descriptors
	VISUAL
0.0	No Visual Activity
1.0	Visually Register/Detect (detect occurrence of image)
3.7	Visually Discriminate (detect visual difference)
4.0	Visually Inspect/Check (discrete inspection/static condition)
5.0	Visually Locate/Align (selective orientation)
5.4	Visually Track/Follow (maintain orientation)
5.9	Visually Read (symbol)
7.0	Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions)
	AUDITORY
0.0	No Auditory Activity
1.0	Detect/Register Sound (detect occurrence of sound).
2.0	Orient to Sound (general orientation/attention)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect occurrence of anticipated sound)
4.9	Interpret Semantic Content (speech)
6.6	Discriminate Sound Characteristics (detect auditory differences)
7.0	Interpret Sound Patterns (pulse rates, etc.)
	COGNITIVE
0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
3.7	Sign/Signal Recognition
4.6	Evaluation/Judgment (consider single aspect)
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion
	PSYCHOMOTOR
0.0	No Psychomotor Activity
1.0	Speech
2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight controls, sensor control)
4.6	Manipulative
5.8	Discrete Adjustive (rotary, vertical thumbwheel, level position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)

One of the limitations of VACP is that it does not account for interference across multiple channels. Performing tasks that require multiple channels can produce different levels of conflict, depending on the channel. For example, auditory and psychomotor tasks are relatively non-competing, thus there is relatively little conflict with walking and listening at the same time. However, performing visual and audio tasks, especially in both involve verbal data, can have a high degree of conflict. For example, there is a high conflict with reading a book and listening to a lecture at the same time. Table 4 and Table 5 display examples of conflict matrices that have been developed to capture the relative conflict across resource channels.

Table 4: Workload Conflict Matrix (recreated from North & Riley, 1989)

		Task "B" Resources			
	Response	**Visual**	**Auditory**	**Manual**	**Verbal**
Task "A" Resources	**Visual**	HIGH CONFLICT (.7-.9) Directly competing resources (e.g. tow search tasks; less if tasks are adjacent or on same display areas).			
	Auditory	LOW CONFLICT (.2-.4) Noncompeting resources (e.g. search and listening).	HIGH CONFLICT (.7-.9) Highly competitive resources; some time-sharing if discriminability between inputs is high.		
	Manual	LOW CONFLICT (.1-.3) Noncompeting resources.	LOW CONFLICT (.1-.3) Noncompeting resources.	HIGH CONFLICT (.7-.9) competitive resources such as two tracking tasks or discrete choice tasks have shown high dual-task decrements.	
	Verbal	LOW CONFLICT (.1-.3) Noncompeting resources.	MEDIUM CONFLICT (.4-.6) More interfering if task requires voiced output.	LOW CONFLICT (.2-.4) Nonoverlapping resources showing little dual-task decrement in studies of tracking and voice input.	HIGH CONFLICT (1.0) Requires complete serial output (e.g. giving two messages or voice commands).

Table 5: Workload Conflict Matrix 2 (recreated from Wickens, 2002)

		Perceptual				Cognitive		Response	
		Visual	Visual	Auditory	Auditory	Cognitive	Cognitive	Response	Response
		Spatial	Verbal	Spatial	Verbal	Spatial	Verbal	Spatial	Verbal
Visual	Spatial	0.8	0.6	0.6	0.4	0.7	0.5	0.4	0.2
Visual	Verbal		0.8	0.4	0.6	0.5	0.7	0.2	0.4
Auditory	Spatial			0.8	0.4	0.7	0.5	0.4	0.2
Auditory	Verbal				0.8	0.5	0.7	0.2	0.4
Cognitive	Spatial					0.8	0.6	0.6	0.4
Cognitive	Verbal						0.8	0.4	0.6
Response	Spatial							0.8	0.6
Response	Verbal								1.0

2.7. Measuring Workload

For system-initiated adaptive automation to be effective, the system must be able to make an accurate assessment of the operator's workload. Workload is a function of both human capacity and task demands. It is important to note that both of these vary, depending on the person and the context. In terms of mental capacity, not only are there differences from person to person, but capacity can also differ for a specific person depending on training, experience, fatigue, and other emotional factors (Arciszewski et al., 2009; Kantowitz, 1987). As for task demands, workload can be affected by both quantity and complexity; although, there is some evidence that complexity has more of an impact on workload than quantity (De Visser &

Parasuraman, 2011). Thus, to be effective, measures of cognitive workload need to be responsive to changes in task and personal factors.

2.7.1. Measurement Considerations

Eggemeier, Wilson, Kramer, and Damos (1991) provide six considerations for evaluating workload measurement tools: sensitivity, diagnosticity, intrusiveness, reliability, implementation requirements, and operator acceptance (Eggmeier, Wilson, Kramer, & Damos, 1991). Sensitivity and diagnosticity relate to task-aspects of the measurement tool. Sensitivity is the ability of the measurement tool to identify changes in workload levels. For adaptive automation purposes, sensitivity to workload changes will directly impact the sensitivity of the invoking mechanism. In addition to being sensitive to changes in workload, diagnosticity captures the ability to differentiate between types of workload described by Wicken's multiple resource theory. Diagnosticity is important for adaptive systems because the type of workload may dictate which tasks are to be automated or the permissible level of automation.

While sensitivity and diagnosticity focus on the task-measurement aspects of the tool, intrusiveness and operator acceptance focus on user aspects of the tool. Intrusiveness captures the degree to which the measurement device interferes with the operator's ability to perform the task. For example, if an adaptive system used a galvanic skin response (GSR) system attached the operator's arm for invoking automation, the equipment should not impact the operator's motor-skills in executing the task. Operator acceptance, on the other hand is the degree to which the operator is willing to use the tool. Operator acceptance can be impacted by personal issues such as trust and self-confidence or physical issues such as comfort and intrusiveness. Furthermore, reliability and implementation requirements also need to be considered. Reliability is the degree to which the tool produces consistent measurement of workload; implementation requirements include software, hardware, training, and time (Eggmeier et al., 1991). In order for the adaptive system to be implemented and effective, the system must be reliable and the implementation requirements must be feasible and practical.

2.7.2. Subjective Measures

Over the past 50 years, researchers have used a variety of subjective measurement tools to measure cognitive workload. These subjective measures involve asking the subject their rate their mental effort across a number of spectrums. Due to the nature of these tools, the measurements are typically conducted just after the conclusion of the task. Thus, the measurements do not provide real-time feedback on the subject's mental workload, nor do they capture changes in workload over the course of the task. Furthermore, since there is just a single post-treatment observation, these measures do not provide a baseline measurement, thus there is no means of comparing differences in workload. The most widely used subjective measurement tools include the NASA-TLX (Hart & Staveland, 1988; Hart, 2006), SWAT (Luximon & Goonetilleke, 2001; Reid & Nygren, 1988), Cooper-Harper (Cooper & Harper, 1969; Cooper & Harper, 1997), MRQ (Boles & Adair, 2001a; Boles & Adair, 2001b), Bedford Scale (Roscoe & Ellis, 1990), Overall Workload (Jung, 2001), Workload Profile (Rubio, Diaz, Martin, & Puente, 2004; Tsang & Velazquez, 1996), and the Integrated Workload Scale (Pickup, Wilson, Norris, Mitchell, & Morrisroe, 2005).

Subjective workload tools have been applied to such varied areas as agriculture (Dey & Mann, 2010), road vehicle driving (Baldauf, Burgard, & Wittmann, 2009; Or & Duffy, 2005), memory tasks (Adams & Biers, 2000; Biers & Anthony, 2000), telecommunications (Whitaker, Hohne, & Birkmire-Peters, 1997), driving while talking on a cell phone (Waugh et al., 2000), military ground operations (Dorneich et al., 2007), visual pursuit tracking (Ellis, Dorighi, Menges, Adelstein, & Jacoby, 1997; Park & Park, 2007), lunar habitat hatch size (Thompson, Litaker, Archer, & Howard, 2008), head mounted displays (Litaker, Thompson, & Archer, 2008), differences in personality types (Cui, Sun, & Yu, 2008), speech intelligibility (Urquhart, 2003), virtual environments (Ng & So, 2003), unmanned vehicle operations (Pepe et al., 2008), video games (Boles, Phillips, Perdelwitz, & Bursk, 2004), vigilance tasks (Finomore, 2006), dual-task performance (Boles, Bursk, Phillips, & Perdelwitz, 2007; Phillips & Boles, 2004), simulated robotic-assisted surgery (Klein et al., 2008), and medical clinical work (Horner et al., 2011).

Reviewing the six measurement considerations of Eggemeier et al. (1991), subjective measures perform well in terms of implementation requirements and operator acceptance because they are frequently simple pen and paper (or mouse and monitor) tools that require minimal training or equipment. Tools such as NASA-TLX and MRQ that are well-vetted and capture different aspects of workload also rate fairly well in terms of reliability and diagnosticity, respectively.

Subjective measures tend to be sensitive to changes in workload when comparing two separate tasks using the same operator. However, they are not able to effectively capture changes in workload that occur during the course of performing a single task. Sensitivity can be traded-off with intrusiveness, by gathering data continuously during the performance of the task; however, the measurement is likely to interfere with the task. Querying a subject for their perceived workload requires auditory or visual perception, consciously determining one's own workload requires cognitive resources, and expressing that response utilizes auditory, visual, and/or psychomotor resources. Thus, subjective measurement tools become secondary tasks themselves. Thus, studies typically avoid interference by obtaining this information after the task is performed. In which case, the workload data will not be continuous, will not be real-time, may be unable to discern changes in workload throughout the task, and may be subject to memory distortions. Measuring workload by directly measuring the individual's behavior or performance alleviates some of these issues.

2.7.3. Behavioral/Performance Measures

Behavior or Performance Measures of cognitive workload monitor task performance such as error frequency, number of errors, response time, and speed to determine the level of cognitive workload. Direct measurement of operator performance can be conducted by the system in a way that is transparent to the user, and it can be conducted continuously throughout the performance of the task. Thus, behavioral measures are not intrusive and are more sensitive to changes in workload than subjective measures. Behavioral measures do have some implementation requirements since real-time, automated data collection will require the system to be modified to collect the required data. These implementation requirements can be traded-off with sensitivity by just collecting cumulative performance statistics at the end of the trial, rather than real-time data.

The main weaknesses for behavioral measures are diagnosticity and reliability. These measures have low diagnosticity, since collecting performance data such as speed or error rate will not provide any indication as to the type of workload the operator is experiencing. Similarly, reliability is an issue because there is not a direct linear relationship between performance and workload. Instead, both high and low workload are associated with poor performance. Low workload contributes to boredom and distraction, especially in sustained-attention scenarios where very little change occurs (baggage screening, etc.). Low workload scenarios prevent the operator from maintaining their full attention on the task, eventually resulting in errors. High workload tasks place undue pressure on limited cognitive resources. Overtaxing the operator's limited resources will also result in declining performance. However, when workload is maintained at a manageable level, the operator is able to achieve their best performance (Young & Stanton, 2002). By altering the human's workload, automated systems impact system performance in a similar manner. Thus, too much automation can be expected to result in under-load for the operator and declining system performance. Similarly, too little automation can be expected to result in over-load for the operator and, thus, also resulting in declining system performance.

Based on a series of experimental studies of mental workload and performance, Cassenti and Kelley (2006) describe a segmented workload-performance curve consisting of four segments: undertax, ceiling performance, steady decline, and floor performance. This curve, depicted in Figure 4, reveals that, subjects experienced a slight dampening of their performance at low levels of workload, peak performance at moderate levels of workload, steadily declining performance at moderately-high levels of workload, and then minimum performance levels at high workload (Cassenti, Kelley, & Carlson, 2010; Cassenti & Kelley, 2006; Cassenti, Kelley, Colle, & McGregor, 2011). This curve demonstrates the uneven tradeoff between workload and performance, revealing that high levels of workload impact performance more severely than low levels of workload.

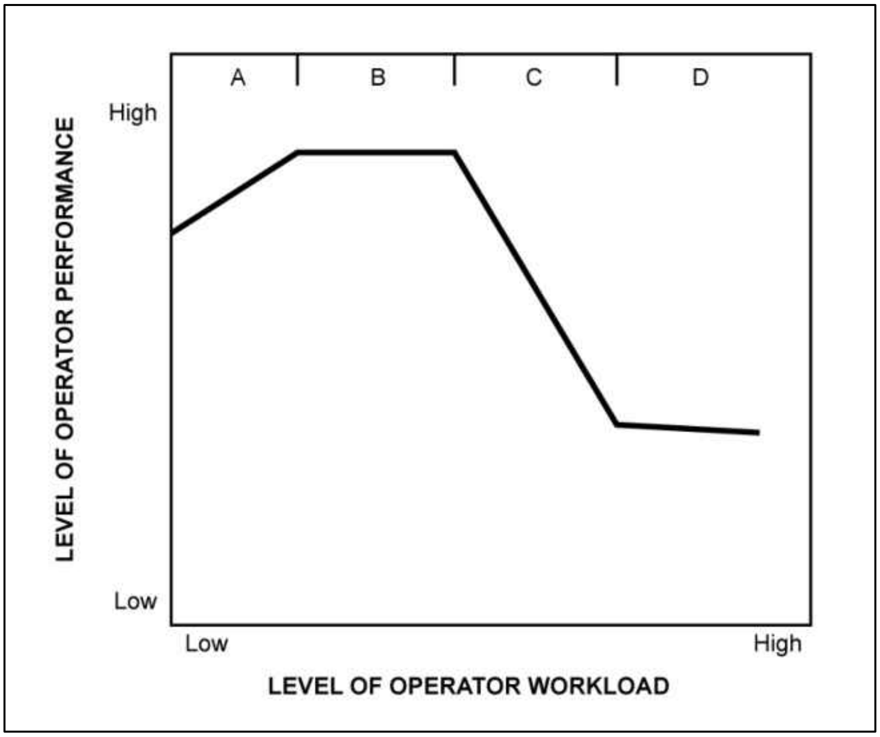

Figure 4: Workload-Performance Curve (recreated from Cassenti et al., 2011)

This uneven tradeoff between workload and performance erodes both the reliability and the sensitivity of behavioral measures. Furthermore, despite behavioral measures ability to monitor real-time changes in performance, these changes are likely a lagging indicator of actual workload. That is, an operator may experience excessively high workload for a considerable period of time before this becomes apparent in their performance. This lagging indicator tendency is especially detrimental to adaptive systems, because the intent is to prevent performance degradation, thus behavioral measures provide information too late. These issues, combined with technological advances, have led adaptive systems to turn to physiological measures for assessing operator workload.

2.7.4. Physiological Measures

Physiological measures use biological feedback to estimate cognitive workload. Common physiological measures include heart rate, heart rate variability, respiration, skin response, pupil dilation, eye movement/fixation, blink rate, and brain activity. Physiological measures are particularly well suited for adaptive automation purposes because they provide immediate feedback, are highly sensitive to change, and can be designed for minimal to no intrusiveness in primary task performance. Further details about the most common physiological measures are provided below.

Table 6 displays the relative performance for each type of workload tool (subjective, behavioral, and physiological) in terms of Eggemeier et al.'s (1991) six measurement considerations, plus two additional considerations not discussed by Eggemeier et al.: validity and timeliness. Validity is the degree to which the tool measures what it is supposed to measure (i.e. workload), and timeliness captures the proximity in time between the workload event and the collection of the data. A quick review of this table easily reveals that each measurement tool has its strengths and weaknesses and that no particular type of tool dominates another in all categories. For this reason, most workload studies choose to use multiple tools. In time, physiological measures are likely to become the tool of choice; for, as technology progresses, these measurement tools will improve in all of the measurement consideration categories. Subjective and behavioral measures, on the other hand, are not likely to experience any drastic improvements.

Table 6: Measurement Considerations

Measurement Consideration	Subjective Measure: NASA-TLX	Subjective Measure: MRQ	Behavioral/ Performance Measures	Physiological Measures
Sensitivity (with-in task)	**Poor**	**Poor**	**Moderate**	**Good**
Diagnosticity	**Poor**	**Good**	**Poor**	**Moderate**
Intrusiveness	**Good**	**Good**	**Good**	**Moderate**
Reliability	**Moderate**	**Moderate**	**Good**	**Good**
Implementation Requirements	**Good**	**Good**	**Moderate**	**Poor**
Operator Acceptance	**Good**	**Good**	**Good**	**Moderate**
Validity	**Moderate**	**Moderate**	**Poor**	**Moderate**
Timeliness	**Poor**	**Poor**	**Moderate**	**Good**

One of the most common physiological measures of cognitive workload is the measurement of brain activity. This can be accomplished through the use of electroencephalography (EEG), functional magnetic resonance imaging (fRMI), or transcranial Doppler (TCD) sonography. Researchers have successful correlated EEG, fMRI, and TCD results with different resource channels, thus enabling the tracking of specific types of tasks. These measures have been used to differentiate between various tasks language, spatial, problem solving (Just, Carpenter, & Miyaki, 2003), between high task load and resting states (Dussault, Jouanin, & Guezennec, 2004; Taylor et al., 2011), as well as to differentiate between memory, attention, reaction, and alertness (Dussault et al., 2004; Ryu & Myung, 2005; Warm & Parasuraman, 2007; Wilson, Caldwell, & Russell, 2007).

In addition to brain activity, eye activity has also been proposed for predicting mental workload. Eye activity includes pupil dilation, eye movements, and blink rate. While eye activity has been shown to be a successful predictor of cognitive workload these measures are affected by physical, psychological, and environmental factors, which can lead to conflicting interpretations (David, Mollard, Cabon, & Farbos, 2000; Di Nocera, Camilli, & Terenzi, 2006;

Naatanen, 1992; Recarte & Nunes, 2003; Recarte, Pérez, Conchillo, & Nunes, 2008; Taylor et al., 2011).

As cognitive workload increases, the body and mind experience increased arousal and stress. Thus, measures of cardiovascular activity such as heart rate, heart rate variability, and blood pressure are also potential indicators of change in cognitive workload (Hjortskov et al., 2004; Wilson, 1992). Studies have had mixed success with using heart rate and heart rate variability to predict cognitive workload. While some studies find clear results of these measures tracking as expected with changes in workload (Dussault et al., 2004; Hjortskov et al., 2004), others find no difference or even contradictory results with the measure tracking in the opposite direction (Rottger et al., 2009; Taylor et al., 2011; Wilson et al., 2007).

A number of studies have combined physiological measures to produce results more successful than those found with a single measure. Successful combinations include: heart rate variability and electro-dermal response (Haarmann et al., 2009), alpha band, blink rate and heart rate variability (Ryu & Myung, 2005), and multi-band EEG, eye movement, and heart rate (Wilson & Russell, 2004; Wilson & Russell, 2007).

2.8. Modeling Cognitive Workload Using Simulation

Modeling cognitive workload is still in its infancy, thus most research in the field focuses on creating, modifying, and validating workload models and simulation tools. Unfortunately, this means that few studies have used these models and tools to conduct research experiments. Over the past decade, the discrete event simulation software produced by Micro Analysis and Design (acquired by Alion Science and Technology Corporation) and the U.S. Army Research Laboratory has reached a state of maturity that it is now starting to be used by researchers and practitioners. This discrete event simulation is a natural fit for human factors researchers accustomed to conducting Task Analyses because it centers on Task Networking Modeling. Task Network models decomposes human performance into a series of functions that are further decomposed into tasks. Interfaces, resource channels, and workload values can then be assigned directly to these lowest level tasks. Benefits of task network models include their ability to interface with system models, their relative low cost and time effectiveness when compared with human trials, and their ability to be used in a discrete event simulation to answer performance questions (Laughery, 1999). The current version of ARL and Alion Science and Technology Corporation's discrete event simulation tool is the Improved Performance Research Integration Tool (IMPRINT). This tool is designed to predict the effects of human performance and system performance due to changes in crew size, technology, job assignments, environmental factors, and personal protective equipment, as well as monitor the outputs of user-defined variables and stressors (Allender, 2000b; Wojciechowski, 2007).

The predominance of cognitive workload research using IMPRINT has been to determine manpower requirements. These studies have successfully evaluated task performance, group workload, and individual workload in order to establish the recommended crew size or operator to system ratio. Perhaps due to the accessibility of IMPRINT through ARL, all of the studies have a military focus, such as crew requirements for operating UAVs (Hunn & Heuckeroth, 2006; McGrogan et al., 2011), UGVs (Wojcik, 2002) , manned ground vehicles (Mitchell & McDowell, 2008; Mitchell, 2008), a U.S. Navy Destroyer bridge (Allender, 2000a), and a communications center (Allender, 2000b).

Besides manpower requirements, IMPRINT has also been used to evaluate cognitive workload differences between human-robot teams and human-human teams (Harriott, Zhang, & Adams, 2011), performance enhancements and degradation from personal protective equipment

(Allender, 2000a; Allender, 2000b), performance impacts due to environmental stressors (Wojciechowski, 2007), and workload issues arising from interface design (Wojcik, 2002), duty assignments, and work/rest schedules (Allender, 2000b; Mitchell, 2008).

2.9. Future Work in Adaptive Automation

Adaptive automation is still a relatively new area of research and there are many opportunities for future work. While not within the scope of this research investigation, much of the current body of knowledge is focused on the physical implementation and design of adaptive systems, especially those based upon physiological measures. Despite this focus, the literature show no clear performance, workload, or situational awareness benefits for choosing physiological measures over other system-based invoking methods. While a number of studies show positive results from critical event and performance-based invoking methods, the review of the current literature review finds no recent studies using performance models. This is surprising considering the widespread use of modeling and simulation in other areas.

Another area requiring further study is the difference between user and system-initiated automation. With the exception of Bailey et al. (2006), studies exclusively focus on a single type of invoking method, without attempting to empirically validate that one method provides improved performance, workload, or situational awareness benefits over the other. Furthermore, while attention has focused on how to best invoke adaptive automation, there is scant research on recovery and hand-off of tasks to the human.

Adaptive automation research would also benefit from additional investigation into the types of tasks that are adaptively automated and the levels of automation. In addition to further studying which types of information processing tasks benefit from adaptive automation, studies should also examine how performance and situational awareness are affected when these task types are primary versus secondary tasks.

2.10. Summary

Research in the area of adaptive automation has made significant progress since the term "adaptive aiding" was first coined by W. B. Rouse in 1988. The past decade, especially, has seen substantial advances as technological innovation has changed implementation of adaptive automation from a theoretical academic concept to an achievable reality. Continued miniaturization and cost decreases in computerized technologies will eventually allow for widespread use of adaptive automation. However, in order to reach widespread implementation, researchers must first resolve the issues discussed in this chapter.

This review reveals that a major limitation of the current adaptive automation research is the failure to evaluate the relative effectiveness of adaptive automation revoking methods. Important aspects of revoking that need to be addressed include timing, duration, and situational awareness. For timing, the revoking method should seek to maintain peak engagement levels, thus returning system control to the user prior to an under-load situation. The duration of the automation should be long enough to allow the user to recover from an over-taxed state, while short enough to prevent loss in situational awareness. The revoking method should also avoid excessive handoffs and illogical task interruption. Additionally, there are likely differences between perceived workload and actual workload, as well as differences between the physiological measures and actual workload. Determining the appropriate measure of workload and the triggering thresholds will key to earning and maintain user acceptance and trust.

CHAPTER 3
RESEARCH METHODOLOGY

3.1. Introduction

This chapter begins with an overview of the phases of the study, including the inputs to and outputs from each phase. Subsequent sections provide further details for each phase including data sources, model expectations, validation criteria, and analysis techniques.

3.2. Overview of Research Methodology

This research study is divided into five distinct phases:

1. Baseline Model Construction
2. Baseline Model Validation
3. Baseline Model Workload Evaluation
4. Adaptive Automation Experiments
5. Revoking Strategies Experiments

Figure 5 shows these five phases pictorially. Further descriptions of each phase are provided in the subsequent sections of this chapter. Note that Phase 2 of the methodology addresses Sub-Question 1, Phase 3 addresses Sub-Questions 2 and 3, Phase 4 addresses Sub-Question 4, and Phase 5 addresses Sub-Question 5.

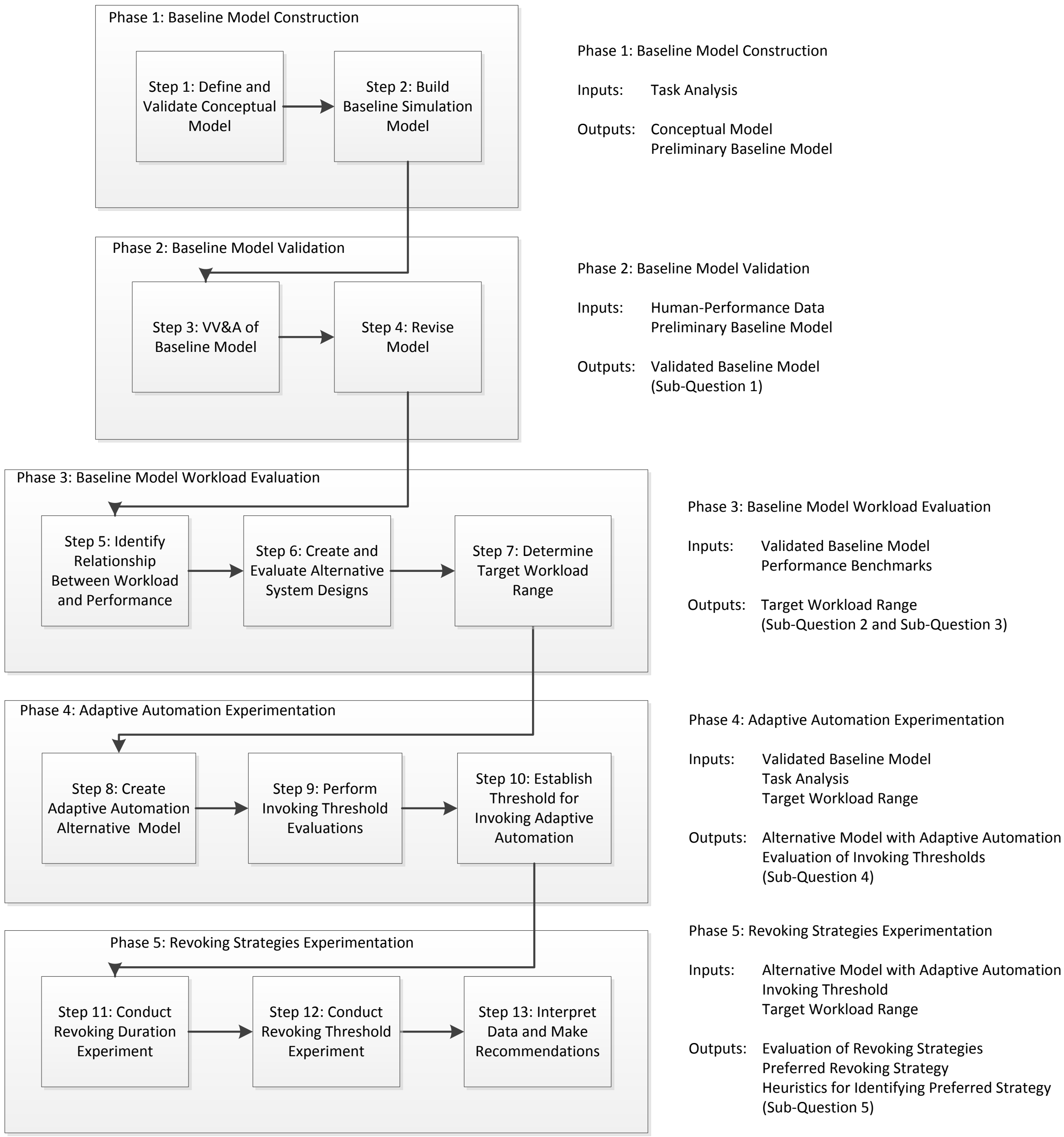

Figure 5: Methodology Flow Chart

3.3. Phase 1: Baseline Model Construction

Overview: The first phase, Baseline Model Construction, includes establishing the conceptual model of the specific tasks of interest and building a discrete event simulation based on this conceptual model. The case scenario for this particular study is a human supervisory control situation in which a human operator receives and interprets intelligence outputs from multiple unmanned assets, and then identifies and reports potential threats and changes in the environment. The tasks are called intelligence, surveillance and reconnaissance (ISR) tasks. Specifically, the ISR tasks are comprised of virtually-simulated unmanned vehicle (UV) Change Detection and Threat Detection tasks. The Change Detection and Threat Detection tasks are performed by human participants as part of an ongoing research project performed by the Applied Cognition and Training in Immersive Virtual Environments Laboratory in the Institute for Simulation & Training at the University of Central Florida (UCF IST ACTIVE Lab) studies for the United States Army Research Laboratory (ARL). The ACTIVE Lab conducts state-of-the-art basic and applied research centered on human performance using live, simulated, and virtual environments. The Lab's core research areas are Human-Robot Interaction, Physiological Assessment and Human Factors, and Training and Education. The ACTIVE Lab partners extensively with defense organizations including the Office of Naval Research (ONR), ARL, the Defense Advanced Research Projects Agency (DARPA), U.S. Joint Forces Command (USJFCOM), and the Office of the Secretary of Defense (OSD) (Applied Cognition & Training in Immersive Virtual Environments Lab, 2011). Although the human participatns are interacting with a virtual simulated system, the cognitive workload the participants experience is real and that which is the subject and focus of the conceptual model and the baseline model.

Inputs: This phase requires, as input, a detailed task analysis capturing sub-tasks, sub-task sequencing, events, event triggers, cognitive resources engaged, user responses, and process times.

Outputs: The outputs of this phase include: 1) a conceptual model of the Change Detection and Threat Detection scenarios that have been validated by UCF IST ACTIVE Lab subject matter experts, and 2) a preliminary baseline model. The preliminary baseline model includes all sub-tasks identified during the task analysis and provides performance measures based on input data. The performance measures include task accuracy for each variation of the Change Detection and Threat Detection tasks.

Detailed Description: The purpose of Phase 1 is to establish the conceptual model of the specific tasks being considered and build a preliminary discrete event simulation based on this conceptual model. This phase is composed of two steps: 1) Define and Validate the Conceptual Model, and 2) Build Baseline Simulation Model.

Defining the conceptual model begins with a detailed task analysis of the ISR tasks to be modeled. As mentioned herein, the tasks consist of virtually-simulated UV Change Detection and Threat Detection tasks performed by human participants as part of an ongoing research project performed by UCF IST ACTIVE Lab studies for the U.S. Army Research Laboratory. While Threat Detection and Change Detection are the primary tasks, these tasks each involve a finite set of sub-tasks in order to accomplish them. The task analysis consists of a complete listing of each sub-task performed by the human participant, including each event (e.g., icon change), the mechanism for triggering the event (e.g., time), and the possible responses to each event. The task analysis also captures the process flow for the sub-tasks and whether sub-tasks are performed in parallel or in series. The task analysis captures the corresponding resource channel(s) used to perform each sub-task (e.g., visual, cognitive, motor) and the degree to which each channel is used (e.g., for cognitive - alternative selection, for evaluation/judgment, for encoding/decoding, for calculation).

The human experiments also provide rich samples of process times for each sub-task and performance data for the main tasks. Each sample of processing times is used to generate an empirical estimate of the underlying theoretical probability distribution that describes the behavior of the processing times of the corresponding sub-task. After the probability distributions are analyzed for goodness of fit using appropriate statistical tests, these probability distributions are then used in the simulation to generate stochastic sub-task processing times. The performance data for the main tasks consist of the percent accuracy for each task by task load level. This information is used to verify the model during the Build phase.

The conceptual model is validated using subject matter expert (SME) reviews of the model. The SMEs are researchers from the UCF IST ACTIVE Lab, who are the designers and builders of the virtual simulated system and the original scenarios used in the human experiments. The SME review of the conceptual model includes: presence of all essential sub-tasks, accurate capture of level and changes in task load, accurate reflection of task duration, correct sequencing of events, and accurate depiction of event timing and concurrence. The conceptual model is updated using the SME feedback. The validated conceptual model is used to construct the baseline simulation model using the Improved Performance Integration Research Tool (IMPRINT) Pro 3.5 (Alion Science and Technology Corporation, Inc., 2013). IMPRINT, developed by the United States Army Research Laboratory (ARL), is a discrete event simulation software that specifically accounts for human factors modeling such as workload, training, and stressors (United States Army Research Laboratory, 2010).

The preliminary baseline model simulates a single human operator performing all of the sub-tasks required to complete the Threat Detection and Change Detection tasks. Each sub-task is associated with a workload score for each resource channel (i.e., visual, auditory, cognitive, and psychomotor), as applicable, with the value of the score accounting for the level of capacity required to perform the task, as described by Bierbaum et al. (1989). The simulation mimics the human experiment scenarios by varying the task load using the same pattern.

The results from the baseline simulation model provide cognitive workload values by resource channel for each event in the simulation event log, as well as aggregated time-averaged workload scores for each single- and dual-task scenario by task load.

3.4. Phase 2: Baseline Model Validation

Overview: This phase entails revising and validating the preliminary baseline model based on human-performance and cognitive workload data for the main tasks.

Inputs: Validation of the preliminary model is conducted through a comparison of the model outputs (workload) to the human-performance data collected by the UCF IST ACTIVE Lab. In order for the preliminary model to be considered valid, the model outputs should follow similar patterns of difficulty level by scenario and task load as characterized by the human-performance data. After this comparative analysis, the model is revised as necessary, to reflect real-world observations.

Outputs: This phase concludes with a validated baseline model. This phase answers the first research sub-question.

Detailed Description: In order to validate the preliminary baseline model, the study conducts a correlation analysis of the model workload outputs to the human-performance data collected by the ACTIVE Lab. The human-performance data include a number of continuous physiological measures including brain activity, heart rate, blood oxygenation in the brain, and eye activity. After the ACTIVE Lab collects these physiological data, the data are compared with subjective workload data and the intended task load in order to identify the measure(s) that most accurately reflect changes in workload. The preliminary baseline model validation utilizes both subjective and physiological measures. The preliminary baseline model is considered valid if the model's predicted workload correlates to the measured workload from the human experiments. This validated model provides the answer to research Sub-Question 1, "Can simulation modeling predict cognitive workload as well as established measures of cognitive workload?"

3.5. Phase 3: Baseline Model Workload Evaluation

Overview: The purpose of this phase is to characterize the relationship between the workload and performance from both the human-performance data and the predicted workload from the validated baseline DES model in order to identify a desired workload range that achieves peak performance. This phase also includes the development of alternative design configurations to the original virtual simulated system configuration in order to demonstrate the DES model's effectiveness for use in system design configuration evaluation.

Inputs: This phase requires the validated baseline model providing predicted workload data. Establishing a target workload range also requires performance benchmarks/standards for the specific tasks. These benchmarks are generated from the ACTIVE Lab's human-performance data.

Outputs: This phase produces an established target workload range based on the cognitive workload score(s) associated with peak task performance. This phase answers research Sub-Questions 2 and 3.

Detailed Description: This phase consists of three steps: 1) identify the relationship between workload and performance, 2) create and evaluate alternative system designs, and 3) determine a target workload range based on the cognitive workload score.

The first step is to identify the relationship between workload and performance. First, the performance and workload data from the human experiments are analyzed to determine the shape of the task load tradeoff curve. Next, the workload outputs from the discrete event simulation baseline model are analyzed to determine the corresponding performance profile that aligns with that from the human experiments.

The second step of this phase is to create models of alternative system design configurations and evaluate these alternative designs using the cognitive workload outputs from the simulation. This evaluation demonstrates the utility of using DES for system design evaluation based on cognitive workload. This is important since the study's goal is to evaluate the relative performance of adaptive automation system designs. This evaluation answers research Sub-Question 2, "Can computer simulation modeling be used to evaluation system designs based on predicted cognitive workload?"

The last step of this phase is to establish a target cognitive workload range based on desired performance. The analysis of the human-performance data provides performance standards for each task. The desired performance range is translated into the equivalent cognitive workload scores, which is the desired target workload range to be achieved by the adaptive automation. This target workload range answers research Sub-Question 3, "How can simulation modeling be used to determine the target level or range of cognitive workload scores for adaptive automation?"

3.6. Phase 4: Adaptive Automation Experiments

Overview: The purpose of this phase is to incorporate adaptive automation into the baseline model, and to evaluate adaptive automation invoking thresholds.

Inputs: This phase requires the validated baseline DES model in order to build the adaptive automation model. The task analysis is used to identify the sub-tasks to be adaptively automated. The target workload range combined with the alternative model's workload and situational awareness results are used to evaluate the invoking thresholds.

Outputs: This phase establishes an alternative model with adaptive automation, including a triggering threshold for invoking automation. The main product of this phase is workload and situational awareness metrics for the invoking thresholds based on time and workload score. This phase addresses research Sub-Question 4.

Detailed Description: During Phase 4, adaptive automation logic is incorporated in the DES model. This phase begins with the creation of alternative simulation models that incorporate adaptive automation. Thus, these models alter the baseline model by dynamically allocating certain tasks to the automated system from than the human operator.

The first adaptive alternative model enables the evaluation of various triggering thresholds for invoking the adaptive automation. Simulation experiments are performed using the DES models, where invoking thresholds are varied, producing varied workload and situational awareness impacts. The relative performance of these triggering thresholds are then evaluated in order to establish a preferred triggering threshold, which addresses research Sub-Question 4, "How can simulation modeling be used to determine a preferred invoking threshold for adaptive automation?" It is likely that establishing an invoking threshold will not be sufficient in obtaining the target workload range threshold established in Phase 3, and that varying the criteria for the revoking strategy produces further workload gains.

3.7. Phase 5: Revoking Strategies Experimentation

Overview: The purpose of this phase is to conduct revoking strategies experiments using simulation to evaluate the relative workload impacts from varying automation durations and revoking thresholds. This phase also analyzes and interprets the data produced in the revoking strategies experiments in order to identify a preferred revoking strategy.

Inputs: This phase requires the alternative DES model with adaptive automation, including the established invoking threshold. The target workload range is utilized to analyze which revoking strategies are preferred. The experiment results are also analyzed using analysis of variance (ANOVA) statistical tests to determine which values within a particular strategy produce a statistically significant difference in workload and situational awareness.

Outputs: Completion of this phase produces a recommendation for a preferred revoking strategy based on statistical analyses and heuristics for identifying preferred revoking strategies. This phase addresses the fifth research sub-question.

Detailed Description: Once the invoking threshold is established, alternative models are then updated to capture various revoking strategies. These revisions include: (1) revoking after a duration of time and (2) revoking based on cognitive workload score threshold. The analysis entails evaluating which revoking strategies are able to meet the target workload range. Additionally, ANOVA is used to determine which values within a particular strategy produce a statistically significant difference in cognitive workload and situational awareness. The research hypothesis is that the revoking strategies differ in workload and situational awareness. This research is exploratory, and, thus, there is no directional hypothesis regarding the revoking strategies. This analysis addresses research Sub-Question 5, "How can simulation modeling be used to determine a preferred revoking strategy for adaptive automation?" Upon completion of Phase 5, this study will have addressed the overall research question: "Can simulation-based modeling of cognitive workload be used for evaluating adaptive automation invoking and revoking strategies?"

3.8. Conclusion

This chapter details the methodology for answering this study's primary research question and sub-questions through the use of discrete event simulation to model human performance and cognitive workload while performing Threat Detection and Change Detection tasks. CHAPTER 4 describes the conceptual model and the baseline simulation model in detail.

CHAPTER 4
DEVELOPMENT OF THE BASELINE DISCRETE EVENT SIMULATION MODEL

4.1. Introduction

The general case scenario is a human supervisory control situation that involves a system operator who receives and interprets intelligence outputs from multiple unmanned assets, and then identifies and reports potential threats and changes in the environment. This scenario is common and quite relevant in the military context, specifically in tactical-level counterinsurgency intelligence, surveillance, and reconnaissance (ISR) operations. The specific scenario involves a system operator who is responsible for continuously monitoring and processing information transmitted by multiple remotely-controlled unmanned systems. The operator views a live feed of enemy and friendly activity in a hostile environment from each system. This live feed is broadcast to a configuration of computer screens. The operator must identify and react to any threats and anomalies so that potential risks to U.S. assets and interests are mitigated or neutralized as soon as possible.

The future of ISR increasingly involves the strategic and tactical use of unmanned assets requiring warfighters to use and interface with complex computing and information technology systems. In order to ensure that unmanned system operators perform effectively, the Department of Defense (DoD) must design systems that balance cognitive workload while maintaining a high level of performance. This makes ISR operations an appropriate and worthwhile case scenario for this study.

4.2. PRIME 2 Study Description

This section describes the controlled human-performance study that supports the construction of the baseline discrete event simulation (DES) model. The human-performance study is the Physiologically-based Robot Interaction as Multimodal Exchanges Phase 2 (PRIME 2) study. PRIME is an on-going research project for the U.S. Army Research, Development and Engineering Command (RDECOM) as part of the Human Robot Interaction Analysis for Training Simulations & Operational Neuroscience (HATS-ON) that is performed by the Applied Cognition and Training in Immersive Virtual Environments (ACTIVE) Laboratory in the Institute for Simulation & Training at the University of Central Florida.

4.2.1. Case Scenario – Mission Description

In the PRIME 2 study, participants represent military operators responsible for monitoring and reporting information received from multiple autonomous unmanned vehicles. One of the autonomous vehicles is an unmanned ground vehicle (UGV), which has a front-mounted camera. This vehicle follows a pre-programmed route through a hostile urban environment. The UGV provides the participant with a video feed from its camera, which allows the participant to see both friendly and enemy forces within the field of view of the UGV. The participants are responsible for identifying threats that appear on the UGV camera feed.

The participants are also presented with an aerial map displaying points of interest. This aerial map is updated based on intelligence provided by unmanned aerial vehicles (UAVs) that are conducting surveillance. Thus, points of interest appear, disappear, or move as new information is received. Participants are responsible for reporting any changes that occur on this aerial map.

During the study, participants will perform each of these tasks (reporting threats for the UGV feed and identifying changes on the UAV area map) individually in single-task scenarios, as well as simultaneously in dual-task scenarios.

4.2.2. Study Participants

As part of the PRIME 2 Study, 150 undergraduate students serve as the main participants in the study, performing the single- and dual-task scenarios. The participants range in age from 18-45 with a mean age of 19.6. There are 85 male participants and 65 female participants. All participants are right-handed.

4.2.3. Study Equipment and Experimental Setup

In the PRIME 2 study, the environmental conditions are modeled using a virtual environment of realistic simulated operations. The experimental setup of the study consists of an Operator Control Unit (OCU) that is comprised of a single computer monitor and a computer mouse for a single human participant. The participants perform the tasks on a standard desktop personal computer with a 3.2 GHz Intel Core i7 processor. The OCU is viewed on a 22” monitor with a 16:10 aspect ratio. The participants use a standard desktop mouse to interface with the OCU, and self-assessment ratings are recorded using a standard external desktop computer microphone (see Section 4.2.5 for description of self-assessments).

A number of physiological measures are collected for each participant. Eye tracking is collected using Seeing Machines’ FaceLAB 5 desk-mounted eye tracking system. Electroencephalography (EEG) and electrocardiography (ECG) measurements are taken with Advanced Brain Monitoring’s B-Alert X10 EEG system, which includes a nine-channel sensor strip with that is attached to the participants scalp, two reference electrodes attached to the participant's mastoid bones, and two EKG sensors, one attached to the right collar bone and the other attached to the left lower rib bone. Functional near-infrared (fNIR) spectroscopy captures the blood oxygenation in the frontal lobe and is measured using Somantics' Invos Cerebral/Somatic oximeter, model 5100c, which is attached to the participant's forehead. Finally, transcranial Doppler (TCD) sonography is used to measure the blood flow into the brain, by attaching sensors from Spencer Technologies' ST^3 Digital Transcranial Doppler, model PMD150, to the right and left sides of the participant's skull near the temple bones.

Prior to running the experimental scenarios, the participant receives a personal training session that explains: (1) each change type in the Change Detection task and (2) each actor image presented in the Threat Detection task. The training session provides detailed descriptions on identifying changes and distinguishes threats from non-threats, with numerous examples, to ensure that participants understand the mission. The training also explains how to interact with the OCU, and the participants practice performing both single-task scenarios and dual-task scenarios using the OCU. Participants are encouraged to ask questions throughout the training and are given an opportunity for additional practice, if desired.

4.2.3.1. Operator Control Unit

The task is set in a virtual environment in which the participant views information presented by simulated unmanned vehicles. The participant views this environment on the OCU displayed on a computer monitor (Figure 6). The OCU consists of three main windows: the Route Map, the Street View, and the Situation Map. The Route Map is located in the upper left-hand corner of the screen and displays the location of the participant's unmanned ground vehicle. The Street View is located in the upper-center/right portions of the screen and displays the live video camera feed from the unmanned ground vehicle. The Situation Map is located in the lower portion of the screen and displays currently intelligence information on locations of military troops and other items of interest (received from one or multiple UAVs).

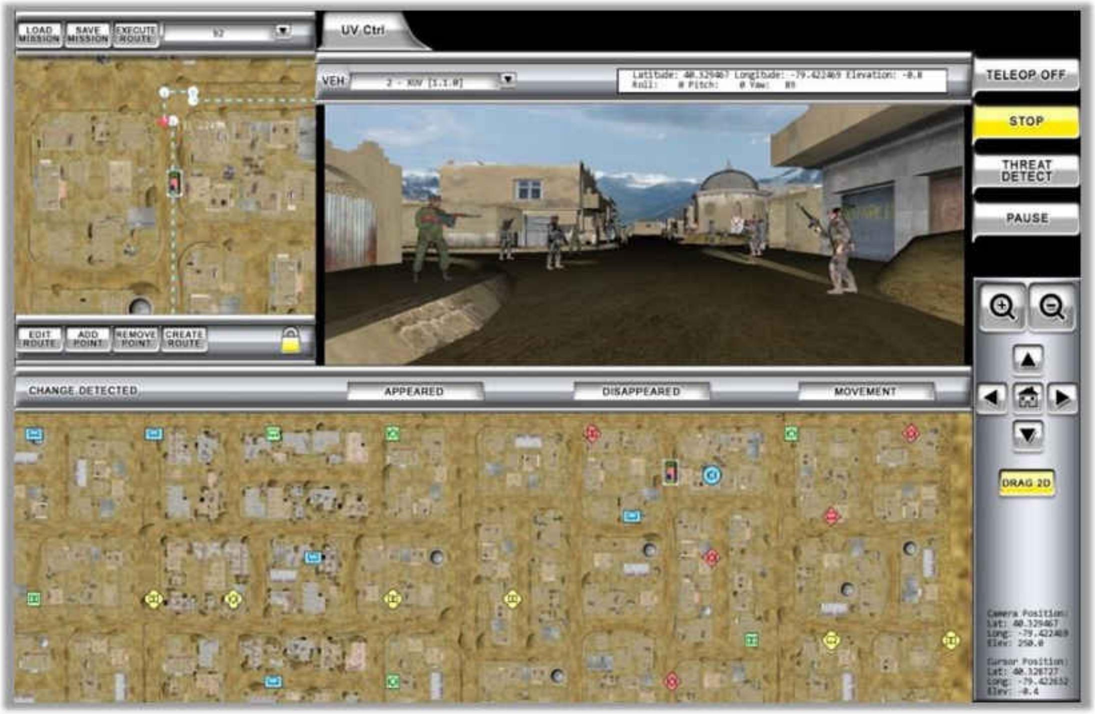

Figure 6: Operator Control Unit

Figure 7 displays the Route Map. The “A” Segments are low task load, the “B” segments are medium task load, and the “C” segments are high task load. Each scenario consists of traveling the entire route map. The participants experience the segments in a random order with the routes randomly starting at any one of the six segments. Those with a “1” proceed in a clockwise direction; those with a “2” proceed in a counter-clockwise direction. For example, if participant starts at B1, the route would be: B1, B2, C1, C2, A1, A2. If a participant starts at A2, then the route would be: A2, A1, C2, C1, B2, B1. The driving task is automated. Therefore, the participant has no control over the vehicle’s route or speed.

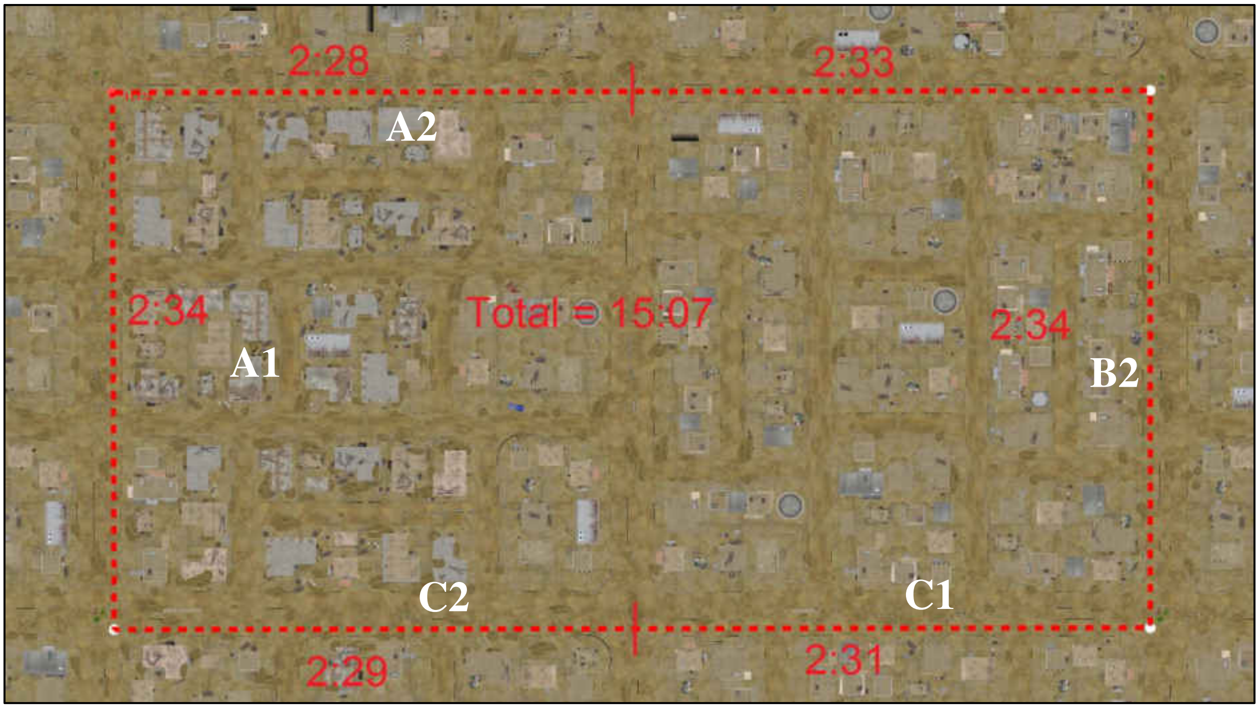

Figure 7: Route Map

4.2.3.2. <u>Change Detection</u>

During the Change Detection task, the participant monitors the Situation Map and identifies any changes that occur. The Situation Map displays an average of 24 colored icons at a time (Figure 8). At specific times, two icons appear, disappear or move. Movements are short distances (approximately 1 inch). The participants indicate that they perceived a change by using their computer mouse to select the button that corresponds with the perceived change. The participant is required to respond to the change before the next change occurs. In other words, if subsequent change event occurs before the participant responds to the previous event, then the participant is unable to respond to that previous event. During the training session before each participant begins the experiment, the participant is advised that it is better to indicate a change rather than do nothing; thus, if the participant is unsure of which type of change has occurred, he or she is expected to guess.

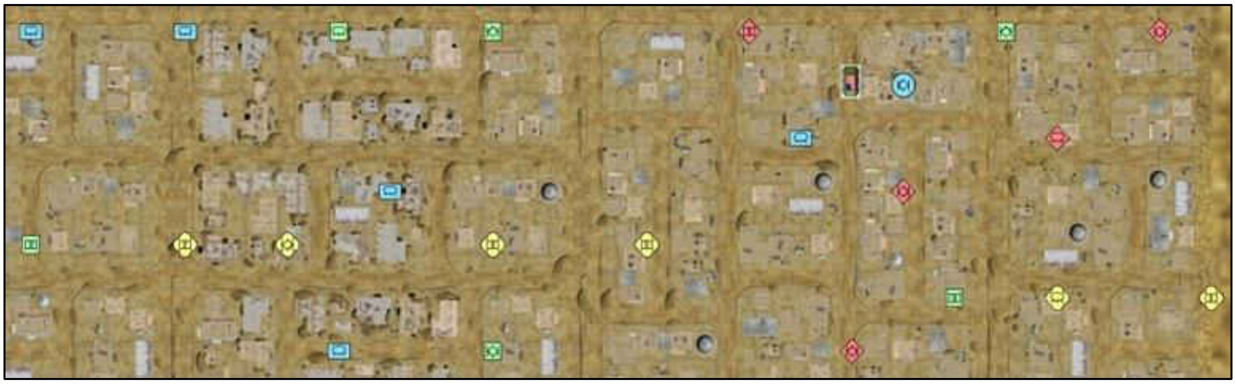

Figure 8: Change Detection Map

Figure 9 is a task flow diagram of the Change Detection task. Each task represented in the diagram requires a unique set of cognitive resources. The workload and resources associated with each of these tasks are presented in Section 4.7 and discussed in detail in Appendix A. Note that the participant may not see all changes that occur. Furthermore, even if the participant does see a change, he or she may not identify it correctly.

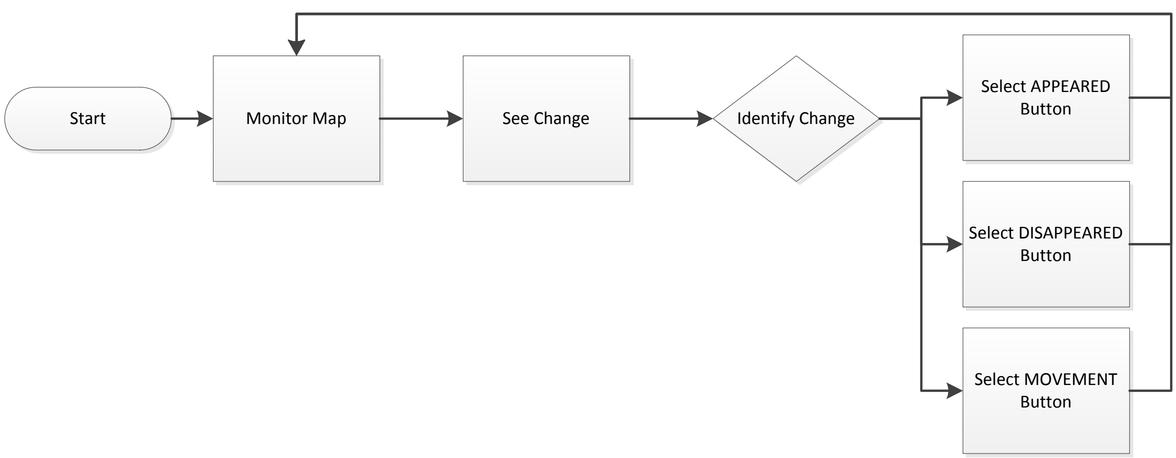

Figure 9: Change Detection Task Sequence

4.2.3.3. Threat Detection

During the Threat Detection task, the participant views the live camera feed on the Street View in order to identify whether the actor images that appear on the screen are threats. There are four types of actors: Friendly Soldiers, Friendly Civilians, Enemy Soldiers, and Armed Civilians (Insurgents). Figure 10 through Figure 13 show example images of the actors. In the PRIME 2 study, both Enemy Soldiers and Armed Civilians are considered threats.

Figure 10: Friendly Soldiers

Figure 11: Friendly Civilians

Figure 12: Enemy Soldiers

Figure 13: Armed Civilians (Insurgents)

The Street View displays streets lined with these actors (Figure 14). When the participant perceives a threat, he or she identifies the threat by using the computer mouse to select the `THREAT DETECT` button and then click on the perceived threat using the mouse. Since the driving task is automated, the participant must identify any threats before the vehicle passes them and the threats disappear from the Street View portion of the monitor. If there are multiple threats on the monitor simultaneously, and the participant may identify them in any order. Note that the participant must select the `THREAT DETECT` button before identifying each threat. Once the participant selects the `THREAT DETECT` button, it is active and it is highlighted. When the participant clicks on the threat in the Street View, the `THREAT DETECT` button is deactivated and it is no longer highlighted. There is no additional feedback that a threat has been selected (e.g., the actor does not change color, become highlighted, or disappear). The participant must click directly on the actor in order to select the actor. If the participant clicks near the actor but not on it, then the actor is not selected. In the event a participant believes that he or she may have not clicked directly on the actor, the participant is permitted to re-select the actor.

Figure 14: Street View

Figure 15 is the flow diagram for the Threat Detection task. The workload and resources associated with each of these tasks are presented in Section 4.7 and discussed in detail in Appendix A. Note that while the participant is not required to take any action for non-threats, the presence of these actors creates visual clutter and additional workload for the participant. Additionally, the participant may not identify all actors correctly. As described in Section 4.5, participants sometimes pre-load the `THREAT DETECT` button. In these cases, the participant selects the `THREAT DETECT` button before he or she sees or identifies an actor as a threat. The logic for the baseline DES model accounts for both the intended task flow and this exception.

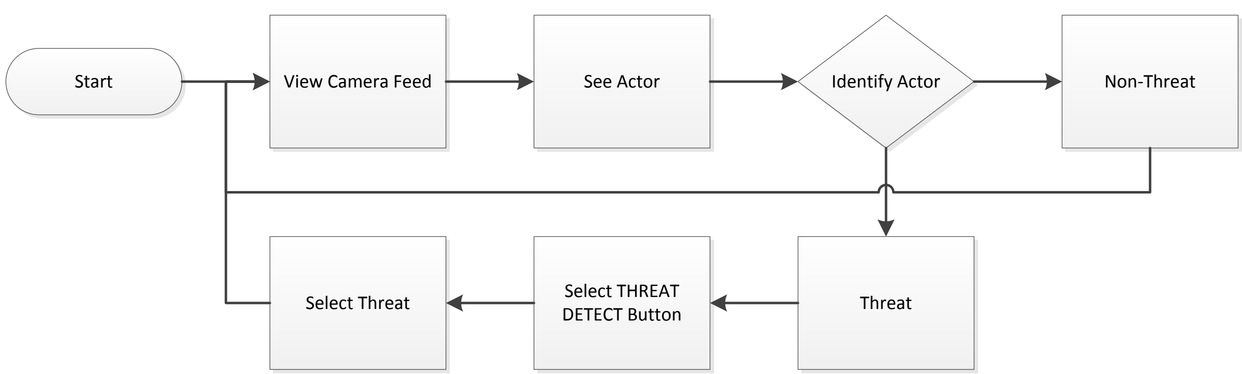

Figure 15: Threat Detection Task Sequence

4.2.3.4. Dual-Task

For the dual-task scenarios, the participant performs both the Change Detection and Threat Detection tasks simultaneously. The participants are told that the tasks are both equally important. There is no change in the task flows between the single- and dual-task modes. The participants are free to move back and forth between the two tasks at any point; they can even select the `THREAT DETECT` button on the Threat Detection task, and then select one of the Change Detection buttons before clicking on the actor that has been identified as a threat.

4.2.4. Study Procedures and Experimental Design

A five-minute resting baseline is recorded for each study participant to obtain at-rest physiological data. Then, the participants perform a 30-minute training session, which includes instructions on how to identify threats, how to identify changes, and how to report threats and changes using the OCU.

The participant performs four scenarios: one scenario consists solely of the Change Detection task (Scenario 1), one scenario consists solely of the Threat Detection task (Scenario 3), and the remaining two scenarios consist of both tasks performed simultaneously. One dual-task scenario has variable Change Detection rates with constant Threat Detection rates (Scenario 2). The second dual-task scenario has variable Threat Detection rates with constant Change Detection rates (Scenario 4). Each scenario consists of six segments: two low task load segments, two medium task load segments, and two high task load segments. Table 7 provides the event rates for each segment in each scenario. Since segments of the same task load within a scenario are always performed back-to-back, the combination of these is referred to as the segment letter without the segment number. Thus, Segment A is comprised of A1 and A2. If the number is not included, the term "segment" refers to the entire task load (portions 1 and 2). A particular task load within a particular segment is referred to as a "scenario-segment combination"; for example, Scenario 4, Segment A is a scenario-segment combination. Thus, there are 4 (scenarios) x 3 (task load segments) = 12 scenario-segment combinations. Each segment is approximately 5 minutes; thus, each scenario is approximately 15 minutes, and the entire mission is 60 minutes.

Table 7: Scenario Task Loads ("CD" means Change Detection and "TD" means Threat Detection)

Task Load	Scenario 1	Scenario 2	Scenario 3	Scenario 4
A1 Low	CD: 6 changes/min TD: none	CD: 6 changes/min TD: 28 actors/min	CD: none TD: 14 actors/min	CD: 12 changes/min TD: 14 actors/min
A2 Low	CD: 6 changes/min TD: none	CD: 6 changes/min TD: 28 actors/min	CD: none TD: 14 actors/min	CD: 12 changes/min TD: 14 actors/min
B1 Med	CD: 12 changes/min TD: none	CD: 12 changes/min TD: 28 actors/min	CD: none TD: 28 actors/min	CD: 12 changes/min TD: 28 actors/min
B2 Med	CD: 12 changes/min TD: none	CD: 12 changes/min TD: 28 actors/min	CD: none TD: 28 actors/min	CD: 12 changes/min TD: 28 actors/min
C1 High	CD: 24 changes/min TD: none	CD: 24 changes/min TD: 28 actors/min	CD: none TD: 56 actors/min	CD: 12 changes/min TD: 56 actors/min
C2 High	CD: 24 changes/min TD: none	CD: 24 changes/min TD: 28 actors/min	CD: none TD: 56 actors/min	CD: 12 changes/min TD: 56 actors/min

4.2.5. Assessment Methods of Study Participant Workload and Performance

Subjective workload measurements are also recorded during the mission via questionnaires. First, at approximately 75% through each scenario-segment combination, the participants experience an audio prompt to rate their workload using an Instantaneous Self-Assessment (ISA) approach, and the participants provide an oral response to the prompt. Upon completion of each scenario-segment combination, each participant reports his or her subjective workload by completing the National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire, Multiple Resource Questionnaire (MRQ), and Dundee Stress State

Questionnaire (DSSQ). For the purposes of the baseline model evaluation, only the ISA and NASA-TLX are utilized, because these instruments provided the clearest measures of cognitive workload.

Objective measures of workload are recorded through the use of various physiological measures as surrogates. The most significant physiological measures include eye tracking Index of Cognitive Activity (ICA), Heart Rate Variability (HRV), TCD mean flow velocity for channel 1 (left side) and channel 2 (right side), alpha EEG frequency at the Poz site, theta EEG frequency at the Fz site, alpha EEG frequency for the occipital lobe, theta EEG frequency for the frontal lobe, and the fNIR mean rs02 for channels 1 and 2, which are left and right, respectively.

4.3. Assumptions

This section includes a list of assumptions for the baseline discrete event simulation models.

4.3.1. General Model Assumptions

Table 8 lists assumptions that apply to the overall baseline model. The table also contains a justification that explains why the assumption is reasonable.

Table 8: General Assumptions for the Baseline Model

Assumption	Rationale
All operators have the same level of experience	PRIME 2 participants all have the same level of experience, obtained through standardized training.
Data used from PRIME 2 participants are representative of the population.	PRIME 2 participants are similar in age and education to military enlisted personnel.
Workload is independent of the order in which the scenarios (Scenario 1: Change Detection, Scenario 3: Threat Detection,	The PRIME 2 study is designed to have clear differences in task load across scenarios. Analysis of data reveals minimal to no order

Assumption	Rationale
Scenario 2 & Scenario 4: dual-task) are performed.	effects. Furthermore, the PRIME 2 study is designed to minimize order effects by using a randomized assignment of scenarios.
Workload is independent of the order in which the segments (A: low, B: medium, C: high) are performed	The PRIME 2 study is designed to have clear differences in task load across segments. Analysis of data reveals minimal to no order effects. Furthermore, the PRIME 2 study is designed to minimize order effects by using a randomized assignment of segments.
No fatigue or learning impacts on workload scores during the performance of the task	Each segment is only five mins long, thus each scenario is 15 min long, and the entire study is 1 hour. These short durations assist in preventing fatigue or learning impacts.
Not modeling individual differences	Individual differences are captured implicitly in probability distributions, but not explicitly through separate models
Not modeling workload impacts of personal factors (emotional stress, fatigue) or environmental factors	Individual differences are captured implicitly in probability distributions, but not explicitly through separate models
Not modeling workload impacts of environmental factors	The PRIME 2 study is conducted in a controlled laboratory setting in which environmental factors are kept constant.
Not modeling subjective/perceived workload	Model's objective workload scores are evaluated against subjective workload, but purpose of model is not to predict subjective workload.
Not modeling impacts of training and/or experience	All PRIME 2 participants have same (limited) training and no prior experience
VACP demand values are accurate	VACP is an accepted measure of objective workload.
Operators make every effort to perform the task correctly.	Military operators would be highly motivated to perform task correctly. PRIME 2 participants fill out a pre-screening questionnaire where they acknowledge that it is unethical to provide random or purposefully incorrect responses to research studies.

4.3.2. ISA Task Model Assumptions

Table 9 lists assumptions and justifications for the ISA portions of the baseline model. The ISA task occurs in Segments A2, B2, and C2 of both the single- and dual-task scenarios.

Table 9: ISA Task Assumptions for Baseline Model

Assumption	Rationale
The participant listens to the ISA prompt for 1.7 seconds.	This is the duration of the audio prompt.
The participant begins to decide an ISA value once the audio prompt concludes.	Prompt initiates the decision cycle.
Audible "umms" are part of the decision time but not part of the oral response.	Even though the participants are using their voices, they are most likely still trying to make a decision.
The probability distribution for the time to decide an ISA value is representative of the population.	The Kolmogorov-Smirnov Test p-value suggests to reject the null hypothesis that the data are from the hypothesized probability distribution.
The fitted distribution for the time to speak the ISA value is representative of the population.	The distribution is a discrete empirical distribution of the sample.
Operator will always complete this task.	PRIME 2 study participants always completed this task.
ISA and primary tasks are performed simultaneously.	PRIME 2 study designed for ISA responses to be given during Change Detection/Threat Detection tasks.

4.3.3. Change Detection Task Model Assumptions

Table 10 lists assumptions and justification for the Change Detection task in the baseline model. The Change Detection task occurs in all segments of Scenario 1 (Change Detection only), Scenario 2 (variable Change Detection, constant Threat Detection) and Scenario 4 (constant Change Detection, variable Threat Detection).

Table 10: Change Detection Task Assumptions for Baseline Model

Assumption	Rationale
Workload is independent of the change event sequence (order of appearance, disappearance, and movements).	Change detection task is simple enough to not have residual workload after a change is identified.
A “no response” in the PRIME 2 study data is interpreted as meaning the operator did not see the change.	PRIME 2 participants are instructed to “guess” if they see a change, but are unsure of its classification. Thus, it is reasonable to assume that non-responses are not due to lack of indecision. Non-responses due to not getting to the button in time before the next change are equivalent to missing the change.
Population non-response rate is equivalent to the non-response rate from the PRIME 2 sample for Scenario 1, Change Detection only.	Higher non-response rates for the dual-task scenarios are captured by the model through workload impacts.
Operators are responding to the current event.	PRIME 2 study participants are instructed that they can only respond to the most recent event. In cases where this assumption appears to be violated, the data in question are removed from the sample, see Section 4.8.
IMPRINT micromodels for task times are accurate.	Micromodels are taken directly from published time-motion studies.
The probability distribution for the time to identify a change is representative of the population.	The Kolmogorov-Smirnov Test p-value suggests to reject the null hypothesis that the data are from the hypothesized probability distribution.
PRIME 2 sample probabilities for selecting a change type, given the actual change event, are representative of the population.	
Monitoring the situation map and identifying/responding to a change are not performed simultaneously.	
No false positives. That is, operators do not identify a change event when no event has occurred.	PRIME 2 subjects had a 4.56% false positive rate. For segments A and B, the rates are a negligible 1.29% and 1.12%, respectively. In segment C, the rate is 6.95%. This high false positive rate is due to a time lag in clicking button when two changes occur close together. Thus, it is due to participants not responding to the most recent event, rather than a true false positive, see Section 4.8 for more detailed

Assumption	Rationale
	discussion.
Operators do not intentionally misidentify changes.	PRIME 2 participants fill out a pre-screening questionnaire where they acknowledge that it is unethical to provide random or purposefully incorrect responses to research studies.
Operators rest their mouse cursor on the `DISAPPEARED` button.	Located in the middle of the other buttons.

4.3.4. Threat Detection Task Model Assumptions

Table 11 lists assumptions and justification for the Threat Detection task in the baseline model. The Change Detection task occurs in all segments of Scenario 2 (variable Change Detection, constant Threat Detection), Scenario 3 (Threat Detection only), and Scenario 4 (constant Change Detection, variable Threat Detection).

Table 11: Threat Detection Task Assumptions for Baseline Model

Assumption	Rationale
IMPRINT micromodels for task times are accurate.	Micromodels are taken directly from published time-motion studies.
The probabilities identified for “pre-loading” the `THREAT DETECT` button are accurate.	“Pre-loading” is identified from the bi-modal nature of the response time distribution for the Threat Detection task.
The “Reaction Time” in the study is the time to locate, track, align, and select.	“Reaction Time” is calculated as the difference between the time when the `THREAT DETECT` button is selected and the threat is selected.
The fitted distribution for the time to select a threat is representative of the population.	The Kolmogorov-Smirnov Test p-value is insufficient to reject the null hypothesis that the data are from the fitted distribution.
The task time and workload to select a non-actor is equivalent to the time to select a threat.	PRIME 2 subjects accidently select non-actors, intending to select threats.
Any actor that is not identified by the operator as a threat was deliberately identified as a non-threat.	
PRIME 2 sample probabilities for identifying	

Assumption	Rationale
an actor as a threat, given the actor type, are representative of the actual population.	
PRIME 2 sample probabilities for identifying a non-actor as a threat are representative of the population.	
Each actor is identified only once.	
Operators do not intentionally identify non-actors or non-threats as threats.	PRIME 2 participants fill out a pre-screening questionnaire where they acknowledge that it is unethical to provide random or purposefully incorrect responses to research studies.
Actors appear on the screen at a visible fixed distance of 14.14 meters.	This is based on the 80th percentile of responses from the pilot study. Equates to the 74th percentile for the first 30 participants in the PRIME 2 Study.
Once the participant starts the "Select `THREAT DETECT` button, Select Actor" sequence, they are able to complete it. Thus, if a threat is visible before the `THREAT DETECT` button is selected, the threat will still be visible for selection after selecting the button.	Threats are on the screen for an average of approximately 33 seconds, pressing the button takes 0.4 sec, selecting the threat takes an average of less than 2 seconds. The proportionally short duration of the "Select `THREAT DETECT` button, Select Actor" sequence compared with the time that the actor is visible on the screen makes this assumption reasonable.

4.3.5. Dual-Task Model Assumptions

Table 12 lists assumptions and justification for the dual-task scenarios in the baseline model. The dual-tasks are Scenario 2 (variable Change Detection, constant Threat Detection) and Scenario 4 (constant Change Detection, variable Threat Detection).

Table 12: Dual-Task Scenario Assumptions for Baseline Model

Assumption	Rationale
Distributions fitted for the single-task scenarios are applicable for the dual-task scenarios.	Individually, there is no difference between performing change (or threat) detection in the single-task vs. dual-task scenarios. Impacts in participant response times are due to performing the two tasks simultaneously, not due to a change in the nature of the task.
All assumptions listed in Table 10.	Change Detection task is performed the same way for both single- and dual-task scenarios.
All assumptions listed in Table 11.	Threat Detection task is performed the same way for both single- and dual-task scenarios.

4.4. Change Detection Model

This section describes the discrete event simulation model for Scenario 1, the single-task scenario in which operators perform the Change Detection task at variable event rates. As discussed in CHAPTER 3, all DES models are constructed using the Improved Performance Research Integration Tool (IMPRINT) Pro version 3.5. IMPRINT Pro is a discrete event simulation tool based on the MicroSaint Sharp environment. IMPRINT Pro is developed and supported by Alion Science and Technology for the United States Army Research Laboratory.

The Change Detection discrete event simulation model is based upon the task sequence shown in Figure 9, as well as observations from the performance of the human participants. Each of the six segments shown in Table 7 is represented by a function node within the discrete event simulation (Figure 16). These functions each contain a task network with the specific tasks carried out by the system and the operator during that segment of the task.

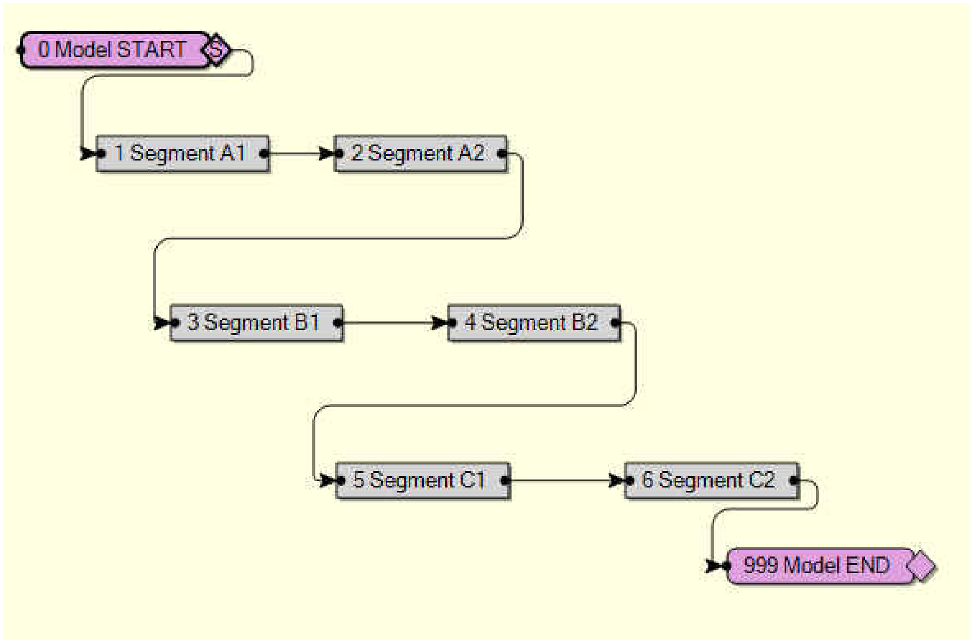

Figure 16: Change Detection Discrete Event Simulation Function Nodes

Figure 17 displays the task network for Segment A1, which is similar to the task sequence from Figure 9. The purple nodes are the tasks that are performed by the operator. For clarity, “participant” refers to human subjects in the PRIME 2 study, and “operator” refers to simulated humans in the discrete event simulation models. The purple nodes are thus tasks that consume time and occupy the operator’s cognitive resources. The green nodes represent tasks that are performed by the system and other internal logic. These tasks may have time durations or may be instantaneous, and they do not consume any of the operator’s cognitive resources.

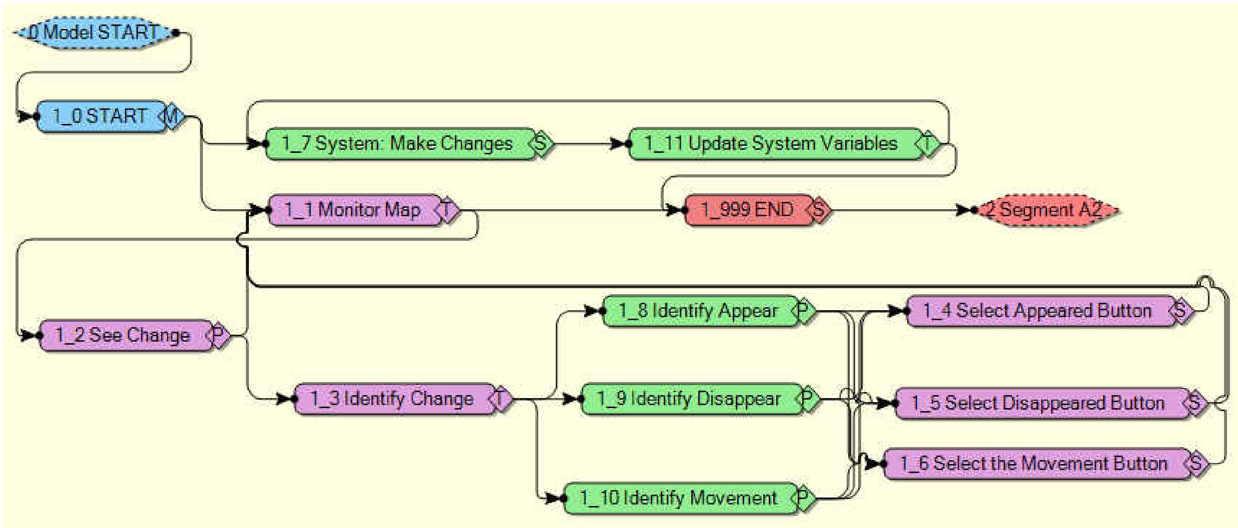

Figure 17: Change Detection Task Network for Segment A1

When the simulation begins it creates two entities. One entity represents the computer system and is responsible for generating the various icon changes that the operator experiences on the computer screen. This entity proceeds from the Start node to the System: Make Changes node and the Update System Variables node. The system loops continuously through these two nodes (described in detail in Appendix A) until the allotted time for that segment has passed.

A second entity is generated at the start of the simulation, which represents the human operator, and proceeds to the Monitor Map task. The operator remains in this task until the system generates a change. When a change occurs, the entity moves to the See Change node. This node then contains probabilistic branching logic to determine if a change is seen by the operator. The probabilities for each segment are based on the human-performance data obtained from the PRIME 2 study. If the change is perceived, then the entity moves to the Identify Change node. In this node, operators use their cognitive processes to discriminate which type of change they observed. The entity then exits this node and is routed based on the type of change that actually occurred. The green nodes Identify Appear, Identify Disappear, and Identify Movement are system logic that determines the probability that the operator selects a particular change type based on the actual change type. These probabilities are based on the data obtained from the PRIME 2 study. From there, a particular button is selected to indicate the change type observed (nodes Select Appeared, Disappeared, and Movement Button). After a button is selected, the entity returns to the Monitor Map task to await the next change. After the allotted time for that segment passes, the entity moves to the End node, which leads into the next segment.

The remaining segments in Figure 16 have almost identical task sequences and logic as that of A1 shown in Figure 17, with a few exceptions. These exceptions can be seen in Figure 18, the task network for Segment A2. For all segments after A1, an additional system task System Wait is added. This task delays the start time of the system changes and represents the transition time that is experienced by the operator between segments.

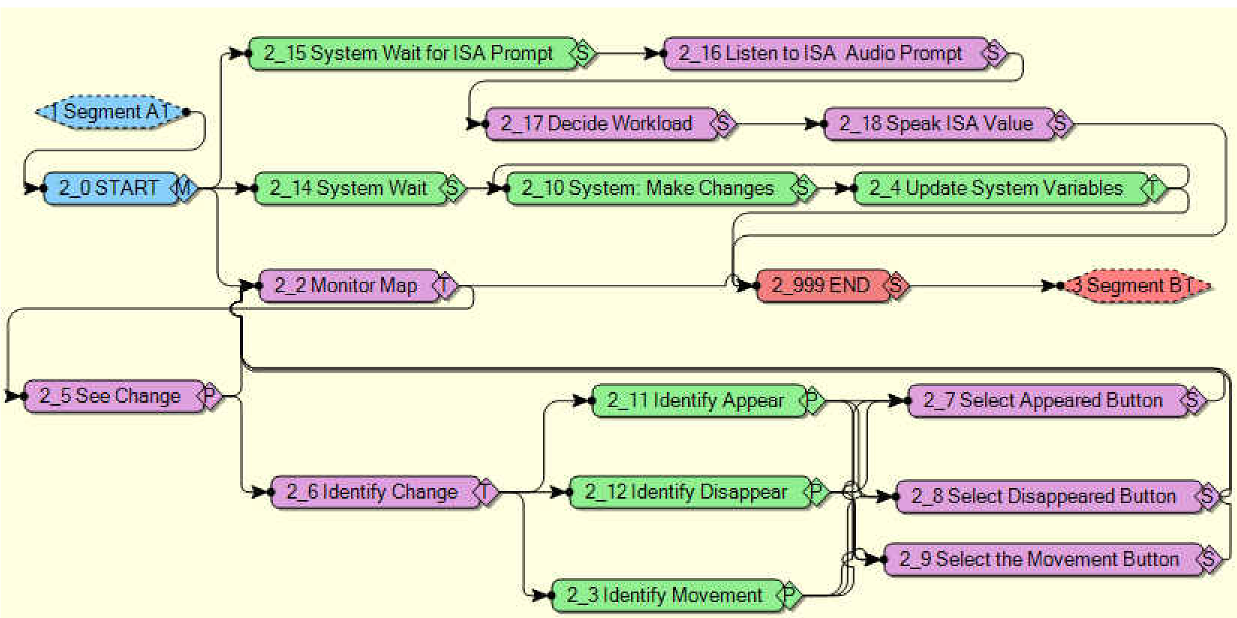

Figure 18: Change Detection Task Network for Segment A2

Additionally, Segments A2, B2, and C2 include the Instantaneous Self-Assessment (ISA) task sequence. The ISA is a measure of global workload on a scale of 1-5; see Table 13 for rating descriptions. In this task, the operator hears an audio prompt that says “Please rate your workload." The operator then responds orally with a number between 1 and 5 that best matches his or her perceived workload level. The green node System Wait for ISA prompt is a time placeholder that delays the audio prompt until the correct time. The operator hears the prompt during the Listen to ISA audio prompt, then uses his or her cognitive processes to choose the correct workload value at the Decide Workload node, and finally responds orally during the Speak ISA Value node.

Table 13: Instantaneous Self-Assessment Ratings (adapted from Jordan and Brennen, 1992)

Rating	Workload	Description
1	Under-utilized	Nothing to do. Rather boring.
2	Relaxed	More than enough time for all tasks. Active on the task less than 50% of the time.
3	Comfortably Busy Pace	All tasks well in hand. Busy but stimulating pace. Could keep going continuously at this level.
4	High	Non-essential task suffering. Could not work at this level very long.
5	Excessive	Behind on tasks; losing track of the full picture.

Table 14 lists all of the nodes found in the Change Detection Model along with the purpose of these nodes. For a detail description of the individual task nodes, see Appendix A.

Table 14: Change Detection Nodes

Node	Performed by	Purpose
Start Node	System	System logic node that generates entities necessary to start the segment
Monitor Map Node	Operator	Sub-task in which the operator monitors the Situation Map for changes
See Change Node	Operator	Sub-task to determine whether or not a change event is detected by the operator
Identify Change Node	Operator	Sub-task in which the operator uses

Node	Performed by	Purpose
		cognitive process to categorize a change as either an appearance, disappearance, or movement
Identify Appear, Identify Disappear, and Identify Movement Nodes	System	System logic nodes to determine which change type button will be pressed given a particular change event
Select Appeared, Select Disappeared, and Select Movement Button Nodes	Operator	Sub-task in which the human operator presses the `APPEAR`, `DISAPPEAR`, or `MOVEMENT` buttons
System Wait Node	System	System logic node that prevents the change events from starting prematurely
System: Make Changes Node	System	System logic node that generates the change events
Update System Variable Node	System	System logic node that updates system counters and batching variables
System Wait for ISA Prompt Node	System	System logic node to delay ISA prompt until the appropriate time in the segment
Listen to ISA Prompt Node	Operator	Sub-task in which the operator listens to the ISA prompt
Decide Workload Node	Operator	Sub-task in which the operator decides the ISA score that corresponds to his or her current level of workload
Speak ISA Value Node	Operator	Sub-task in which the operator speaks his or her ISA score out loud
End Node	System	System logic node to batch entities, end the current Segment, and begin the next Segment

4.5. Threat Detection Model

This section describes the discrete event simulation model for Scenario 3, the single-task scenario in which operators perform the Threat Detection task at variable event rates. The Threat Detection discrete event simulation model is based upon the task sequence shown in Figure 15 as well as observations from the performance of the human participants. Each of the six segments shown in Table 7 is represented by a function node within the discrete event simulation (Figure 19). These functions each contain a task network with the specific tasks carried out by the system and the operator during that segment of the task.

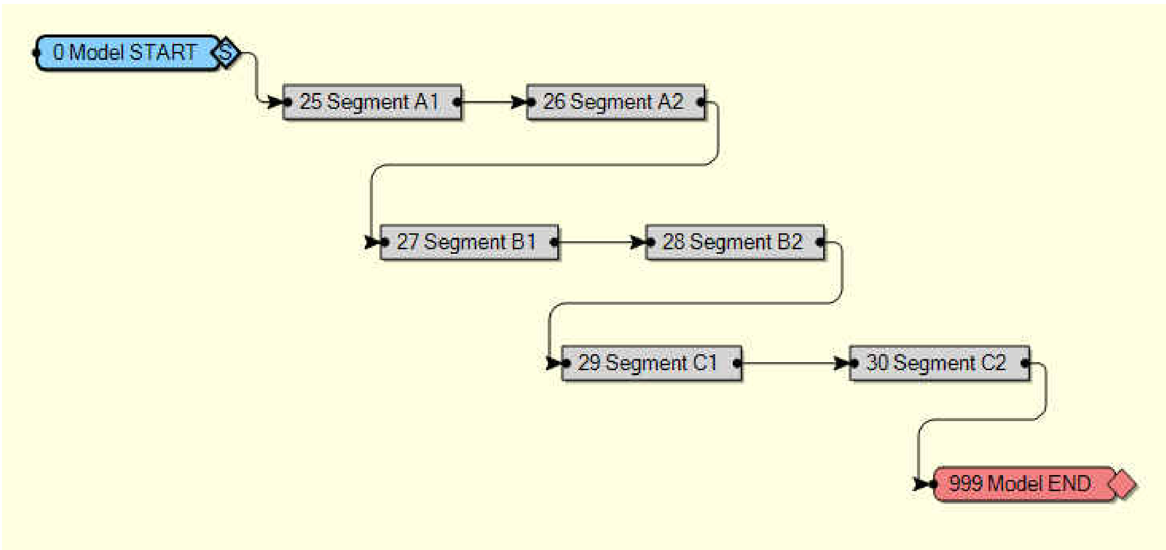

Figure 19: Threat Detection Discrete Event Simulation Function Nodes

Figure 20 displays the task network for segment A1. The purple nodes are the tasks that are performed by the operator. These tasks consume time and occupy the operator's cognitive resources. The green nodes represent tasks that are performed by the system and other internal logic. These tasks may have time durations or may be instantaneous, and they do not consume any of the operator's cognitive resources.

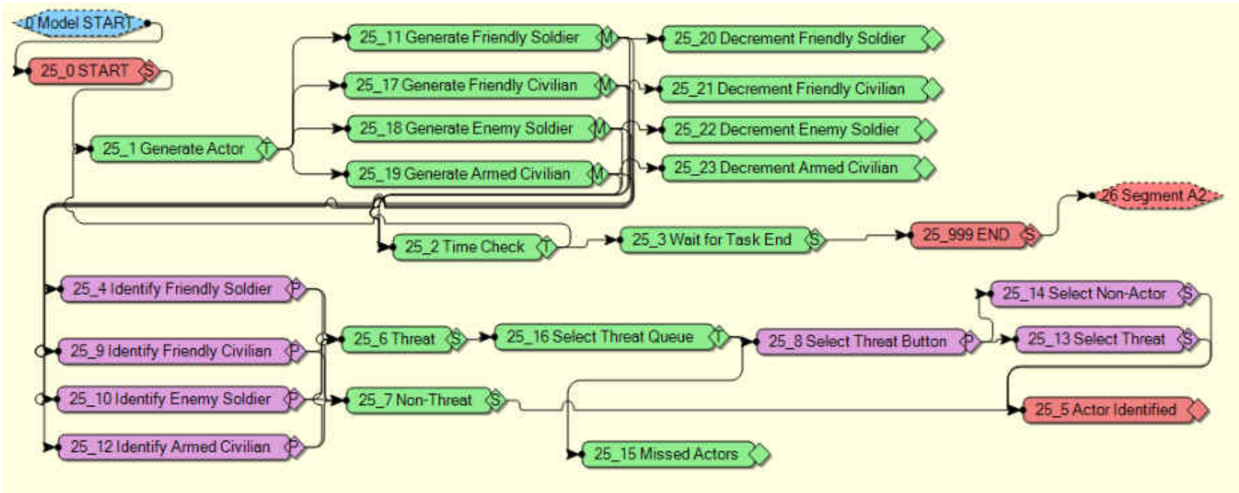

Figure 20: Threat Detection Task Network for Segment A1

When the simulation begins, it creates a single entity that proceeds from the START node to the Generate Actor node. The Generate Actor node contains the system logic necessary to generate the threats that the human operator will respond to. This node determines which type of actor will be generated next. The entity exits the Generate Actor node and is routed according to the actor type to be generated. The entity then proceeds to the Generate Friendly Soldier, Generate Friendly Civilian, Generate Enemy Solider, or Generate Armed Civilian node. These nodes increase system counters in order to keep track of the number of actors of that type present on the screen. These nodes also tag the entities, so that the entities can be tracked for use by the system in the Select Threat Queue node.

From theses nodes, three entities are created using a multiple branching logic. The first entity proceeds to the Decrement Friendly Solider, Decrement Friendly Civilian, Decrement Enemy Soldier, or Decrement Armed Civilian node. These nodes hold the entities representing the actors visible on the screen, until the visible duration of the actor has passed, then the entities are disposed and the entity counters are reduced.

The second entity proceeds to the Time Check node, where the entity verifies whether this entity creation was the last for the segment, or whether additional entities still need to be created. If additional entities still need to be generated, then the entity proceeds from the Time Check node back to the Generate Actor node, to repeat the above sequence. Otherwise, the entity proceeds to the Wait for Task End node, where the entity waits for the Clock time to reach the segment end time, whereupon the entity then proceeds to the End node, which Ends the segment and begins the next segment.

The third entity flows to the Identify Friendly Soldier, Identify Friendly Civilian, Identify Enemy Soldier, or Identify Armed Civilian node that corresponds to the actor type for that entity. In these nodes, the human operator identifies the type of actor that is on the screen. The entity then uses probabilistic branching logic to proceed to the Threat or Non-Threat node, based upon the operator's identification of the actor. The probabilities used are based upon data from human participants in the PRIME 2 study.

If the actor is identified as a non-threat, then no action is necessary, thus the entity flows to the Actor Identified disposal node. If the actor is identified as a threat, then the entity flows to the Select Threat Queue. This node holds entities waiting to be identified. In order to report a threat, the human operator must select the `THREAT DETECT` button and then click on the actor identified as a threat. This Select Threat Queue ensures that the `THREAT DETECT` button is select followed by selecting an actor (or non-actor), before the `THREAT DETECT` button is selected again. The Select Threat Queue also verifies that the actor to be identified is still visible in the system, by referencing the tag generated in the Generate Friendly Soldier, Generate Friendly Civilian, Generate Enemy Solider, and Generate Armed Civilian nodes. If the actor is no longer visible in the system, then the entity will proceed to the Missed Actors disposal node. Otherwise, the entity will proceed to the Select Threat Button node.

During the PRIME 2 study, approximately 30% of the time the participants would "pre-load" the `THREAT DETECT` button. That is, they would select the `THREAT DETECT` button, even though they were not ready to report a threat. This assists them by allowing them to select the threat immediately once a threat appears. To account for this in the model, if an entity arrives at the Threat node and there are no entities in the Select Threat Queue, then the entity has an approximately 30% chance of having a task time of zero for the Select Threat Button node. The actual percentage is based upon the study data and varies across segments, see Section 4.8.3.2. After selecting the `THREAT DETECT` button, the entity then faces a probabilistic branching logic, where it either proceeds to Select Threat or Select Non-Actor. These probabilities also vary across segments and are discussed in Section 4.8.3.3. Entities then flow to the Actor Identified disposal node.

The remaining segments in Figure 19 have almost identical task sequences and logic as that of A1 shown in Figure 20, with a few exceptions. As with the Change Detection task, Segments A2, B2, and C2 of the Threat Detection task also include the ISA task sequence. In this task, the operator hears an audio prompt that says "Please rate your workload." The operator then responds orally with a number between 1 and 5 that best matches his or her perceived workload level. There is a System Wait for ISA Prompt is a time placeholder, which delays the audio prompt until the correct time. The operator hears the prompt during the Listen to ISA Audio Prompt, then uses his or her cognitive processes to choose the correct workload value during the Decide Workload node, and finally responds out loud during the Speak ISA Value node.

Table 15 lists all of the nodes found in the Threat Detection Model along with the purpose of these nodes. For a detail description of the individual task nodes, see Appendix A.

Table 15: Threat Detection Nodes

Node	Performed by	Purpose
Start Node	System	System logic node that generates entities necessary to start the segment
Generate Actor Node	System	System logic node that the generates actors that appear on the UGV video feed of the OCU
Generate Friendly Soldier, Generate Friendly Civilian, Generate Enemy Soldier, and Generate Armed Civilian Nodes	System	System logic node that updates counter variables and creates entity tags for use in the Select Threat Queue Node
Decrement Friendly Soldier, Decrement Friendly Civilian, Decrement Enemy Soldier, and Decrement Armed Civilian Nodes	System	System logic node that serves for a holding place for entities (actors) that are visible, and then decrements counter variables when entities (actors) are no longer visible
Time Check Node	System	System logic variable that determines whether the segment is complete or additional actors are to be generated
Wait for Task to End Node	System	System logic variable that prevents the Segment from ending prematurely
Identify Friendly Soldier, Identify Friendly Civilian, Identify Enemy Soldier, and Identify Armed Civilian Nodes	Operator	Sub-task in which the operator categorizes actors on the screen into threats or non-threats
Non-Threat Node	System	System logic node to route entities (actors) identified as non-threats
Actor Identified Node	System	System logic node to dispose of identified actors
Threat Node	System	System logic node to route entities (actors) identified as threats
Select Threat Queue Node	System	System logic node to hold identified threats waiting to be reported
Missed Actors Node	System	System logic node to collect and dispose of identified threats that become non-visible before the operator has reported them

Node	Performed by	Purpose
Select Threat Button Node	Operator	Sub-task in which the operator selects the threat button
Select Threat Node	Operator	Sub-task in which the operator selects an identified threat
Select Non-Actor Node	Operator	Sub-task in which the operator mistakenly selects a non-actor instead of a threat
System Wait for ISA Prompt Node	System	System logic node to delay ISA prompt until the appropriate time in the segment
Listen to ISA Prompt Node	Operator	Sub-task in which the operator listens to the ISA prompt
Decide Workload Node	Operator	Sub-task in which the operator decides the ISA score that corresponds to his or her current level of workload
Speak ISA Value Node	Operator	Sub-task in which the operator speaks his or her ISA score out loud
End Node	System	System logic node end the current Segment, and begin the next Segment

4.6. Dual-Task Models

This section describes the discrete event simulation models for the dual-task scenarios. Scenario 2 is the dual-task scenario in which the operator performs the Change Detection task at variable event rates while simultaneously performing the Threat Detection task at a constant event rate. Scenario 4 is the dual-task scenario in which the operator performs the Change Detection task at a constant event rate while simultaneously performing the Threat Detection task at a variable event rate.

The top-level function nodes for these models are the same as those for the single-task models as shown in Figure 21. Segment A1 for the dual-task scenarios is shown in Figure 22 and Segment A2 is shown in Figure 23. These task networks are a combination of Change Detection and Threat Detection task networks. There are no unique nodes in the dual-task scenarios that did not appear in the single-task scenarios. Since reporting changes and reporting threats both require the use of the computer mouse, additional release condition logic is included in both the Change Detection and Threat Detection task networks to ensure that the operator can only select one item using the mouse at a time.

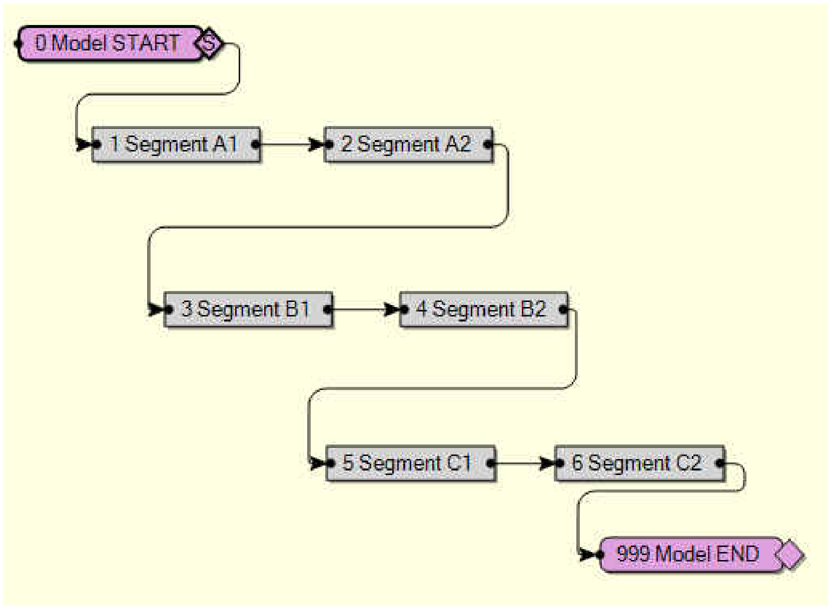

Figure 21: Dual-Task Scenario Discrete Event Simulation Function Nodes

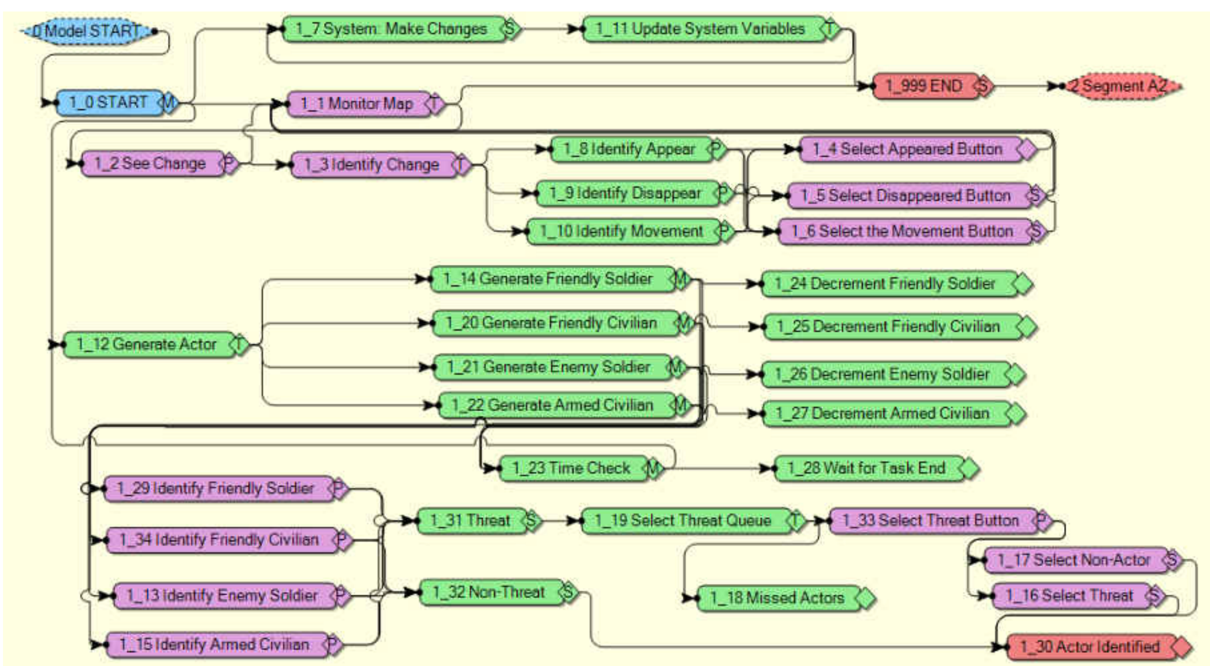

Figure 22: Dual-Task Network for Segment A1

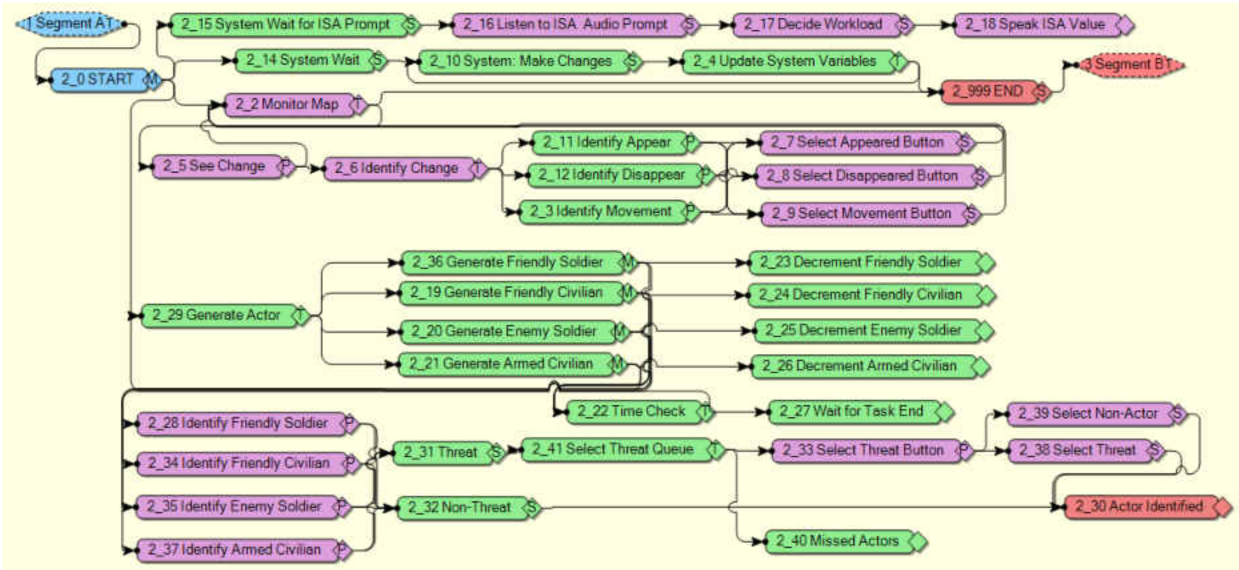

Figure 23: Dual-Task Network for Segment A2

The Select Threat Button, Decide Workload, and Speak ISA Value nodes differ in their task times from the single-task nodes, as described in Appendix A. All other task times, probabilities, and workload values remain constant from the single-task scenarios. Thus, degradations in performance are not built into the dual-task models, but rather will be outputs due to workload overload. The Change Detection event logs and Threat Detection actor generation logs are updated in the dual-task scenarios to reflect the timing of these events under dual-task conditions. See Appendix B for the Change Detection event logs and Appendix C for the Threat Detection event logs.

4.7. Workload Summary

Table 16 through Table 18 display the VACP workload values experienced by the human operator. Table 16 contains the VACP workload values for the Change Detection task. The tasks listed in this table are experienced during every segment of Scenario 1 (Change Detection only), Scenario 2 (variable Change Detection, constant Threat Detection) and Scenario 4 (constant Change Detection, variable Threat Detection).

Table 16: VACP Workload Values for Change Detection Baseline Model Tasks

Task	Visual	Auditory	Cognitive	Psychomotor (Fine Motor)	Speech
Monitor Map	6.0				
See Change	5.0		1.2		
Identify Change	5.0		6.8		
Select Appeared Button	4.0			2.2	
Select Disappeared Button	4.0			2.2	
Select Movement Button	4.0			2.2	

Table 17 contains the VACP workload values for the Threat Detection task. The tasks listed in this table are experienced during every segment of Scenario 2 (variable Change Detection, constant Threat Detection), Scenario 3 (Threat Detection only), and Scenario 4 (constant Change Detection, variable Threat Detection).

Table 17: VACP Workload Values for Threat Detection Baseline Model Tasks

Task	Visual	Auditory	Cognitive	Psychomotor (Fine Motor)	Speech
Identify Friendly Soldier	5.0		1.2		
Identify Friendly Civilian	5.0		1.2		
Identify Enemy Soldier	5.0		1.2		
Identify Armed Civilian	5.0		1.2		
Select Threat Button	4.0			2.2	
Select Threat	4.4			4.6	
Select Non-Actor	4.4			4.6	

Table 18 contains the VACP workload values for the ISA task. The tasks listed in this table are experienced during Segments A2, B2, and C2 of all Scenarios. Note that this task utilizes the auditory and speech channels, which are not used for the Change Detection or Threat

Detection tasks. The ISA task was intentionally designed to use these channels in order to minimize workload conflict.

Table 18: VACP Workload Values for ISA Baseline Model Tasks

Task	Visual	Auditory	Cognitive	Psychomotor (Fine Motor)	Speech
Listen to ISA Audio Prompt		3.0			
Decide Workload			6.8		
Speak ISA Value					2.0

4.8. Probability Distributions

This section describes the derivation of the probability distributions used in the task nodes for the Change Detection and Threat Detection tasks. Statistical software packages used to analyze and fit distributions include: MathWave Technologies Easy Fit Professional version 5.5 (Mathwave Technologies, 2010), Rockwell Automation Input Analyzer version 13.5 (Rockwell Automation, Inc., 2010), and MiniTab version 16.2 (MiniTab, Inc., 2013). Random numbers are generated using the Random Integer Set Generator from Random.org (Random.org, 2012).

Table 19 provides a complete listing of all of the probability distributions in the baseline model, along with their respective parameter values and Kolmogorov-Smirnov p-values. To verify the suitability of these distributions, a simple model with a single task was created in IMPRINT in order to generate 50 random values using the distribution. Based on the low variability, it is determined that 50 replications are sufficient. The descriptive statistics for this simple model are then compared with the descriptive statistics of the original sample data to ensure that the distribution produces reliable data points, in all cases the fitted distribution are found to be suitable.

Table 19: Summary of Probability Distributions Used in Baseline Models (level of significance $\alpha = 0.05$)

Task	Node	Segment	Distribution	Parameters	K-S p-value
ISA – Single-Task	Decide Workload	All	LogLogistic	shape $\alpha = 4.3682$ scale $\beta = 0.72904$	0.367
ISA – Dual-Task	Decide Workload	All	Pearson5	shape $\alpha = 3.7144$ scale $\beta = 2.9176$	0.30882
Change Detection	Identify Change	A	LogNormal	mean $\mu = 1.64$ st.dev. $\sigma = 0.61$	0.0977
Change Detection	Identify Change	B	LogLogistic	shape $\alpha = 6.4634$ scale $\beta = 1.3856$	0.4678
Change Detection	Identify Change	C	Logistic	mean $\mu = 1.17$ st.dev. $\sigma = 0.23$	0.10656
Threat Detection	Select Threat	A	Pearson5	shape $\alpha = 10.991$ scale $\beta = 16.37$	0.976
Threat Detection	Select Threat	B	Pearson5	shape $\alpha = 7.7987$ scale $\beta = 11.542$	0.914
Threat Detection	Select Threat	C	Pearson5	shape $\alpha = 7.2163$ scale $\beta = 9.7574$	0.459

4.8.1. ISA Probability Distributions

For the ISA task sequence, two task times were derived from PRIME 2 participant data: the amount of time deciding the workload score and the amount of time to speak the response. During the experiment, the participant hears a pre-recorded audio prompt that states “Please rate your workload.” The ISA prompts always occur at clock times: 3:50, 8:50, and 13:50, in segments A2, B2, and C2, respectively. The participant then rates his or her workload on a 1-5 scale and states the rating out loud, which is recorded through a microphone and transcribed by a research assistant.

The Decide Workload task time is the time that the participant spends deciding on the appropriate workload score. This decide time begins when the audio prompt ends and the decide time ends when the participant beings to speak their score. The Speak ISA Value task time is the duration of the participant's audio response.

To obtain task times for the Decide Workload and Speak ISA Value nodes, a random sample of 10 participants are selected from the first 30 participants. Each participant selected had completed the full study. Thus, each participant provided 12 ISA score responses – one for each segment A2, B2, and C2 in each of the four scenarios. This sample provides 120 data points for each task time. To obtain the times, each audio file is reviewed using Audacity 2.0 (Audacity, 2013), and times are recorded to the nearest tenth of a second.

4.8.1.1. Decide Workload Task Time

Table 20 summarizes the descriptive statistics for the Decide Workload task. Figure 24 is the histogram of the data, and Figure 25 displays the shape of the data distributions when divided by scenario. A review of the descriptive statistics and the individual distribution shapes data reveals that Scenario 1 and Scenario 3, the single-task scenarios, have similar durations and Scenario 2 and Scenario 4, the dual-task scenarios, have similar durations.

Table 20: Descriptive Statistics for Decide Workload Task

	Scenario 1: Single-Task Variable Change Detection	**Scenario 2: Dual-Task Variable Change Detection**	**Scenario 3: Single-Task Variable Threat Detection**	**Scenario 4: Dual-Task Variable Threat Detection**
Minimum	0.3	0.4	0.3	0.4
Maximum	2	3.1	1.7	3.6
Mean	0.84	1.07	0.77	1.08
Median	0.7	0.85	0.7	0.8
St Dev	0.39	0.65	0.30	0.81

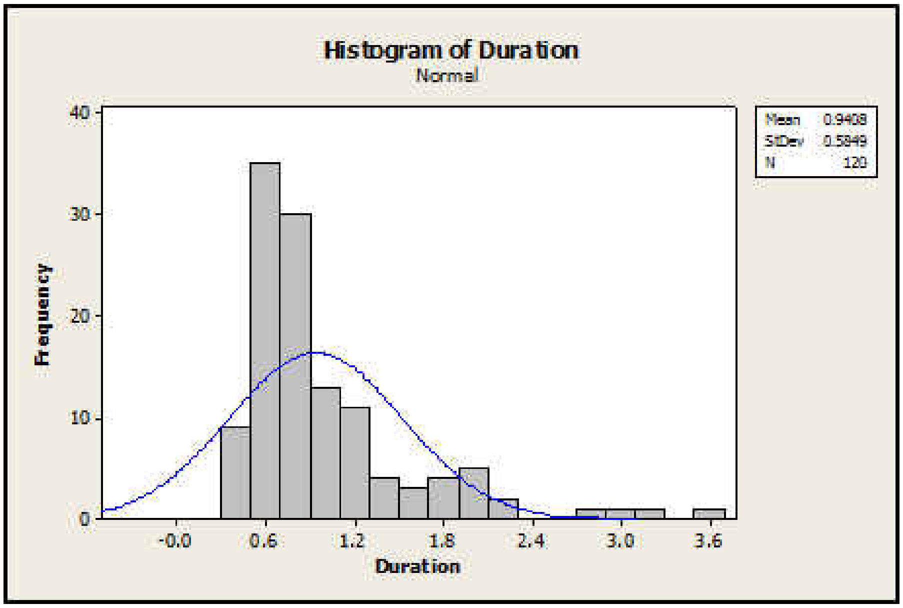

Figure 24: Histogram of Decide Workload Task Time and Distribution Shape by Scenario

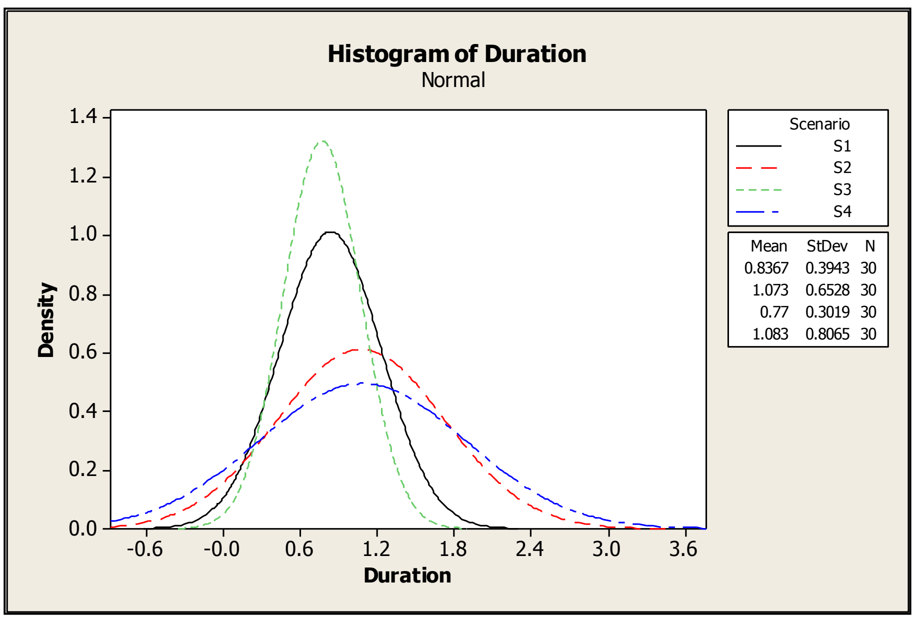

Figure 25: Distribution Shape of Decide Workload Task Time by Scenario

ANOVAs are performed to confirm this insight that Scenario 1 and Scenario 3 are from the same distribution, and that Scenario 2 and Scenario 4 are from the same distribution, but that the single- and dual-task scenarios are from different distributions. Figure 26 summarizes one-way ANOVA results for Decide Workload task time by each scenario. Note that there are 30 data points for each scenario. From the confidence intervals it is clear that the single-task scenarios are from the same distribution and the dual-task scenarios are from another distribution.

One-way ANOVA: Duration versus Scenario

```
Source     DF      SS     MS     F      P
Scenario    3   2.337  0.779  2.35  0.076
Error     116  38.373  0.331
Total     119  40.710

S = 0.5752   R-Sq = 5.74%   R-Sq(adj) = 3.30%


                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level   N    Mean   StDev   --+---------+---------+---------+-------
S1     30  0.8367  0.3943      (----------*---------)
S2     30  1.0733  0.6528                  (----------*---------)
S3     30  0.7700  0.3019   (---------*----------)
S4     30  1.0833  0.8065                   (---------*----------)
                            --+---------+---------+---------+-------
                            0.60      0.80      1.00      1.20

Pooled StDev = 0.5752
```

Figure 26: One-Way ANOVA for Decide Workload Time by Scenario

Since the single-task scenario and dual-task scenario distribution have some overlap, another ANOVA was run in which the data from Scenario 1 and Scenario 3 were combined, and the data from Scenario 2 and Scenario 4 were combined. This provides two datasets, each with 60 points. Figure 27 provides this revised ANOVA. The 95% confidence intervals provide a clear visual picture that reveals that single- and dual-task scenarios are from separate distributions. The ANOVA clearly rejects the null hypothesis of the two datasets being from the same distribution with a p-value of 0.009.

One-way ANOVA: Duration versus Scenario Type

```
Source          DF      SS     MS     F      P
Scenario Type    1   2.269  2.269  6.96  0.009
Error          118  38.441  0.326
Total          119  40.710

S = 0.5708   R-Sq = 5.57%   R-Sq(adj) = 4.77%


                              Individual 95% CIs For Mean Based on
                              Pooled StDev
Level    N    Mean   StDev   ------+---------+---------+---------+---
Dual    60  1.0783  0.7275                       (---------*---------)
Single  60  0.8033  0.3498   (---------*--------)
                             ------+---------+---------+---------+---
                                 0.75      0.90      1.05      1.20

Pooled StDev = 0.5708
```

Figure 27: One-Way ANOVA for Decide Workload Time by Scenario Type

Based on these ANOVAs, two distributions were fitted for the Decide Workload task time: one for the single-task scenarios and one for the dual-task scenarios. The best distribution fit for the single-task scenarios is a LogLogistic distribution, with a shape parameter $\alpha = 4.3682$ and a scale parameter $\beta = 0.72904$. The Kolmogorov-Smirnov p-value for this distribution is 0.3674. Figure 28 provides the histogram of the single-task Decide Workload distribution along with its fitted curve.

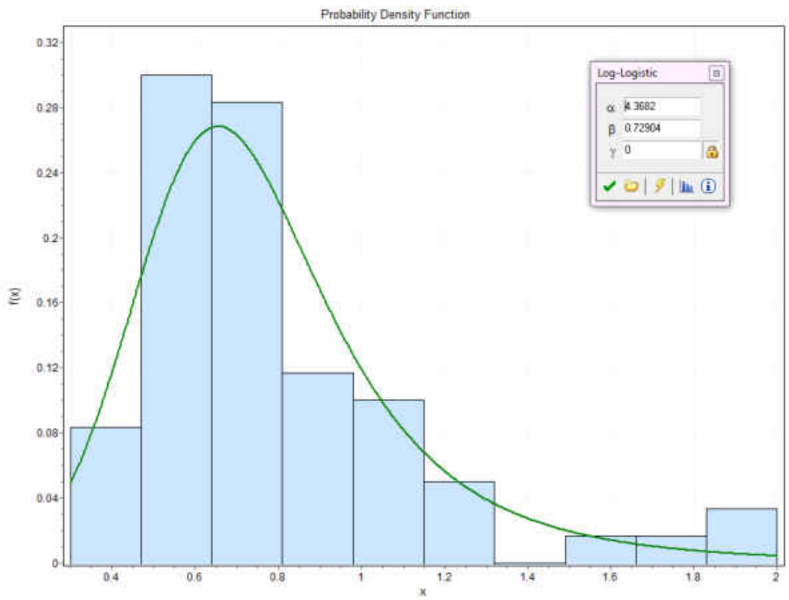

Figure 28: Fitted Distribution for Single-Task Scenarios Decide Workload

The best distribution fit for the dual-task scenarios is a Pearson5, with a shape parameter α = 3.7144 and a scale parameter β = 2.9176. The Kolmogorov-Smirnov p-value for this distribution is 0.3088. Figure 29 provides the histogram of the dual-task Decide Workload distribution along with its fitted curve.

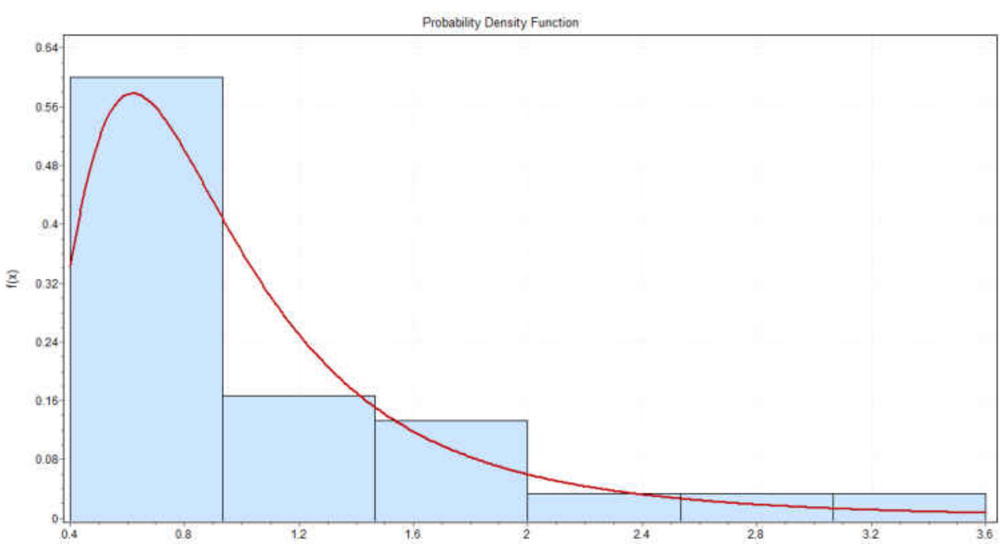

Figure 29: Fitted Distribution for Dual-Task Decide Workload

4.8.1.2. <u>Speak ISA Value Task Time</u>

Table 21 summarizes the descriptive statistics for the Speak ISA Value task. Figure 30 displays a histogram of the data, and Figure 31 displays the shape of the data distributions by scenario. A review of the descriptive statistics and the individual distribution shapes data reveals that Scenario 1 and Scenario 3, the single-task scenarios, have similar durations, and Scenario 2 and Scenario 4, the dual-task scenarios, have similar durations. However, the visual differences between these two potential distributions are not as pronounced as those of the Decide Workload task.

Table 21: Descriptive Statistics for Speak ISA Value Task

	Scenario 1: Single-Task Variable Change Detection	**Scenario 2: Dual-Task Variable Change Detection**	**Scenario 3: Single-Task Variable Threat Detection**	**Scenario 4: Dual-Task Variable Threat Detection**
Minimum	0.20	0.10	0.10	0.20
Maximum	0.40	0.40	0.40	0.40
Mean	0.24	0.27	0.25	0.27
Median	0.20	0.30	0.20	0.30
St Dev	0.06	0.07	0.07	0.07

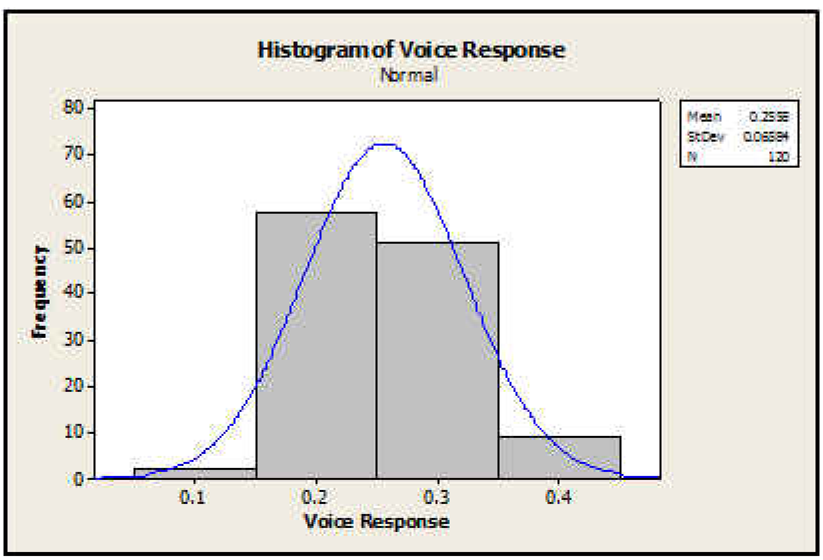

Figure 30: Histogram of Speak ISA Value Task Time and Distribution Shape by Scenario

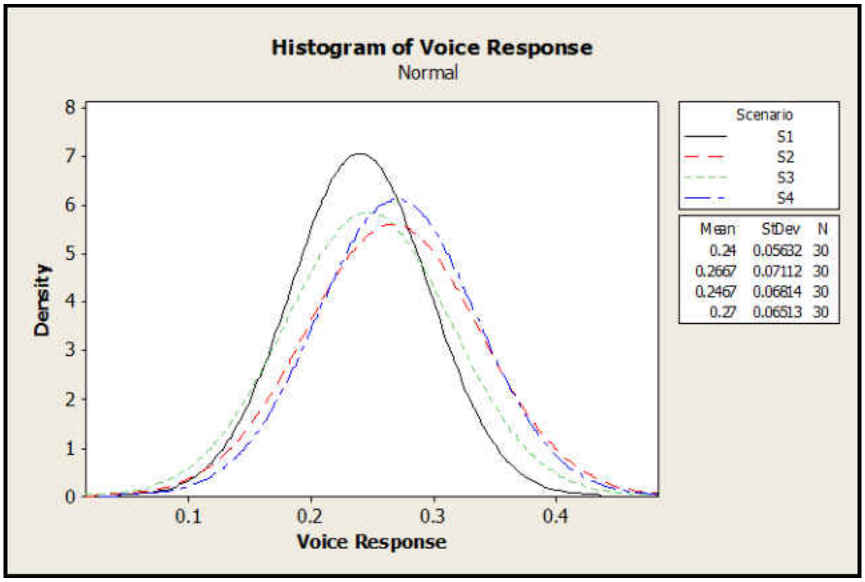

Figure 31: Distribution Shape of Speak ISA Value Task Time by Scenario

As with the Decide Workload task, ANOVAs are conducted to confirm if Scenario 1 and Scenario 3 are from the same distribution and Scenario 2 and if Scenario 4 are from the same distribution, but that the single- and dual-task scenarios are from different distributions. Figure 32 are the results from a one-way ANOVA for Decide Workload task time by each scenario. Note that there are 30 data points for each scenario. From the confidence intervals, it can be seen that the single-task scenarios are from the same distribution, and the dual-task scenarios are from another distribution.

One-way ANOVA: Voice Response versus Scenario

```
Source     DF       SS       MS     F      P
Scenario    3  0.01958  0.00653  1.53  0.212
Error     116  0.49633  0.00428
Total     119  0.51592

S = 0.06541   R-Sq = 3.80%   R-Sq(adj) = 1.31%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N     Mean    StDev  --+---------+---------+---------+-------
S1     30  0.24000  0.05632  (-----------*-----------)
S2     30  0.26667  0.07112               (----------*-----------)
S3     30  0.24667  0.06814     (----------*-----------)
S4     30  0.27000  0.06513                (-----------*-----------)
                             --+---------+---------+---------+-------
                             0.220     0.240     0.260     0.280

Pooled StDev = 0.06541
```

Figure 32: One-Way ANOVA for Speak ISA Value Take Time by Scenario

Since the single-task and dual-task distributions overlap, another ANOVA is performed in which the data from Scenario 1 and Scenario 3 are combined, and the data from Scenario 2 and Scenario 4 are combined. This provides two datasets, each with 60 points. Figure 33 summarizes the results from this revised ANOVA. The 95% confidence intervals reveal that single- and dual- task scenarios are from separate distributions. Hence, the null hypothesis stating that the two datasets are from the same distribution is rejected.

```
One-way ANOVA: Voice Response versus Scenario Type

Source          DF       SS       MS     F      P
Scenario Type    1  0.01875  0.01875  4.45  0.037
Error          118  0.49717  0.00421
Total          119  0.51592

S = 0.06491   R-Sq = 3.63%   R-Sq(adj) = 2.82%


                                 Individual 95% CIs For Mean Based on Pooled StDev
Level    N     Mean    StDev  ---------+---------+---------+---------+
Dual    60  0.26833  0.06763                    (----------*----------)
Single  60  0.24333  0.06207  (----------*----------)
                              ---------+---------+---------+---------+
                                     0.240     0.255     0.270     0.285

Pooled StDev = 0.06491
```

Figure 33: One-Way ANOVA for Speak ISA Value Take Time by Scenario Type

Based on these ANOVAs, two distributions are fitted for the Decide Workload task time: one for the single-task scenarios and one for the dual-task scenarios. Due to the discrete nature of the data, with only four possible values, the best distribution fit is a discrete empirical distribution. The probabilities assigned to each of the four values in this distribution are derived from the sample data, and are summarized in Table 22.

Table 22: Probabilities for Speak ISA Value Discrete Empirical Distribution

Scenario Type	0.1	0.2	0.3	0.4
Single-Task	2%	58%	35%	5%
Dual-Task	2%	38%	50%	10%

4.8.2. Change Detection Probability Distributions

For the Change Detection Tasks, the task times for identifying a change are derived from PRIME 2 data from the first 30 participants, while performing Scenario 1. Additionally, the probability of seeing a change and the probability of selecting a change type, given a particular type, are also derived from the participant data.

4.8.2.1. Seeing a Change Probability

In the Change Detection task network, after a change occurs, the entity representing the operator's task flow moves from the Monitor Map node to the See Change node. From this node, there is a probabilistic branching path logic that determines whether the operator returns to the Monitor Map node (i.e., "miss" seeing the change) or moves on to the Identify Change node. The second and third columns of Table 23 displays the non-response rate for the first 30 participants and all participants of the PRIME 2 study, respectively, by scenario and segment. Since participants are instructed to identify all changes they see, and to guess on the type of change if they are unsure, it is assumed that a non-response equates to the operator not seeing the change.

Table 23: Probability of Seeing Change – Participant Data

Scenario and Segment	Miss Rate 1st 30 Participants	Miss Rate All Participants	Probability Entered into Model	Average Model Miss Rate
S1_A	22%	24%	22%	22.5%
S1_B	23%	27%	20%	27.5%
S1_C	35%	38%	15%	39%
S2_A	37%	40%	37%	38%
S2_B	39%	43%	35%	41.5%
S2_C	49%	53%	30%	51%
S4_A	39%	40%	30%	39%
S4_B	40%	43%	32%	41%
S4_C	47%	50%	43%	48.5%

In the model, however, an operator can miss a change either due to the probabilistic logic in the See Change node or because the operator was occupied responding to a change, and thus was not available to see a new change occur. The fourth column of Table 23 displays the miss rate probabilities entered into the model in the See Change node, and the fifth column provides the average miss rate from 30 replications.

4.8.2.2. Identifying a Change Task Time

If the operator sees a change, the entity representing the operator's task flow moves from the See Change node to the Identify Change node. In this task, the operator is cognitively processing the change that was seen and attempting to identify it as an appearance, disappearance, or movement. To determine the amount of time required to identify a change, a random sample of 300 response times from the data for the first 30 participants of the PRIME 2 study are analyzed, for each segment type (A, B, and C). These response times are the duration between when a change event occurred and when a button was selected. Note that no response time can be calculated for cases where a change event occurred, but the participant did not respond, or where no change occurred but the participant did respond.

The best distribution fit for the identifying a change for Segment A is a LogNormal, with a mean μ = 1.64 and a standard deviation σ = 0.61. The Kolmogorov-Smirnov p-value for this distribution is 0.0977. Figure 34 provides the histogram of the Identify Change distribution for Segment A along with its fitted curve.

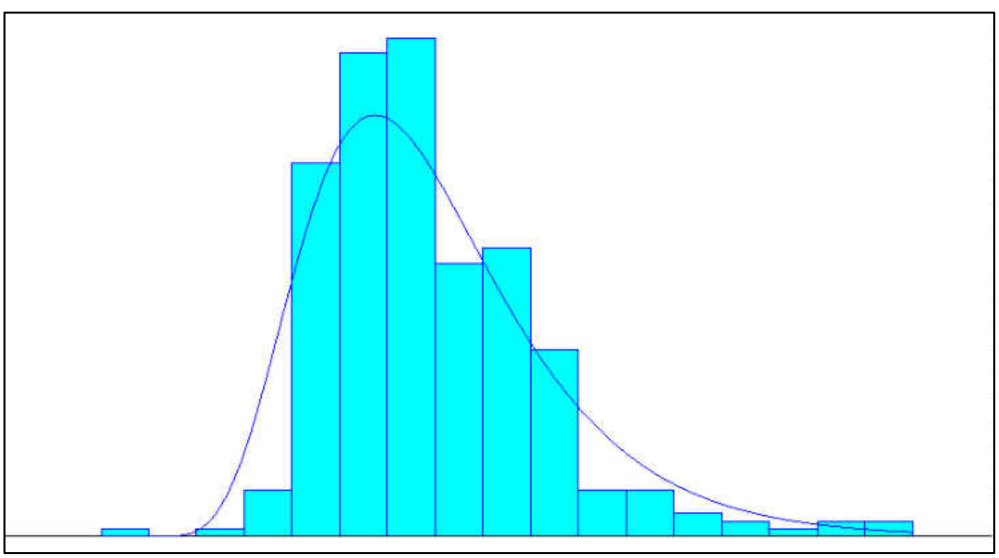

Figure 34: Fitted Distribution for Identifying a Change, Segment A

The best distribution fit for the identifying a change for Segment B is a LogLogistic, with a shape parameter α = 6.4634 and a scale parameter β = 1.3856. The Kolmogorov-Smirnov p-value for this distribution is 0.4678. Figure 35 is the histogram of the Identify Change distribution for Segment B along with its fitted curve.

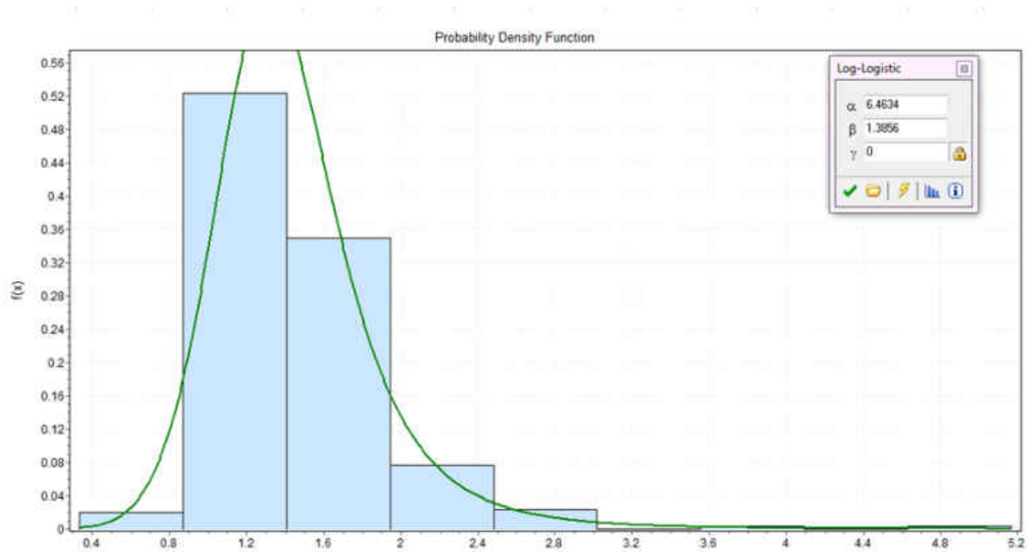

Figure 35: Fitted Distribution for Identifying a Change, Segment B

The response times for Segment C from the PRIME 2 study are bi-modal, as can be seen in Figure 36. After a close examination of the data, this multi-modal distribution appears to stem from participant responding to multiple changes at the same time.

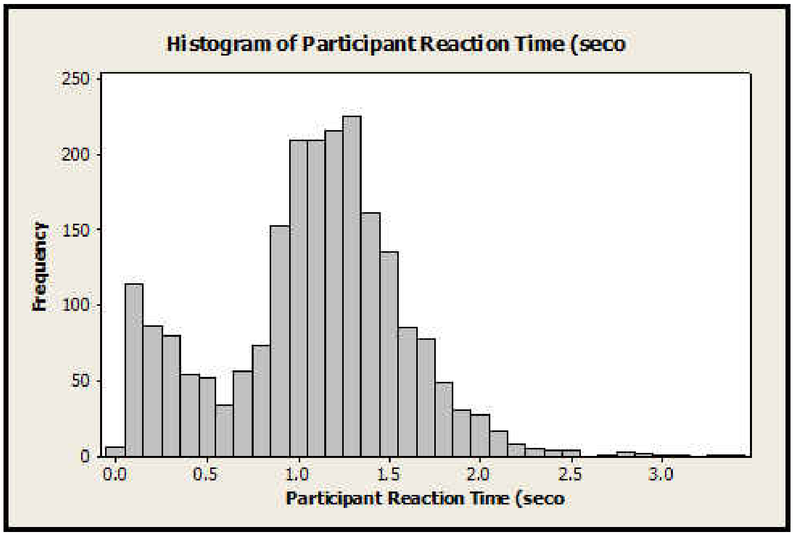

Figure 36: Bi-Modal Distribution of Participant Response Times for Identifying Change, Segment C

Figure 37 provides a representative section of the participant response log for Segment C. the first column displays the actual change event that occurred. The second column displays the participant's response to this change event, and the third column displays the evaluation of the participant's response (Correct, Incorrect, or Time Out). In rows three through five, the actual event sequence was a Movement, a Disappearance, and then No Event (participant hit a button, even though no change had occurred. For those events the participants responses were No Response, Movement, and Disappearance, respectively. Thus, the participant saw a Movement, followed by a Disappearance, and responded in that order, but due to his delay in responding, the participant selected the `MOVEMENT` button after the Disappearance had already occurred. This appears to also be the case for rows 12 through 15, where the actual event sequence is Movement, Appearance, Movement, which is how the participant responded. However, the participants responses are all paired in an off-set manner, such that these show up as incorrect. This occurrence is identifiable by the group start with a No Response (in the second column) and the group ending with a No Event (in the first column).

Event Type "string"	Event Response Type "string"	Event Result Type "string"	Participant Reaction Time (seconds)	Or
Icon Disappeared	No Response	Time Out	0	
Icon Movement	No Response	Time Out	0	
Icon Disappeared	Icon Movement	Incorrect	0.26	
No Event	Icon Disappeared	Incorrect	0	
Icon Appeared	No Response	Time Out	0	
Icon Movement	Icon Disappeared	Incorrect	0.74	
No Event	Icon Movement	Incorrect	0	
Icon Disappeared	Icon Disappeared	Correct	0.97	
Icon Appeared	Icon Appeared	Correct	1.21	
Icon Disappeared	No Response	Time Out	0	
Icon Movement	No Response	Time Out	0	
Icon Appeared	Icon Movement	Incorrect	0.32	
Icon Movement	Icon Appeared	Incorrect	0.5	
No Event	Icon Movement	Incorrect	0	
Icon Appeared	Icon Appeared	Correct	1.57	
Icon Disappeared	Icon Disappeared	Correct	1.15	
Icon Disappeared	Icon Disappeared	Correct	1.31	

Figure 37: Sample Change Detection Response Log, Segment C

Clearly, the data are capturing an unintended consequence of this segment's event rate. While the changes are designed to occur rapidly enough to cause high workload for Segment C, the change frequency is not intended to cause correct responses to be incorrect. This issue flagged the data as problematic, and thus it was decided to remove these sets of responses.

Less clear is the situation in rows 6 through 8, which also have the No Response-No Event pattern for starting and ending the group, but the actual event sequence is Appeared then Movement, whereas the participant responded Disappeared, then Movement. In this case, the task times are used to make a determination. Thus if at least 50% of the response sequence is correct and all response times are less than 1 second, the response set is determined to be a part of this problematic situation and the set is removed. Removal of these problematic response sets creates a relatively unimodal distribution, which is used for determining a distribution for Segment C's Identify Change times.

The best distribution fit for the identifying a change for Segment C is a Logistic, with a shape parameter $\alpha = 1.17$ and a scale parameter $\beta = 0.23$. The Kolmogorov-Smirnov p-value for this distribution is 0.10656. Figure 38 provides the histogram of the Identify Change distribution for Segment C along with its fitted curve.

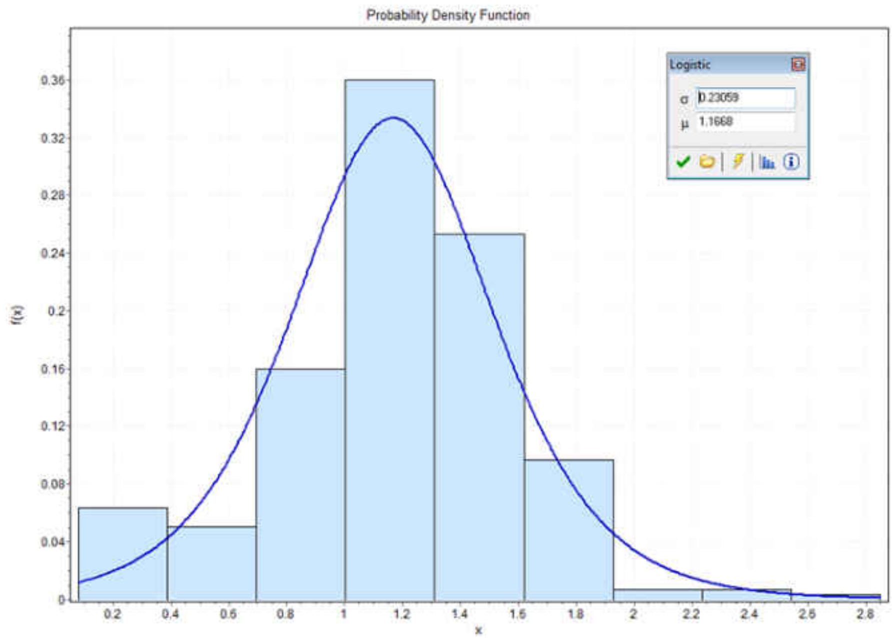

Figure 38: Fitted Distribution for Identifying a Change, Segment C

4.8.2.3. <u>Selecting a Change Type Probability</u>

Even if the operator sees a change that does not necessitate that the operator will identify that change correctly. While the operator is most likely to identify the change correctly, there is a possibility that the operator will identify the change incorrectly. Table 24 displays the probability of selecting a type of change, given a particular change type, by segment, for the first 30 participants while performing the Change Detection single-task scenario. For example, in Segment B, 11% of the participants selected the `MOVEMENT` button when in the actual change was an appearance. These same probabilities are used in the dual-task scenarios, and have been verified for reasonableness against all participant data and dual-task data.

Table 24: Probability of Identifying a Particular Type of Change for the First 30 Participants

Actual, Response	Segment A	Segment B	Segment C
Appear, Appear	92%	86%	69%
Appear, Disappear	1%	3%	11%
Appear, Move	7%	11%	20%
Disappear, Appear	1%	2%	12.4%
Disappear, Disappear	96%	96%	77.3%
Disappear, Move	3%	3%	10.3%
Move, Appear	16%	16%	23%
Move, Disappear	8%	11%	17%
Move, Move	76%	73%	60%

4.8.3. Threat Detection Probability Distributions

For the Threat Detection Tasks, the task times for selecting a threat or non-actor a change are derived from PRIME 2 data from the first 30 participants while performing the Scenario 3. Additionally, the probability of identifying an actor as a threat, the probability of pre-loading the `THREAT DETECT` button, and the probability of selecting a non-actor are also derived from data from the first 30 participants.

4.8.3.1. Identifying an Actor as a Threat Probability

During the Threat Detection task, the operator is presented with four types of actors: Friendly Soldiers, Friendly Civilians, Enemy Soldiers, and Armed Civilians. While only the latter two types of actors are threats, there is a possibility that the operator could detect any of these actor types as threats. Table 25 provides the probability of identifying a particular actor type as a threat, by segment, for the first 30 participants. For example, in Segment B, only 0.2% of Friendly Civilians are identified as threats, while 95.6% of Enemy Soldiers are identified as threats. These probabilities are used in the Identify Friendly Soldier, Identify Friendly Civilian, Identify Enemy Soldier, and Identify Armed Civilian nodes in both the single- and dual-task scenarios.

Table 25: Probability of Identifying an Actor as a Threat

	% Detected as Threats		
Actor Type	**Segment A**	**Segment B**	**Segment C**
Friendly Soldiers	0.2%	0.3%	0.1%
Friendly Civilians	0.1%	0.2%	0.2%
Enemy Soldiers	96.0%	95.6%	97.0%
Armed Civilians	94.2%	93.3%	93.0%

4.8.3.2. Pre-load Probability

During the PRIME 2 study, some participants realized that they could "pre-load" the `THREAT DETECT` button. That is, they could select the `THREAT DETECT` button, even though they were not ready to report a threat. This assists them, by allowing them to select the threat immediately once a threat appears. This pre-loading is accounted for in the model. Based on the analysis of the threat selection task time, it is determined that task times greater than 3.5 seconds would be considered "pre-loads." Table 26 contains the percentage of threat selection task times that are over 3.5 seconds, by segment, for the first 30 participants performing the Threat Detection single-task scenario. These probabilities are used in both the single-task and dual-task scenarios.

Table 26: Pre-load Probabilities by Segment

Segment	% Preloaded
Segment A	38%
Segment B	43%
Segment C	30%

4.8.3.3. Selecting a Non-Actor Probability

During the PRIME 2 study, participants would occasionally select a non-actor (e.g. terrain) when reporting a threat. The model captures this occurrence through probabilistic path logic after selecting the `THREAT DETECT` button that determines whether a threat or a non-actor is selected. Table 27 displays the percent of the total selections that are non-actors, by segment, for the first 30 participants performing the single-task Threat Detection scenario. These probabilities are used to determine the probabilistic path logic of selecting a threat or non-actor in both the single-task and dual-task scenarios.

Table 27: Probability of Selecting a Non-Actor by Segment

Segment	Probability of Selecting Non-Actor
Segment A	3.7%
Segment B	3.3%
Segment C	2.1%

4.8.3.4. Selecting a Threat or Non-Actor Task Time

After the `THREAT DETECT` button is selected, the operator then selects an actor (or non-actor, by mistake). This task includes locating the actor, tracking the actor (since the vehicle is moving through the simulated video feed), aligning the mouse cursor with the actor, and selecting the actor by clicking on the mouse button. To determine the time for this task, response times from the first 30 participants in the PRIME 2 study are used. These response times are the duration from when the `THREAT DETECT` button is selected to when an actor/non-actor is selected. As discussed above, participants would sometimes pre-load the `THREAT DETECT` button before they were ready to identify a particular actor as a threat. Thus, the response times also include these "pre-load" times. The pre-load times can be expected to be the longer durations. Figure 39 displays the distributions of response times across the three segments.

These graphs reveal a cluster of times to the far left of the graph, with a long tail to the right of the graph. These graphs show a bimodal distribution with the tail of the bimodal distribution consisting of an approximate uniform distribution of the pre-load times and the remaining observations representing actual response times. Examination of these distributions reveals that 3.5 seconds is an appropriate separation of these two distributions for each of the three segments.

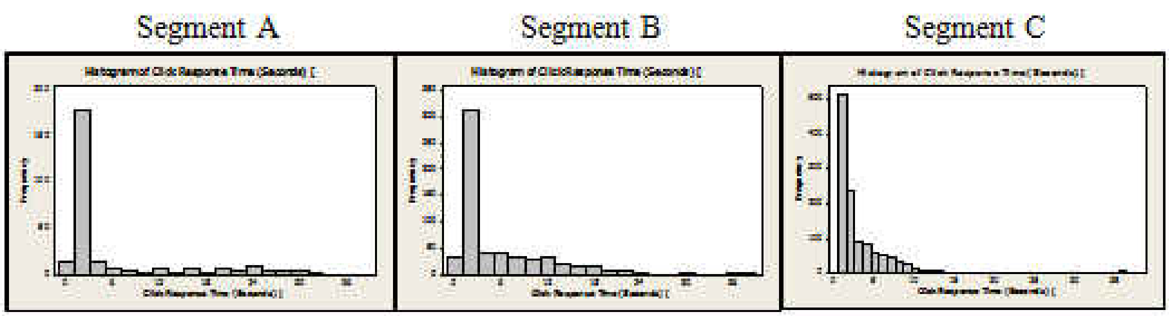

Figure 39: Identify Threat Response Times

Using times that are less than 3.5 seconds provides a sample of 185 data points for Segment A, 342 data points for Segment B, and 821 data points for Segment C. Distributions are then fit for the full data set for Segments A and B and for a random sample of 300 points for Segment C.

The best distribution fit for the selecting a threat for Segment A is a Pearson5, with a shape parameter $\alpha = 10.991$ and a scale parameter $\beta = 16.37$. The Kolmogorov-Smirnov p-value for this distribution is 0.97586. Figure 40 provides the histogram of the Select Threat distribution for Segment A along with its fitted curve.

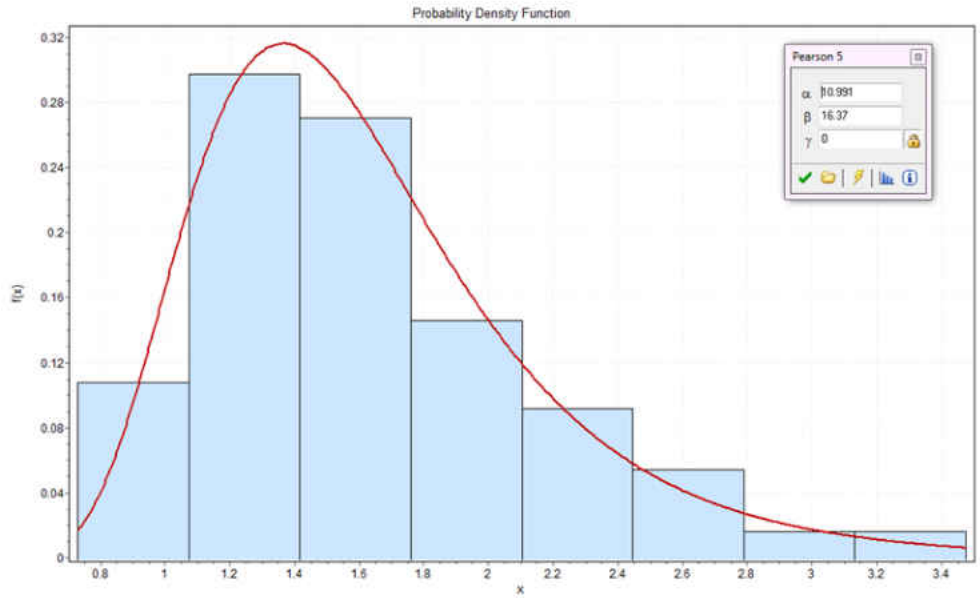

Figure 40: Fitted Distribution for Selecting a Threat, Segment A

The best distribution fit for the selecting a threat for Segment B is a Pearson5, with a shape parameter α = 7.7987 and a scale parameter β = 11.542. The Kolmogorov-Smirnov p-value for this distribution is 0.91436. Figure 41 provides the histogram of the Select Threat distribution for Segment B along with its fitted curve.

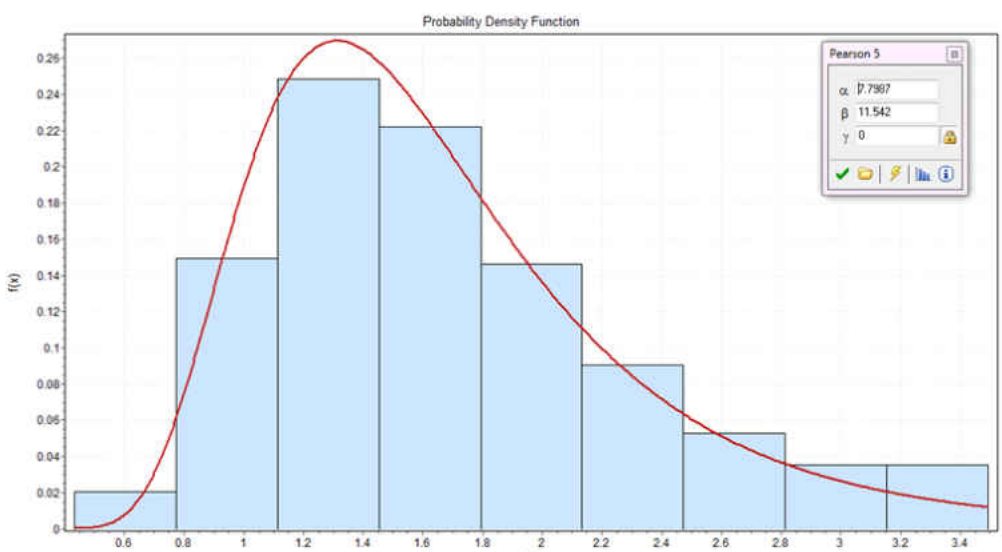

Figure 41: Fitted Distribution for Selecting a Threat, Segment B

The best distribution fit for the selecting a threat for Segment C is a Pearson5, with a shape parameter α = 7.2163 and a scale parameter β = 9.7574. The Kolmogorov-Smirnov p-value for this distribution is 0.45925. Figure 42 provides the histogram of the Select Threat distribution for Segment C along with its fitted curve.

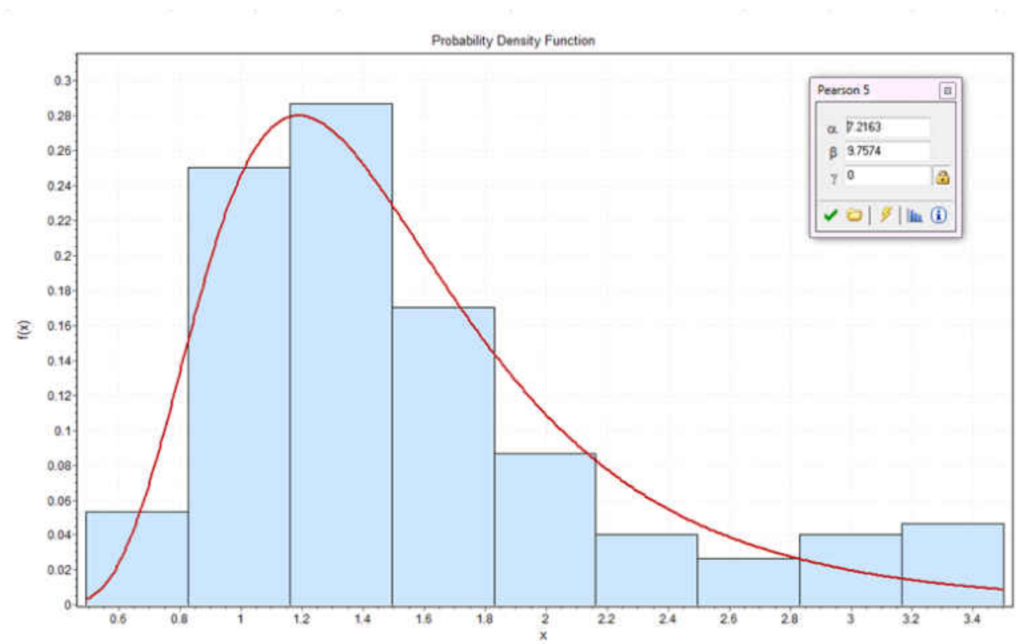

Figure 42: Fitted Distribution for Selecting a Threat, Segment C

4.9. Summary

This chapter describes the creation of the baseline DES model, which is based on a military unmanned vehicle supervisory-control task. The specific case modeled is taken from the PRIME 2 study, which allows for a detailed task analysis of the tasks performed. The PRIME 2 study also provides a rich data set, enabling task time probability distributions and probabilistic decision logic to be based on human participant data. The following chapter describes verification and validation of this baseline model, as well as the application of simulation to system design evaluation.

CHAPTER 5
BASELINE MODEL VERIFICATION, VALIDATION, AND ALTERNATIVE DESIGN CONFIGURATION EVALUATION

5.1. Baseline Model Verification

Verification of the model is conducted using peer walkthroughs with subject matter experts (SMEs). Two SMEs are chosen based on their extensive work on the PRIME 2 study, as well as previous HATS-ON studies. The SMEs are intimately familiar with the experimental design, the virtual environment development, and the scenario details. The SMEs also served as research technicians for the PRIME 2 study, and thus are responsible for experiment setup, participant training, and monitoring participant performance and physiological measurements during the experiments. This knowledge ensures that the SMEs are able to identify errors and assumptions of the baseline model's task network logic, operator behaviors, and workload value assignments. The SMEs provide a number of valuable insights during the peer reviews, with all recommendations being incorporated into the final baseline models.

5.2. Baseline Model Validation

In order to validate the workload values predicted by the DES model, these cognitive workload scores of the model are analyzed for correlation with the cognitive workload outputs obtained from the PRIME 2 study participants. The workload outputs from the PRIME 2 study participants include: (1) participant responses using popular, well-established subjective workload measures used in research and practice, and (2) physiological measures that are commonly used in research and practice as surrogate measures of cognitive workload.

The predicted workload values from the DES model are based the VACP model. Each sub-task is assigned a workload value for each of the relevant channels, based on the level of workload demand given by the VACP charts in Bierbaum, Szabo, and Aldrich (1989). These workload values are then summed within and across channels. The operator's total cognitive workload value is calculated continuously as tasks are performed, these values are then transformed into time-weighted averages for each scenario-segment combination. Each of the 12 scenario-segment combinations is simulated and run using 10 independent replications, with each of the 10 replications using a unique random number seed. Based on the low variability within variant compared to the variability between variants, it is determined that 10 replications are sufficient. The time-weighted averages from these 10 replications are then averaged to produce a score for each of the 12 scenario-segment combinations for use in the correlation analysis.

The subjective workload measures (and abbreviations) from the PRIME 2 study participants used in the correlation analysis include the:

- ISA mean score (ISA),
- NASA-TLX Mental Dimension mean score (NASA TLX Mental),
- NASA-TLX Temporal Dimension mean score (NASA TLX Temporal), and
- NASA-TLX Effort Dimension mean score (NASA TLX Effort).

The physiological measures (and abbreviations) collected from the PRIME 2 study participants include the:

- Heart Rate Variability (Inverse HRV),
- Eye Tracking Index of Cognitive Activity (ICA),

- TCD mean flow velocity for channel 1 (left side) (MeanVelocityCh1),
- TCD mean flow velocity for channel 2 (right side) (MeanVelocityCh2),
- Alpha EEG frequency at the Poz site (AlphaPozInverse),
- Theta EEG frequency at the Fz site (ThetaFz),
- Alpha EEG frequency for the occipital lobe (OccipitalAlphaInverse),
- Theta EEG frequency for the frontal lobe (FrontalTheta),
- fNIR mean rs02 for channel 1 (left side) (MeanRs02Ch1), and
- fNIR mean rs02 for channel 2 (right side) (MeanRs02Ch2).

For each of the subjective workload measures, each participant provides one value per scenario-segment combination. For the physiological measures, the continuous data are converted into a single time-average value per participant per scenario-segment combination. The final values for these physiological measures are then obtained by subtracting the respective resting baseline value, in order to obtain a score that represents the difference from the resting baseline. For each measure, the values from each participant are then averaged, in order to obtain a score for each of the 12 scenario-segment combinations. A correlation analysis of these data provides insight into how predictive each of these measures is in their ability to capture the relative required system-level cognitive workload.

As can be seen from Table 28, the ISA and three NASA TLX dimensions (i.e., Mental, Temporal, and Effort) are highly-correlated with each other, confirming that these measures are a valid standard of workload by which to evaluate the physiological measures, as well as the predictive ability of the baseline model. The correlation analysis also shows that the Inverse HRV physiological measure is the best indicator of workload. The baseline model is highly-correlated (over 0.90) with the ISA and the three NASA TLX dimensions subjective measures, as well as the Inverse HRV physiological measure. This strong correlation not only validates the predicted workload values of the baseline model, but it also suggests that the DES model is equally-predictive of system-level cognitive workload as Inverse HRV and more predictive of system-level workload than all of the other commonly used physiological measures evaluated.

Table 28: Correlation Table

	ISA	NASA TLX Mental	NASA TLX Temporal	NASA TLX Effort	Inverse HRV	ICA	MeanVelocityCh1	MeanVelocityCh2	AlphaPozInverse	ThetaFz	OccipitalAlpha Inverse	FrontalTheta	MeanRs02Ch1	MeanRs02Ch2
NASA TLX Mental	0.983**													
NASA TLX Temporal	0.977**	0.998**												
NASA TLX Effort	0.982**	0.999**	0.996**											
Inverse HRV	0.945**	0.962**	0.952**	0.96**										
ICA	0.229	0.379	0.404	0.358	0.352									
MeanVelocityCh1	-0.531	-0.617*	-0.64*	-0.602*	-0.507	-0.633*								
MeanVelocityCh2	0.347	0.364	0.376	0.342	0.283	0.317	-0.253							
AlphaPozInverse	0.496	0.352	0.32	0.369	0.385	-0.646*	0.254	-0.028						
ThetaFz	-0.626*	-0.571	-0.547	-0.577	-0.665*	0.252	0.156	0.116	-0.714**					
OccipitalAlpha Inverse	0.396	0.233	0.203	0.246	0.237	-0.732**	0.282	-0.172	0.921**	-0.618*				
FrontalTheta	0.668*	0.766*	0.766*	0.757**	0.826**	0.685*	-0.654*	0.225	-0.098	-0.332	-0.251			
MeanRs02Ch1	0.335	0.477	0.504	0.465	0.398	0.900**	-0.652*	0.468	-0.578*	0.267	-0.693*	0.621*		
MeanRs02Ch2	0.486	0.605*	0.633*	0.589*	0.517	0.846**	-0.687*	0.608*	-0.424	0.147	-0.559	0.649*	0.969**	
Baseline Predicted WL	0.929**	0.965**	0.966**	0.96**	0.919**	0.545*	-0.646*	0.386	0.201	-0.415	0.091	0.777	0.615*	0.716**

*p-value <.05

**p-value <.01

5.3. Evaluation of Alternative System Design Configurations

Now that the expected workload is known for the baseline system, DES can be used to evaluate system configuration re-design alternatives for this system, in order to establish which, if any, provide statistically-significant cognitive workload improvements. Three system re-design alternatives to the baseline system are modeled, where each design is expected to have lower cognitive workload than the baseline system. However, the actual level of improvement of each alternative is unknown. Simulation of the designs enables them to be evaluated against the baseline system, as well as each other, for differences in cognitive workload.

5.3.1. Alternative Design 1: Keyboard Response in Change Detection Task

The first alternative design is to replace the mouse response of the Change Detection task with a keyboard response. Currently, the user uses the mouse to select the response that corresponds with the type of change that occurs. Thus, in addition to pressing the left mouse button, the user also tracks and aligns the mouse cursor with the appropriate button the screen. By changing this to a keyboard task, with a different key for each of the three responses, the user simply presses the corresponding button. The Threat Detection task would remain unchanged. This design uses the `A` key, `S` key, `D` key to correspond to an appearance, a movement, and a disappearance, respectively. These keys are selected because a natural typing posture includes resting the fingers of the left hand on these three keyboard keys. By choosing left-handed keys, the right hand is free to rest on the mouse for use during the Threat Detection task during the dual-task scenarios.

5.3.2. Alternative Design 2: Voice Response in Change Detection Task

Alternative Design 2 features the incorporation of voice-recognition software. With this system, the user would perform the Change Detection task by speaking the type of change that has occurred instead of using the mouse to track and align the mouse cursor with the appropriate button on the screen. The Threat Detection task would remain unchanged. It is assumed for this alternative that the times for the oral response are the same as the audio response times from the ISA task from the PRIME 2 study.

5.3.3. Alternative Design 3: Touchscreen Response in Both Change Detection and Threat Detection Tasks

The third alternative design replaces the current system with a touchscreen system. For the Change Detection task, the user touches the button on the screen with his or her finger. For the Threat Detection task, the user selects the threat by touching on the threat on the screen. This design eliminates the need for the `THREAT DETECT` button. The DES model for the design assumes that the Change Detection buttons remain in the same place as in the baseline system, the touchscreen is the same size as the monitor used in the PRIME 2 study, the user is right-handed, and the user rests his or her right hand along the center right edge of the screen. The task times used in this model for selecting the buttons and threats are calculated using Welford's variant of Fitt's Law, i.e.,

$$0.10 \times \log_2(P_1/P_2 + 0.50), \tag{1}$$

where P_1 is the distance (in cm) between the targets and P_2 is the size (in cm) of the target (Welford, 1968). Thus, the Change Detection task times are calculated from the size of the buttons and their distance from the right center edge of the screen. Recall that the user is motivated to select the threat as soon as identifiable in order to select the threat before is no longer visible on the screen, and the images of threat actors grow as it gets closer to the ground vehicle. Therefore, the task times are determined using a discrete empirical distribution based on the actors' distance from the resting place of the right hand.

5.4. Alternative System Designs: Results and Discussion

Figure 43 shows the analysis of variance (ANOVA) results for the first alternative design assuming an $\alpha = 0.05$ level of significance. In the ANOVA, each workload value in the 10 replications per system variant of the baseline DES model is subtracted from the respective replication in the alternative design. Thus, a significant difference in workload is for those system variants whose confidence interval does not include zero. A negative value indicates that the alternative design has lower estimated cognitive workload than the baseline DES model. Alternative Design 1 does not alter the Threat Detection task, thus the system variants for Scenario 3 (Threat Detection only) are not included.

As can be seen from the figure, this alternative design only provides a significant workload improvement for Segment A of Scenario 4, which corresponds to the dual-task scenario containing a medium event rate for Change Detection and a low event rate for Threat Detection.

```
                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level      N     Mean   StDev  ------+---------+---------+---------+---
A1S1_A  10  -0.1170  0.0083            (----*----)
A1S1_B  10  -0.1191  0.1255            (----*----)
A1S1_C  10  -0.2331  0.0554           (----*----)
A1S2_A  10  -0.0323  0.1216             (----*----)
A1S2_B  10   0.0853  0.1916              (----*----)
A1S2_C  10   0.1126  0.1414              (----*----)
A1S4_A  10  -1.1322  1.3600  (----*----)
A1S4_B  10  -0.2032  1.3366           (----*----)
A1S4_C  10   1.8234  1.4106                              (----*----)
                             ------+---------+---------+---------+---
                                -1.0       0.0       1.0       2.0

Pooled StDev = 0.7969
```

Figure 43: ANOVA Results for Alternative Design 1 (level of significance $\alpha = 0.05$)

Figure 44 summarizes the ANOVA results for the second alternative design for the difference from baseline DES model. Alternative Design 2 also does not alter the Threat Detection task; thus, the system variants for Scenario 3 (Threat Detection only) are not included. This design also only provides a significant workload improvement for Segment A of Scenario 4.

```
                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level      N     Mean   StDev  ---+---------+---------+---------+-----
A2S1_A  10  -0.0482  0.1017           (---*---)
A2S1_B  10  -0.0489  0.0914           (---*---)
A2S1_C  10  -0.1085  0.0771          (---*---)
A2S2_A  10  -0.0069  0.2092           (---*---)
A2S2_B  10   0.1153  0.1927            (---*---)
A2S2_C  10   0.1827  0.1505             (---*---)
A2S4_A  10  -1.1283  1.2434  (---*---)
A2S4_B  10  -0.1723  1.2565          (---*---)
A2S4_C  10   1.9312  1.4160                          (---*---)
                             ---+---------+---------+---------+-----
                             -1.2       0.0       1.2       2.4

Pooled StDev = 0.7644
```

Figure 44: ANOVA Results for Alternative Design 2 (level of significance $\alpha = 0.05$)

Figure 45 summarizes the ANOVA results for the third alternative design for the difference from the baseline DES model. This design provides a significant workload improvement for Segment A of Scenario 1 (i.e., a low event rate for Change Detection), all of the segments in Scenario 2 (variable event rate for Change Detection, constant medium event rate for Threat Detection), Scenario 3 (Threat Detection only) and Scenario 4 (constant medium event rate for Change Detection, variable event rate for Threat Detection).

```
                             Individual 95% CIs For Mean Based on Pooled StDev
Level     N     Mean  StDev    -+---------+---------+---------+--------
A3S1_A  10  -0.0110  0.0009                                 (-*)
A3S1_B  10   0.1217  0.1367                                    (*-)
A3S1_C  10   0.2104  0.0528                                     (-*-)
A3S2_A  10  -0.8183  0.1546                 (-*)
A3S2_B  10  -0.6472  0.2110                    (-*-)
A3S2_C  10  -0.4565  0.1238                        (-*-)
A3S3_A  10  -0.3261  0.0521                           (*-)
A3S3_B  10  -0.7368  0.1258                   (*-)
A3S3_C  10  -1.4638  0.1328    (-*)
A3S4_A  10  -0.2768  0.1404                            (*-)
A3S4_B  10  -0.6518  0.1895                    (-*-)
A3S4_C  10  -1.4780  0.1658    (*-)
                               -+---------+---------+---------+--------
                             -1.50     -1.00     -0.50      0.00

Pooled StDev = 0.1368
```

Figure 45: ANOVA for Alternative Design 3 (level of significance $\alpha = 0.05$)

The simulation of these alternative models reveals some unexpected results. All three alternatives are expected to have lower workload; however, only Alternative Design 3 results in consistently lower cognitive workload. While performance analysis is beyond the scope of this evaluation, it must be noted that each of these alternatives improves the overall performance in terms of the response rate. The explanation for the lack of workload improvement in the first two alternative designs is that the time saved by the design changes results in an increase in response by the user. Thus, the result is an improvement in performance, while maintaining the level of workload. Only Alternative Design 3 is able to achieve improved performance and decreased workload. Thus, evaluation of these alternative designs shows that Alternative Design 3 is the best design for reducing workload. These unanticipated results demonstrate the valuable insights provided by DES when it is used for human cognitive workload modeling and analysis.

5.5. Summary

This chapter addresses the research Sub-Question 1, "Can simulation modeling predict cognitive workload as well as established measures of cognitive workload?" and research Sub-Question 2, "Can computer simulation modeling be used to evaluate system designs based on predicted cognitive workload?" The baseline model validation demonstrates that discrete event simulation is a viable alternative to live trials involving human participants. The validation reveals that the simulations results are as predictive of time-weighted average cognitive workload as well-established subjective methods such as the ISA and NASA-TLX, and it is more predictive than many surrogate physiological measures. The alternative system design evaluations then demonstrate that discrete event simulation can be used to compare the relative workload differentials that different design alternative produce, allowing for the identification of a preferred design solution without the costly and time-consuming burden of prototyping and live field testing.

CHAPTER 6
MODELING MULTIPLE CHANNEL INTERFERENCE AND ADAPTIVE AUTOMATION

6.1. Incorporating Multiple Resource Channel Interference

The baseline DES model described in CHAPTER 4 and validated in CHAPTER 5 uses the VACP model to capture the workload values experienced across the various resource channels, and then sums these individual values to produce a total workload score, which is then converted to a time-weighted average. As discussed in CHAPTER 2, one of the limitations of the VACP model is that it does not account for interference between resource channels. To account for this interference, the baseline DES model is revised and enhanced through the use of the conflict value matrix shown in Table 29.

Table 29: Baseline DES Model Conflict Matrix (all matrix values are derived from Wickens, 2002)

	Conflict Values					
RI Pairs	R: Auditory I: ISA Interface	R: Cognitive I: Computer Monitor	R: Cognitive I: ISA Interface	R: Fine Motor I: Computer Mouse	R: Speech I: ISA Interface	R: Visual I: Computer Monitor
R: Auditory I: ISA Interface	0.70	0.40	0.40	0.20	0.40	0.20
R: Cognitive I: Computer Monitor		0.90	0.80	0.10	0.50	0.50
R: Cognitive I: ISA Interface			0.70	0.10	0.50	0.50
R: Fine Motor I: Computer Mouse				0.70	0.20	0.10
R: Speech I: ISA Interface					1.00	0.10
R: Visual I: Computer Monitor						0.80

This matrix is derived from the work of Wickens (2002) and is tailored to the specific tasks. Thus, these values account for information such as the involvement of verbal or tonal information in auditory tasks, the consideration of spatial information in cognitive channels, the presence of verbal information in the fine motor channel, and the involvement of verbal information and/or foveal fixation in the visual channel.

For a series of tasks completed by the operator, a total workload score can be computed. First, the intra-channel conflicts must be determined by summing the total workload for each channel. That sum is, then, multiplied by the respective intra-conflict value (obtained from Table 29) and the total number of conflicts for the channel. It is important to note that the total number of conflicts is n_i-1, where n_i is the number of tasks using channel i. So, for instance, if only one task uses a particular resource channel, then the number of conflicts is zero. If two tasks use the same channel simultaneously, then the number of conflicts is 1. Each intra-channel conflict is then summed to obtain the total intra-channel conflict.

Next, the inter-channel conflict value for each resource channel-pair must be computed by summing the workload values of each channel pair and then multiplying that sum by the conflict value for that resource channel pair. The inter-channel conflict values for each channel pair are then summed to obtain the total inter-channel conflict. Finally, the total conflict value is added to the sum of the workload demand values to compute the total cognitive workload score, *TWL*.

Before presenting the mathematical formulation for computing the total cognitive workload, the relevant notation and parameters are presented. The series of equations (Eq. 2 through Eq. 7) for computing total cognitive workload, *TWL*, with the consideration of interference follows.

Notation and Parameters:

- $\mathbf{T}$: the set of tasks performed by the operator, where $t = 1, \ldots, |\mathbf{T}|$;
- $\mathbf{R}$: the set of resource channels, where 1 denotes the Visual channel, 2 denotes the Auditory channel, 3 denotes the Cognitive channel, 4 denotes the Fine Psychomotor channel, 5 denotes the Gross Psychomotor channel, 6 denotes the Speech channel, and 7 denotes the Tactile channel, where $i = 1, \ldots, |\mathbf{R}|$;
- $\mathbf{P}$: the set of resource channel pairs (i,j), where $i, j = 1, \ldots, |\mathbf{R}|$ and $i \leq j$;
- WL_{ti}: the workload demand for task t for resource channel i, using VACP scores from Bierbaum (1989);
- $WL_{(i,j)}$: workload demand for resource channel pair (i,j), where $i, j = 1, \ldots, |\mathbf{R}|$ and $i \leq j$;
- C_i: the number of conflicts for channel i, where $i = 1, \ldots, |\mathbf{R}|$; $C_i := n_i - 1$, where n_i is the number of tasks using channel i;
- $CV_{(i,j)}$: the conflict value for channel pair (i, j) (using Table 29), where $i, j = 1, \ldots, |\mathbf{R}|$ and $i \leq j$;
- C_{Intra_i}: the intra-channel conflict for channel i, where channel $i = 1, \ldots, |\mathbf{R}|$; and
- $C_{Inter_{(i,j)}}$: the inter-channel conflict for channel pair (i,j), where $i, j = 1, \ldots, |\mathbf{R}|$ and $i \leq j$.

$$C_{Intra_i} = \left(\sum_{i \in \mathbf{R}} WL_{ti}\right) CV_{(i,j)} C_i \qquad \forall t \in \mathbf{T}; \forall i \in \mathbf{R}; \forall (i,j) \in \mathbf{P}; i = j \quad (2)$$

$$TC_{Intra} = \sum_{i \in \mathbf{R}} C_{Intra_i} \quad (3)$$

$$C_{Inter_{(i,j)}} = \left(\sum_{(i,j) \in \mathbf{P}} WL_{(i,j)}\right) CV_{(i,j)} \qquad \forall (i,j) \in \mathbf{P} \quad (4)$$

$$TC_{Inter} = \sum_{(i,j)\in\mathbf{P}} C_{Inter_{(i,j)}} \tag{5}$$

$$TC = TC_{Intra} + TC_{Inter} \tag{6}$$

$$TWL = \left(\sum_{t\in\mathbf{T}} \sum_{i\in\mathbf{R}} WL_{ti}\right) + TC \tag{7}$$

where TC_{Intra} is the total intra-channel conflict, TC_{Inter} is the total inter-channel conflict, TC is the total conflict value and TWL is the total workload value.

The incorporation of the channel interference conflict values increases the total workload values, which produces large spikes in workload at the beginning of each segment for Scenarios 2, 3, and 4. These spikes are due to the large number of actors that appear on the screen at the beginning of a segment during the Threat Detection task. Figure 46 through Figure 49 display these total workload values. Note that the workload scale for Scenarios 2, 3, and 4 differ from Scenario 1.

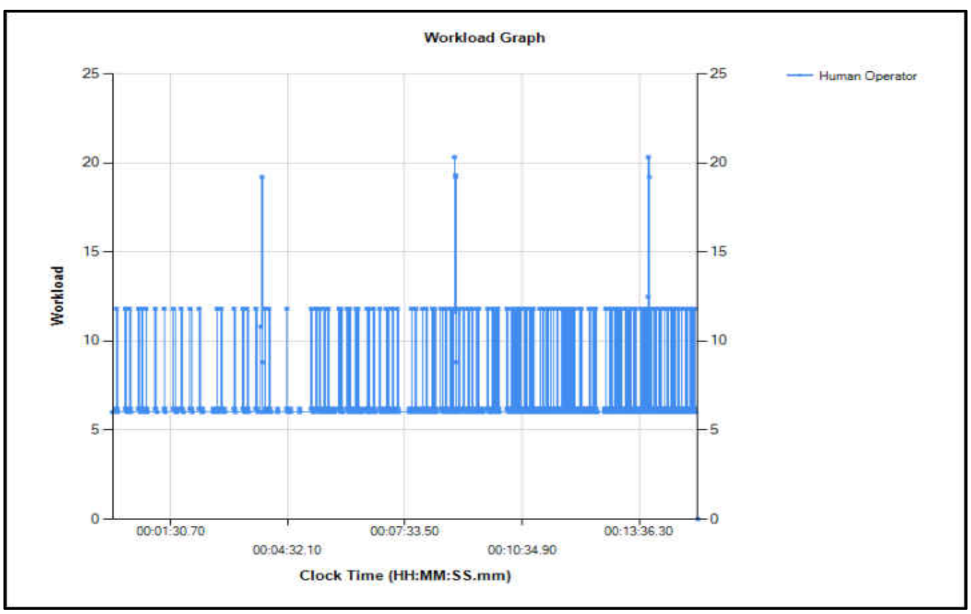

Figure 46: Workload Graph with Interference, Scenario 1

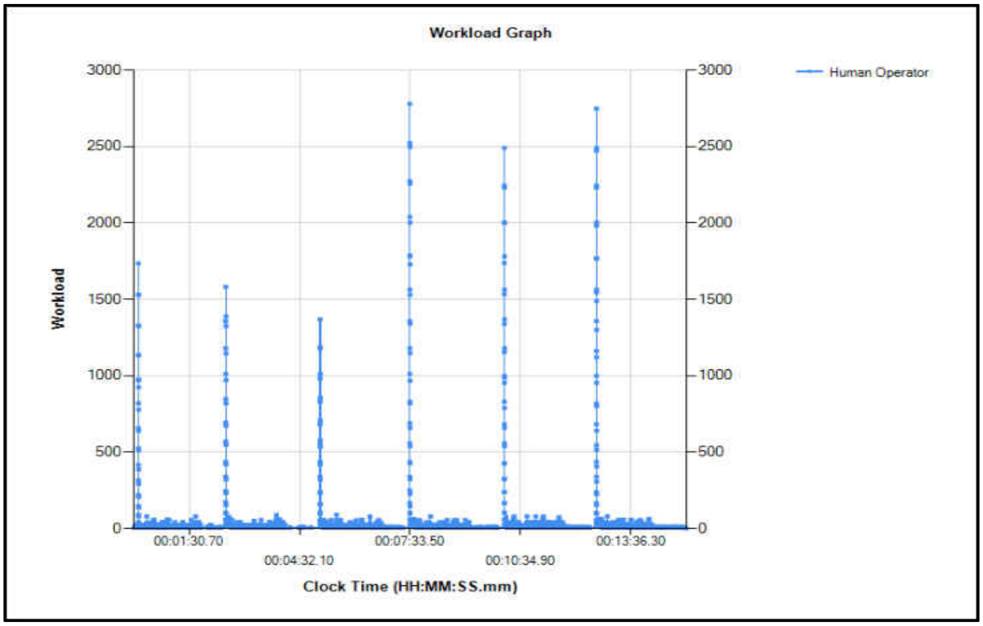

Figure 47: Workload Graph with Interference, Scenario 2

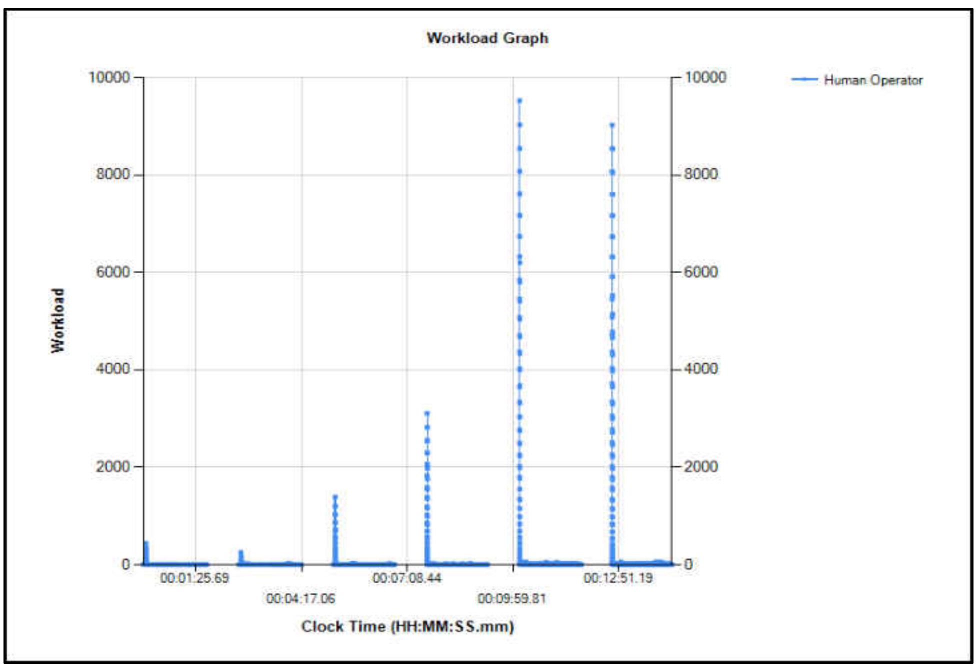

Figure 48: Workload Graph with Interference, Scenario 3

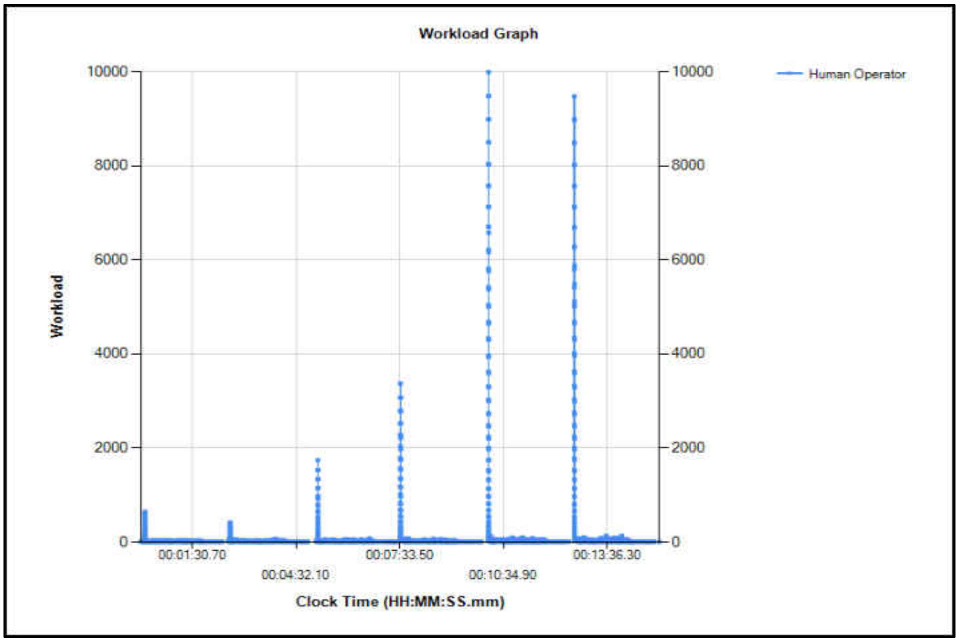

Figure 49: Workload Graph with Interference, Scenario 4

These workload values seem unrealistically high. Therefore, it is reasonable to assume that an operator faced with such a high workload would respond by doing these tasks serially, rather than in parallel. This is addressed by incorporating a workload management strategy into the DES models with channel interference that supports performing tasks serially when above a particular threshold. Analysis of the workload data from the enhanced DES model shows that less than 20% of the workload values are over 80; thus, this value is selected as the threshold for the workload management strategy. Figure 50 through Figure 53 display the updated total workload values once the new workload management strategy is incorporated. The inclusion of this strategy produces more reasonable operator workload levels.

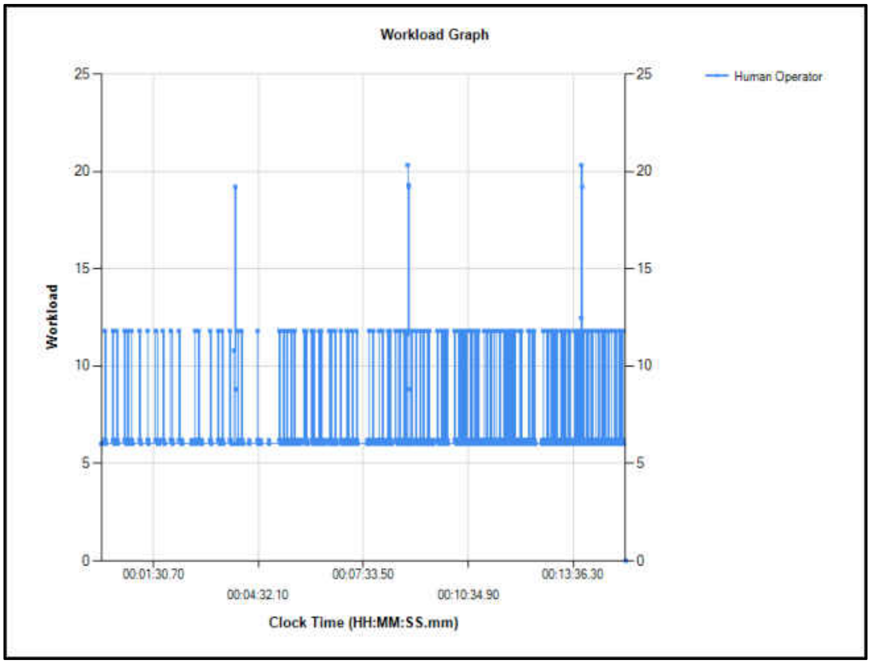

Figure 50: Workload Graph with Interference and Workload Management Strategy, Scenario 1.

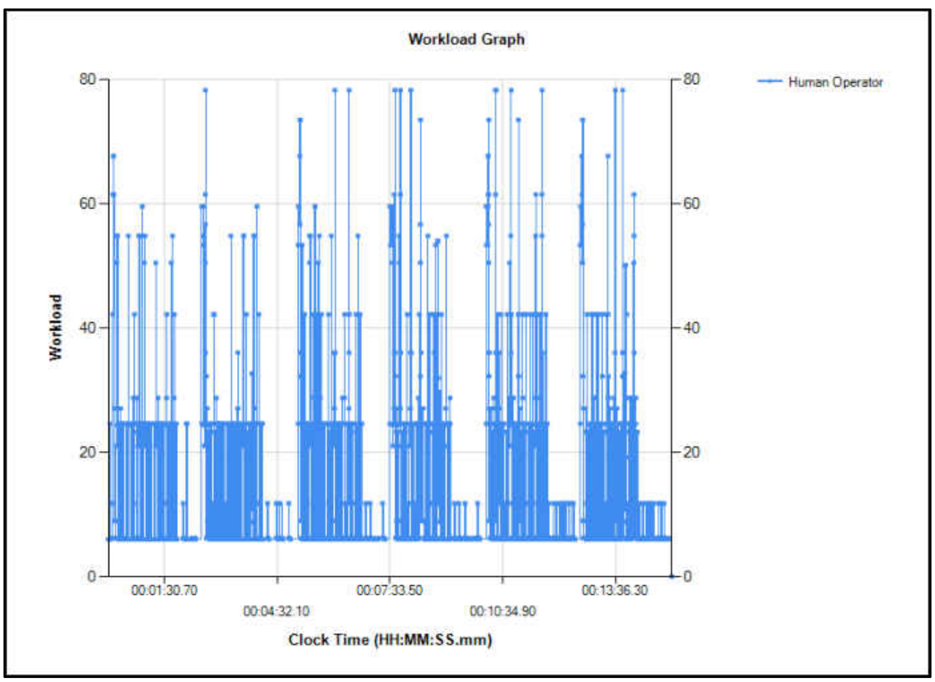

Figure 51: Workload Graph with Interference and Workload Management Strategy, Scenario 2

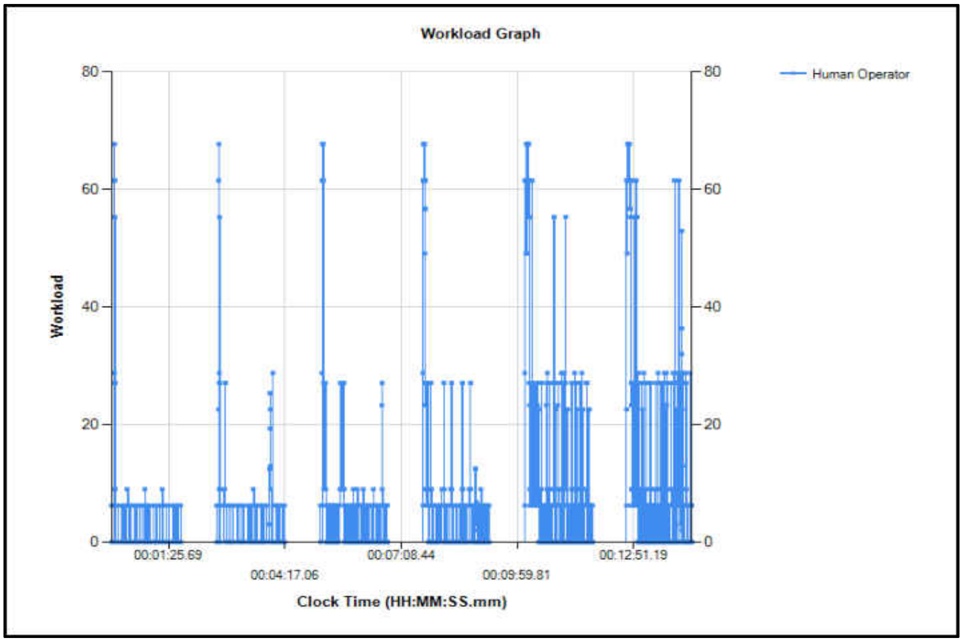

Figure 52: Workload Graph with Interference and Workload Management Strategy, Scenario 3

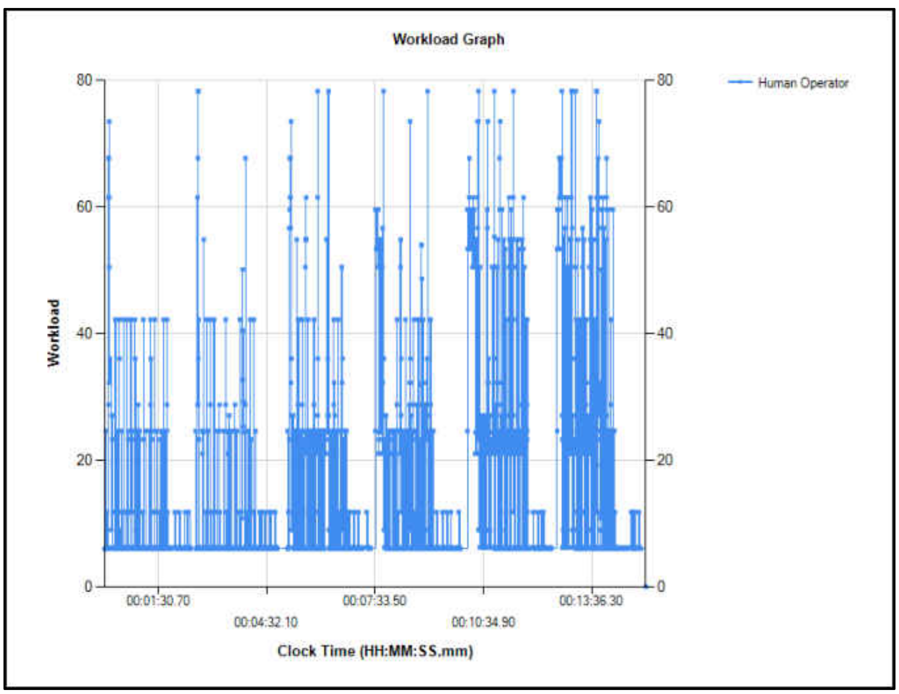

Figure 53: Workload Graph with Interference and Workload Management Strategy, Scenario 4

A correlation analysis is performed in order to assess the impact of these enhancements to the baseline DES model on the validation of the model. Using the ISA scores from the PRIME 2 study, which is shown in CHAPTER 5 to be a valid predictor of cognitive workload, the updated model has a 0.927 correlation with the ISA scores, compared with the 0.929 correlation from the original baseline DES model presented in CHAPTER 5. This similarity in correlation shows that the enhanced DES models are also valid representations of the baseline system.

6.2. Workload-Performance Analysis

Before updating the baseline DES model to include adaptive automation, it is first important to understand the relationship between the workload experienced by the user and the user's performance.

Figure 54 and Figure 55 are graphs of the average ISA score versus performance from the Prime 2 study for the Change Detection tasks and Threat Detection tasks, respectively. Each point on the graphs represents a single scenario-segment combination (e.g., Scenario 1 Segment A) with the scores from that segment averaged from the 150 participants in the Prime 2 study. Figure 54 shows the percent of change identified correctly during the Change Detection task, which is computed by dividing the number of changes identified correctly by the participant by the total number of changes in that scenario-segment combination. From the figure, the relationship between workload and performance is relatively linear and decreasing. Thus, increased workload leads to decreased performance. This trend aligns with Section C of the workload-performance curve of Cassenti et al. (2011) (Figure 4, Section 2.7.3). Figure 55 shows the percent of threats identified correctly during the Threat Detection task, and is computed by dividing the number of threats identified by the total number of threats in each scenario-segment combination. Similar to that during the Change Detection task, the relationship between workload and performance is relatively linear and decreasing during the Threat Detection task. So, it can be concluded that increased workload leads to decreased performance, which is intuitive.

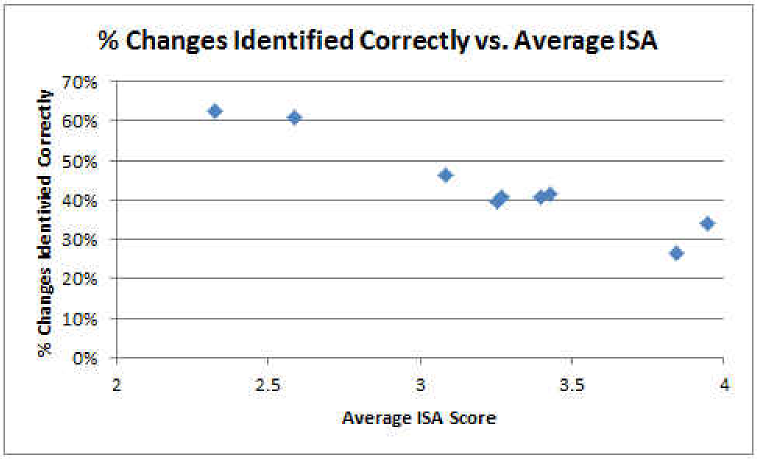

Figure 54: Average ISA Score vs. Percent of Changes Identified Correctly

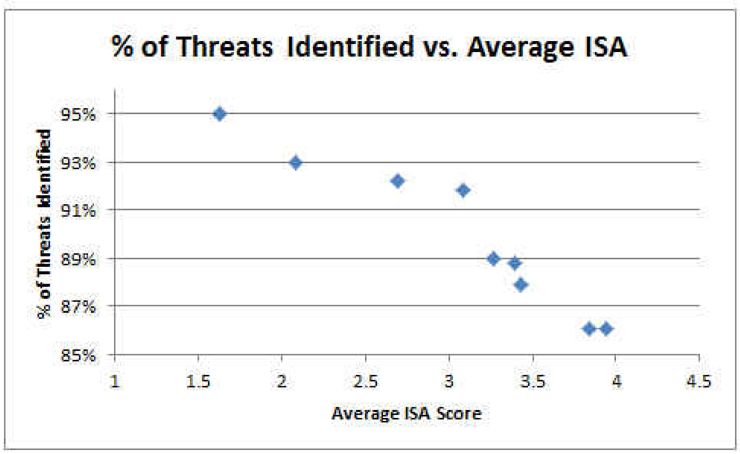

Figure 55: Average ISA Score vs. Percent of Threats Identified

Figure 56 and Figure 57 graph the predicted workload from the enhanced baseline DES model which includes interference versus the Percent of Changes Identified Correctly and Percent of Threats Identified performance measures, respectively. Predicted workload is computed by averaging the time-averaged total workload score for each scenario-segment combination across 10 independent replications. As can be seen from the two figures, the workload-performance relationship using the workload predicted by the DES model is still relatively linear and decreasing; thus, increased workload leads to decreased performance.

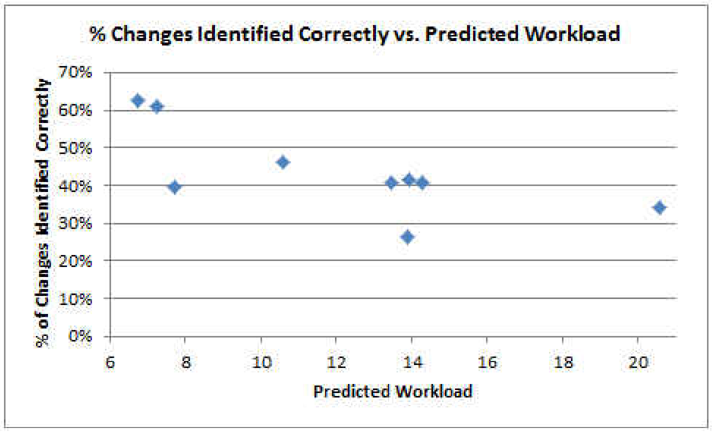

Figure 56: Predicted Workload vs. Percent of Changes Identified Correctly

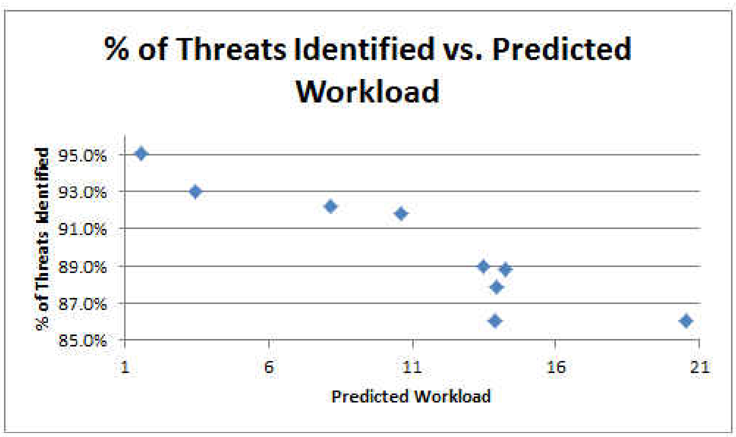

Figure 57: Predicted Workload vs. Percent of Threats Identified

Table 30 summarizes the four figures above in tabular form. From this table, the best performance occurs during Segments A and B (i.e., low and medium workload, respectively) of the single-task scenarios (i.e., Scenario 1 and Scenario 3). For the Change Detection performance measure, Segments A and B of Scenario 1 have similar performance, with a large difference between these and other scenario-segment combinations. This corresponds to workload values of 6.74 and 7.24, respectively, with workload declining steeply for values over 7.7 (Scenario 1, Segment C). For the Threat Detection performance, Segments A, B, and C of Scenario 3 and Segment A of Scenario 4 have the best performance, corresponding to workload values of 1.56, 3.46, 8.15, and 10.59, respectively. Thus, solely Threat Detection tasks, workload values of 10.59 or lower produce high levels of performance. However, the dual-task scenarios also include Change Detection, which experiences performance degradation by workload values of 7.7. For these scenarios, there are no specific performance goals; the goal is simply to achieve the best possible performance. Since these particular tasks occur in the decreasing portion of the workload-performance curve of Cassenti et al. (2011), lower workload values produce increased performance. Therefore, this study sets the target workload range for the dual-task scenarios to be between 1.5 and 7.5. However, if there was a specific performance goal, this could either limit or expand the target range. For example, if there was a Threat Detection performance objective of 93% or better, then the target workload range would be between 1.56 and 3.46.

Table 30: Performance and Workload Scores

Scenario, Segment	% Changes Identified Correctly	% Threats Identified	ISA Mean Value	Predicted Workload
S3 A		95%	1.63	1.56
S3 B		93%	2.08	3.46
S1 A	63%		2.33	6.74
S1 B	61%		2.59	7.24
S3 C		92%	2.69	8.15
S4 A	46%	92%	3.08	10.59
S1 C	40%		3.26	7.70

Scenario, Segment	% Changes Identified Correctly	% Threats Identified	ISA Mean Value	Predicted Workload
S2 A	41%	89%	3.27	13.47
S2 B	41%	89%	3.40	14.27
S4 B	41%	88%	3.43	13.95
S2 C	26%	86%	3.84	13.88
S4 C	34%	86%	3.94	20.55

6.3. Summary

The analysis in this chapter addresses Sub-Question 3, "How can simulation modeling be used to determine the target level or range of cognitive workload scores for adaptive automation?" The target workload range identified for this study (i.e., workload values between 1.5 and 7.5) is specific to the particular tasks performed in these scenarios. However, the methodology is widely applicable. When establishing a target workload range, it is important to characterize the workload-performance relationship in order to determine whether the particular task (or set of tasks) in question falls along an increasing, decreasing, or flat portion of the workload-performance continuum. It is also possible for a task (or set of tasks) to span multiple segments of this continuum. Once this relationship is known, the objective to minimize workload, maximize workload, or maintain workload within a target range should be easier to identify. Another important consideration when establishing the target range is to consider whether there are any objective performance goals (e.g., at least 85% accuracy). These performance goals also influence the identification of the target workload range.

CHAPTER 7
ADAPTIVE AUTOMATION INVOKING THRESHOLDS

7.1. Introduction

This chapter describes enhancements to the baseline DES model from Section 6.1 in order to include adaptive automation to the baseline scenarios. This chapter, then, explores the impact of various invoking thresholds on workload and situational awareness and provides a recommended invoking threshold for this adaptive system.

7.2. Incorporating Adaptive Automation

To incorporate adaptive automation into the baseline DES model, it is necessary to determine which sub-task(s) are to be automated and what mechanism is to be used to trigger the automation. Based on the nature of the tasks and the relative workload across the four scenarios considered in this research investigation, adaptive automation is added to only the dual-task scenarios, i.e., Scenario 2 and Scenario 4.

7.2.1. Determining the Sub-tasks to Automate

Based on the task analysis and observations of participants performing the tasks, and exit surveys completed by the PRIME 2 Study participants, the Change Detection task appears more difficult to perform than the Threat Detection task. During the Change Detection task, a change event occurs at a particular point in time. If the operator does not see the change, or is unsure of which change event occurs, there is no means for the operator to replay the video feed to verify what has occurred. Furthermore, the operator must identify the change type by selecting the appropriate button before the next change occurs. However, the operator does not know when the next change will occur. So, the operator is faced with an unknown and limited amount of time to perform the task. The range of time in which the participant must respond in the PRIME 2 Study is between 1 and 10 seconds.

While the Change Detection task is rigid, the Threat Detection task provides the operator with considerable flexibility. In the Threat Detection task, the threat and non-threat images are stationary. However, the images first appear as small images on the screen displaying the video feed. Then, as the UGV travels forward through the urban environment, the images grow and appear larger as the UGV gets close to the images. Eventually, the images disappear from the video feed as the UGV passes them. In the Threat Detection task, most images are on the screen for over 30 seconds, and the appearance of an additional threat does not prevent the operator from clicking on any previously identified threats remaining on the screen displaying the video feed. As a result, the operator can respond to the visible threats in any order, and can look at the threat as often or for as long as the threat is on the screen. This allows the operator some flexibility in managing his or her time to respond to threats, as well as in assessing whether or not an actor is a threat.

Based on the nature of the Change Detection task and the Threat Detection task, the preferred task to automate in this research investigation is the Change Detection task. This task is also more reasonable to automate since determining the appearance, disappearance, or movement of an icon is a univariate decision, whereas determining whether an actor image is a threat is a multivariate decision, requiring more study and consideration of several characterization variables (e.g. distinguishing between friendly and enemy soldiers based on uniform types and weapon types). Automating the Change Detection task in this study aligns with the work of Arciszewski et al. (2009), which suggests using automation for less critical and repetitive tasks, while reserving higher priority and more engaging tasks for the human. Thus, the univariate Change Detection task is a logical choice for automation, whereas the multivariate Threat Detection task makes it more suitable to be performed by the human operator.

7.2.2. Invoking Automation

Automation is invoked when an operator's workload reaches or exceeds a pre-specified threshold. Establishing the mechanism for triggering the automation involves not only establishing a threshold, but also determining the waypoints and/or at what frequency workload measurements are taken for comparison with the threshold. The Change Detection task sequence involves monitoring the map, seeing a change, identifying a change, and selecting the appropriate button. The response chain of seeing, identifying, and selecting takes less than three (3) seconds, on average, and it is anticipated that it is not logical to invoke automation in the middle of this sequence. Thus, the adaptive system is designed to invoke only during the monitoring map phase. The adaptive system compares the operator's total workload level (including interference) against an established invoking threshold, based on an established frequency.

Figure 58 is the updated Change Detection task network diagram. In this updated model, two additional nodes are added in order to capture the adaptive automation logic. In this updated task network, the system compares the operator's current workload to the value of the invoking threshold. If the operator's workload is greater than or equal to the invoking threshold, then the automation is turned on. When this occurs, the operator does not experience any workload from the Change Detection task. Rather, the operator experiences workload only from the Threat Detection task. The system continues to check the operator's workload while the automation is active, based on a fixed duration. If the operator's workload falls below the revoking threshold, then the automation is turned off, and the operator must perform both the Change Detection and Threat Detection tasks.

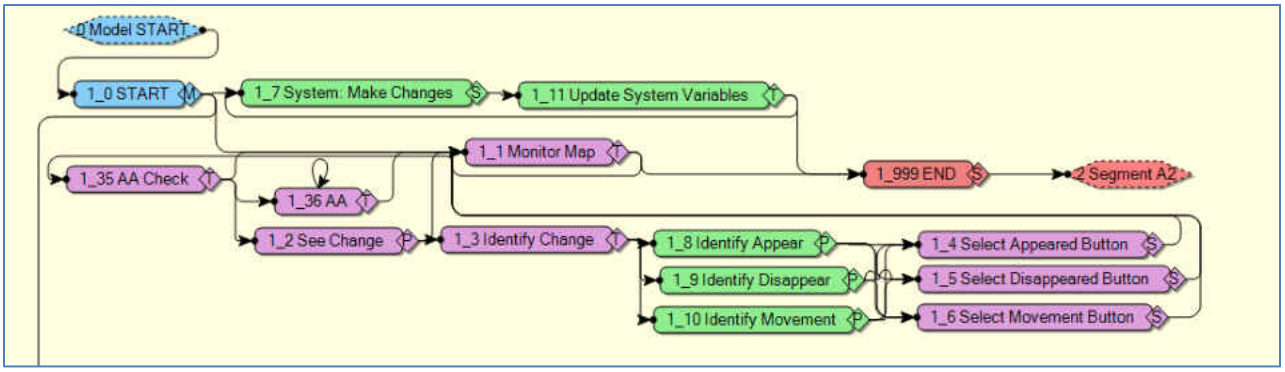

Figure 58: Change Detection Task Network Diagram with Adaptive Automation

7.3. Objective of the Invoking Threshold Experiment

To determine an appropriate automation invoking threshold, this study conducts an Invoking Threshold Experiment that evaluates candidate thresholds. While increasing the invoking threshold is expected to reduce the quantity of automation, the specific cognitive workload and situational awareness effects are unknown. This experiment takes an intra-system approach, specifically evaluating the situational awareness-workload tradeoff that is made when selecting an invoking threshold for an adaptive system.

This study considers three hypotheses. The **first hypothesis** is that the tradeoff between situational awareness and workload, for the system being evaluated, is positively correlated and non-linear. Since the relationship is expected to be positively correlated, when automation is low (i.e., with a high invoking threshold), both workload and situational awareness will be high. Conversely, when automation is high (i.e., with a low invoking threshold), workload and situational awareness will be low. This relationship is hypothesized to be non-linear, rather than directly proportional. Thus, increasing the automation invoking threshold will not tradeoff workload and situational awareness at a constant rate.

The **second hypothesis** is that the tradeoff between cognitive workload and situational awareness depends on the level of task load (low, medium, or high). That is, the tradeoff is more sensitive (steeper curve) at certain levels of task load than at other levels. Thus, the value gained from situational awareness by increasing workload depends on the initial task load level.

The **third hypothesis** is that the tradeoff between workload and situational awareness depends on the congruity between tasks being adaptively automated and those with varying task load. Adaptive automation is anticipated to have a greater impact on the tradeoff between cognitive workload and situational awareness for those tasks that vary task load in the tasks being automated (i.e., when Change Detection is varied, Scenario 2).

7.4. Invoking Threshold Experimental Procedure

This section describes the experimental design, model variables, and simulation outputs for the Invoking Threshold Experiment.

7.4.1. Experimental Design

The Invoking Threshold Experiment uses a 3x2x4 design, with three task load levels, two task congruity aspects, and four invoking threshold levels. The three levels of task load are low, medium, and high, as described in Section 4.2.4. The two task congruity aspects capture whether the task being automated is congruent with the task being varied in task load. Scenario 2 is congruent because the task being adaptively automated (Change Detection) and the task that varies in task load (Change Detection) are the same. Scenario 4 is incongruent task being adaptively automated (Change Detection) and the task that varies in task load (Threat Detection) are not the same. The initial four invoking thresholds are cognitive workload values of 10, 20, 30, and 40. This design serves as a screening experiment because these thresholds cover a large proportion of the expected design space in terms of workload values. Based on the results from this initial screening experiment, additional invoking threshold levels are chosen based on the initial threshold levels that provide additional information regarding workload-situational awareness tradeoffs. Each of the 24 variants (3x2x4) is simulated and run for 10 independent replications, with each replication using a unique random number seed. Based on the low variability within each variant compared to the variability between the variant, it is determined that running 10 replications is sufficient.

7.4.2. Adaptive Automation Model Variables

The adaptive automation DES model contains four variables of interest for the simulation experiments: the Invoking Threshold, the Revoking Threshold, the Adaptive Automation Check Duration, and the Adaptive Automation Duration.

7.4.2.1. Invoking Threshold

The Invoking Threshold is the level at which the automation is activated (i.e., turns on) if the operator's total cognitive workload, including interference, reaches or exceeds its value. Based on the workload management strategy established in Section 6.1, the operator's workload value can vary from 0 to 80, with most values occurring below 50. This experiment begins with Invoking Thresholds of 10, 20, 30, and 40.

7.4.2.2. Revoking Threshold

The Revoking Threshold is the level at which the automation is deactivated (i.e., turns off) if the operator's total cognitive workload, including interference, falls below its value. For the purposes of this experiment, the Revoking Threshold is equal to the Invoking Threshold.

7.4.2.3. Adaptive Automation Check Duration

The Adaptive Automation Check Duration is the frequency at which the adaptive system compares the Invoking Threshold to the operator's total workload. For the purposes of this experiment, the Adaptive Automation Check Duration is equal to 2 seconds. This value is based upon the distribution of time to perform the Change Detection task.

7.4.2.4. Adaptive Automation Duration

The Adaptive Automation Duration is the amount of time that the adaptive system is activated, before checking to see whether or not the automation should be revoked. For the purposes of this experiment, the Adaptive Automation Duration is equal to 2 seconds. This value is based upon the average amount of time to perform the Change Detection task.

7.4.3. Simulation Performance Measures

The simulation experiments produce three performance measures of interest: predicted cognitive workload values and two measures of situational awareness. The predicted workload values are the total workload values including interference, as described in Section 6.1. Two outputs are identified as surrogates for situational awareness. Endsley (1995) describes situational awareness along three increasing levels: Perception (Level 1), Comprehension (Level 2), and Projection (Level 3). The first situational awareness output measure is the percentage of changes detected, i.e., the number of changes identified by the operator divided by the total number of changes, excluding changes identified by the adaptive system. Thus, this measure is a Level 2 situational awareness measure since it requires the operator to see and respond to a change.

The second output measure is the percentage of time the operator spends monitoring the map, i.e., the number of time units spent performing the monitoring map task, divided by the total amount of time for that variant. Monitoring the map excludes any time spent in the adaptive mode, as well as time spent identifying changes. This output is a Level 1 situational awareness measure since the operator is viewing the monitor and not necessarily comprehending the content. The former output measure represents limited focus on a specific task, while the latter measure represents a system-level perspective.

7.5. Invoking Threshold Experiment Results

This section describes the initial screen experiment, addition of invoking thresholds, and experimental results for the Invoking Threshold Experiment.

7.5.1. Screening Experiment & Additional Invoking Thresholds

Figure 59 and Figure 60 show predicted cognitive workload versus the situational awareness output measures for Scenario 2 and Scenario 4, respectively. The graphs display the three task load conditions separately, with each invoking threshold as a single point based on the average value across the 10 independent replications. From these initial experiments, in all cases, the workload-situational awareness graphs show that the points cluster into two groups, with thresholds 10 and 20 very close together, and thresholds 30 and 40 very close together. These results suggest that the tradeoff curve exhibits asymptotic behavior, with a possibility of a non-linear tradeoff between invoking thresholds 20 and 30. The clustering behavior of these points supports the first hypothesis, in that, varying the invoking threshold does not produce a proportional change in situational awareness and cognitive workload. Based on these results, the initial experimental design is expanded to include additional automation invoking levels. Additional levels are chosen based on the midpoint between the two invoking thresholds that produce the largest gap. Thus, three additional invoking thresholds of 23, 24 and 25 are included.

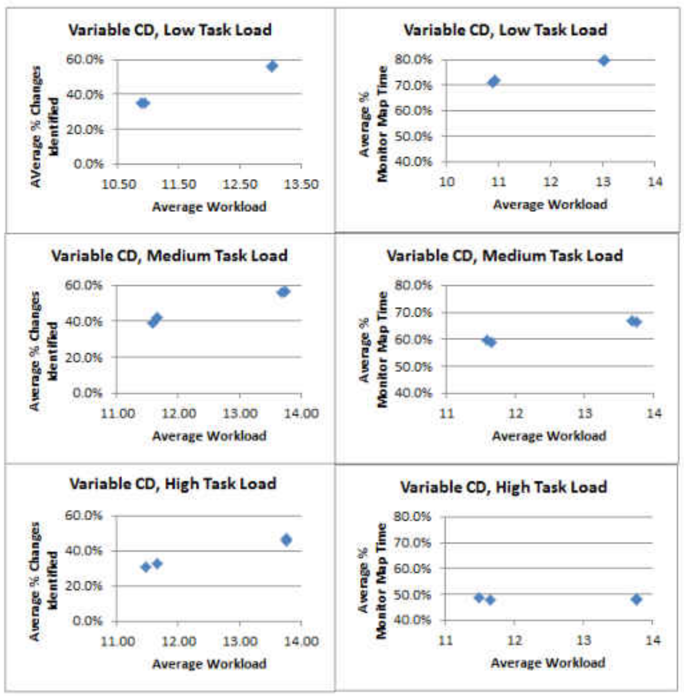

Figure 59: Screening Experiment, Predicted Cognitive Workload vs. Situational Awareness, Scenario 2

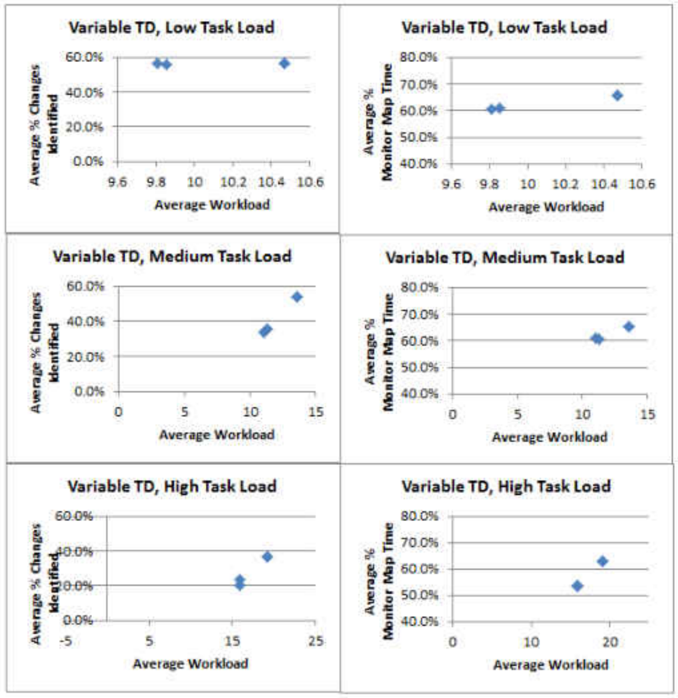

Figure 60: Screening Experiment, Predicted Cognitive Workload vs. Situational Awareness, Scenario 4

7.5.2. Scenario 2 Results

Figure 61 shows predicted cognitive workload versus the situational awareness performance measures for Scenario 2, revealing a gap between invoking thresholds 24 and 25, rather than a continuous concave or convex curve. An Analysis of Variance (ANOVA) at a 5% level of significance reveals that there is a statistically-significant difference between invoking threshold levels 24 and 25 for the predicted cognitive workload and two situational awareness measures, with the exception of the Average % Monitor Map Time for the high task load variant. In almost all cases, there is neither a statistical difference between invoking thresholds 10 to 24 nor between invoking thresholds 25 to 40 (See Appendix D for the detailed ANOVA results).

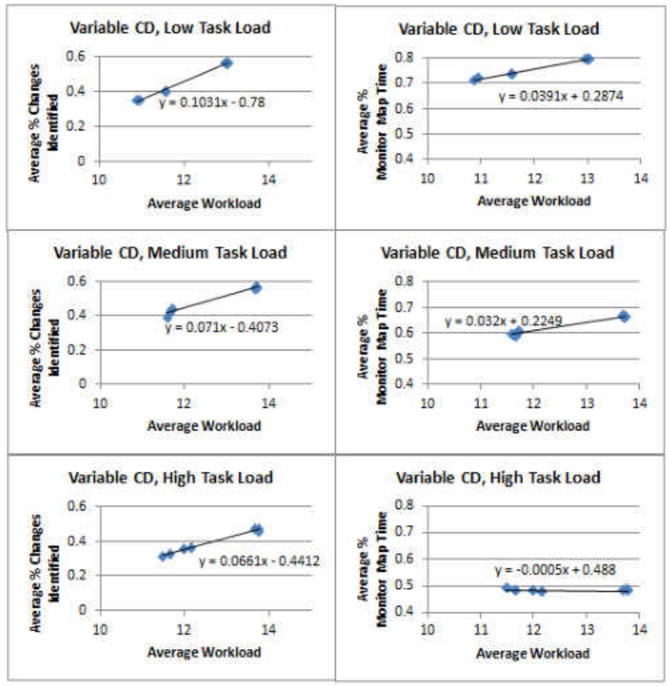

Figure 61: Final Experiment, Cognitive Workload vs. Situational Awareness, Scenario 2

The relationship between cognitive workload and situational awareness appears to be relatively linear. While the tradeoff between workload and identifying changes has the expected positive relationship, the relationship between workload and monitoring the map appears to be relatively flat, with increases in workload having little to no effect on situational awareness. This is especially the case for the high task load variant. This linear relationship is confirmed by the correlations summarized in Table 31. Table 31 shows that workload and identifying changes are positively correlated across all three task load variants. The correlations are each based on 70 points (i.e., ten independent replications for each of the seven invoking threshold levels).

Table 31: Correlations between Cognitive Workload and Situational Awareness, Scenario 2

Task Load	Workload-Changes Identified	Workload-Monitor Map
Low Task Load	0.817	0.654
Medium Task Load	0.801	0.726
High Task Load	0.917	-0.041

7.5.3. Scenario 4 Results

Figure 62 shows the graphs for the predicted cognitive workload versus the situational awareness performance measures for Scenario 4. These graphs also reveal a significant gap between invoking thresholds 24 and 25, rather than a continuous concave or convex curve. ANOVA results assuming a 5% level of significance reveal that there is a statistically-significant difference between thresholds 24 and 25 for the predicted cognitive workload and two situational awareness measures, with the exception of the Average % Changes Identified for the low task load variant and the Average % Monitor Map Time for the medium task load variant. In almost all cases, there is neither a statistical difference between thresholds 10 to 24 nor between thresholds 25 to 40 (See Appendix E for the detailed ANOVA results).

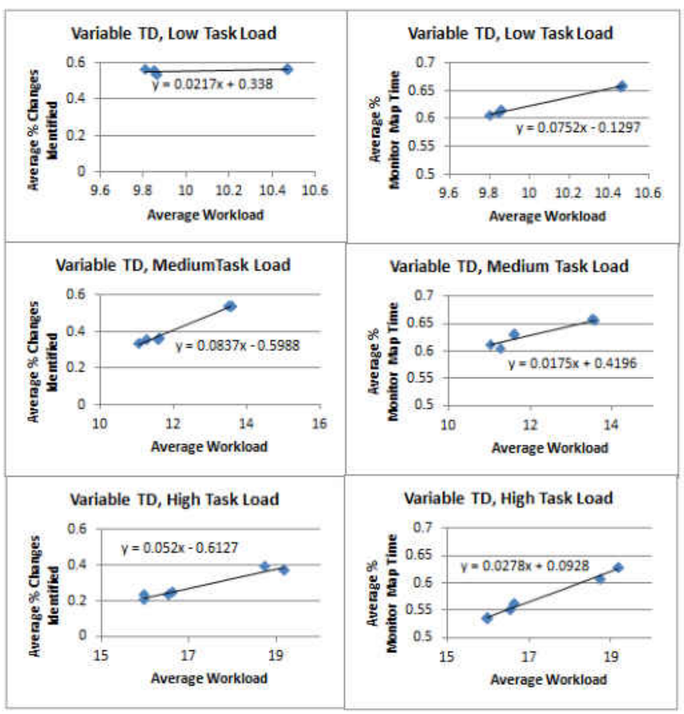

Figure 62: Final Experiment, Predicted Cognitive Workload vs. Situational Awareness, Scenario 4

The low task load variants produce workload scores such that automation is rarely invoked. Thus, there is essentially no change in predicted cognitive workload or situational awareness across the invoking thresholds. Note the scale of the *x*-axis for the low task load charts. For the medium and high task load variants, there appears to be a positive, linear (although discontinuous) relationship between workload and both measures of situational awareness. Table 32 shows that there is no relationship between workload and identifying changes for the low task load variant. However, there is a strong correlation for the medium and high task load variants. Predicted cognitive workload and the monitoring map situational awareness measure have a stronger correlation for the low task load variant, compared to predicted cognitive workload and the identifying changes situational awareness measure. However, the correlation between predicted cognitive workload and the monitoring map situational awareness measure is less strongly correlated for the medium task load variant.

Table 32: Correlations between Predicted Cognitive Workload and Situational Awareness, Scenario 4

Task Load	Workload-Changes Identified	Workload-Monitor Map
Low Task Load	0.295	0.632
Medium Task Load	0.915	0.577
High Task Load	0.858	0.821

7.6. Invoking Threshold Experiment Discussion

This section provides a discussion of the Invoking Threshold Experiment results for each of the three hypotheses.

7.6.1. Hypothesis 1: The tradeoff is positive and non-linear

This study finds that the tradeoff between predicted cognitive workload and situational awareness is positively correlated, for this system. However, the study does not find the tradeoff to be non-linear. Instead, the data reveal a discontinuity in the trend behavior. On each side of the gap caused by the discontinuity, invoking threshold levels produce similar values despite significant differences in those threshold levels. Thus, the difference in workload/situational awareness between a threshold of 10 and 20 is insignificant, as is the difference between 25 and 40. However, the difference between a threshold of 24 and 25 is rather large. This reveals that, while the relationship between workload and situational awareness is relatively linear, the relationship between the invoking threshold and its effects on workload/situational awareness is not. The positive trend indicates that, under most conditions, an increase in cognitive workload also results in an increase in situational awareness. This means operators are primarily working somewhere between the Ideal and Challenged regions of the Endsley (1993) workload-situational awareness continuum. However, the tradeoff depends on the task load.

7.6.2. Hypothesis 2: The tradeoff depends on the task load

For Scenario 2, the slope of the line decreases with increasing task load (see Figure 61). Therefore, at lower levels of task load, additional cognitive workload produces greater increases in situational awareness than it does at higher levels of task load. However, the slope for the monitoring map situational awareness performance measure may indicate that the operator is functioning at a high level of situational awareness, with decreases in automation primarily resulting in increased workload, thus moving the operator from the Ideal to the Challenged region, and for the high task load condition, possibly into the Overload region of the Endsley (1993) continuum. For Scenario 4, there does not appear to be a consistent relationship between the level of the task load and the slope of the workload-situational awareness curve since increasing or decreasing task load does not predictably increase or decrease the slope of the curve. For the low task load, all threshold levels produce similar predicted cognitive workload levels, which is the lowest of all variants, thus indicating that this task combination is between the Vigilance and Ideal regions. The medium and high task load combinations appear to fall between the Ideal and Challenged regions.

7.6.3. Hypothesis 3: The tradeoff depends on the congruency of the automation with the task load

When comparing the two scenarios, across all task load levels, the workload-situational awareness tradeoff curves are more consistent for Scenario 2 for both situational awareness performance measures. In this scenario, the task being adaptively automated and the tasks that are varying in task load are the same (i.e., congruent). The predicted cognitive workload and identifying changes situational awareness measure relationship is consistently increasing and highly positively correlated across all task load variants. The predicted cognitive workload and monitoring map situational awareness measure relationship are relatively flat (i.e., slope < 0.05) for all task load variants, with the lines become flatter as task load increases. For Scenario 4, the tasks being adaptively automated and the tasks that are varying in task load are not the same (i.e., incongruent). For the identifying changes situational awareness performance measure, the slope is relatively flat for low task load, but not the other task loads, while the monitor map situational awareness performance measure is relatively flat for the medium and high task loads, but not the low task load.

7.6.4. Invoking Threshold Experiment Conclusions

Within an adaptive system, predicted cognitive workload and situational awareness are dependent upon the invoking threshold and the specific relationship between situational awareness and workload for that system. In order to choose an appropriate invoking threshold, system designers must be aware of where the operator-task combination falls on the cognitive workload-situational awareness continuum. Depending on the level of task load and the congruency of the automation with the task load, this study discovers instances where increasing the invoking threshold leads to increases in both workload and situational awareness, whereas, in other cases, it leads to increases in cognitive workload with no corresponding increases in situational awareness. Familiarity with which regions the invoking thresholds span allows system designers to effectively balance cognitive workload and situational awareness performance based on the invoking threshold of the adaptive automation.

7.7. Recommended Invoking Threshold

As discussed in this chapter, in many cases, the change in workload due to a change in invoking threshold is negligible. This is especially the case for invoking threshold levels 10 to 24 and for invoking threshold levels 25 to 40. However, there is a measureable decrease in workload between thresholds 24 and 25. Table 33 displays the numerical workload values for the entire set of invoking threshold experiments. This table shows that, for higher thresholds (i.e., 25 through 40), there is only a slight improvement in predicted cognitive workload between adaptive automation and the baseline scenario without adaptive automation. While threshold levels 10 through 24 aid in the goal of reducing cognitive workload, none of these experiments achieve the target workload range between 1.5 and 7.5. However, this is to be expected, since the revoking strategies are not yet evaluated.

The analysis in this chapter also reveals that operators who would complete the Change Detection and Threat Detection tasks are operating between the Ideal, Challenged, and Overload states, depending on the scenario-segment combination. Therefore, it is important to choose an invoking threshold that results in a lower cognitive workload. Since the workload gains are best achieved in thresholds 10-24, and there is minimal difference between these, this study selects an invoking threshold of 24 for the revoking strategies experiment, in order to give ample opportunity to evaluate the revoking strategies.

Table 33: Workload Values by Threshold

Condition	Baseline with Interference	Threshold						
		10	20	23	24	25	30	40
S2_A	13.47	10.88	10.94	11.57	11.58	12.97	13.02	13.04
S2_B	14.27	11.59	11.66	11.70	11.74	13.72	13.74	13.69
S2_C	13.88	11.48	11.66	12.16	12.00	13.68	13.78	13.76
S4_A	10.59	9.81	9.85	9.86	9.86	10.47	10.47	10.47
S4_B	13.95	11.05	11.29	11.63	11.59	13.52	13.58	13.58
S4_C	20.55	15.96	15.97	16.53	16.63	18.75	19.16	19.16

7.8. Invoking Threshold Conclusions

The analysis in this chapter addresses Sub-Question 4, "How can simulation modeling be used to determine a preferred invoking threshold for adaptive automation?" The invoking threshold identified for the particular scenarios in this study is a workload value of 24; however, this invoking value is appropriate and specific to this particular study. When selecting an invoking strategy, it is important to consider the impact of the threshold levels under consideration on both cognitive workload and situational awareness. This consideration requires the system designer to know whether the operator is performing tasks in the Vigilance, Ideal, Challenged, or Overload state of Endsley's (1993) workload-situational awareness continuum.

This knowledge enables the system designer to determine whether workload and/or situational awareness should be increased or decreased. In addition, this knowledge can also help to determine if it is more valuable to trade off workload for situational awareness, or vice versa, if necessary.

When operating in the Vigilance and Ideal states, workload is low and, thus, the need for adaptive automation is low. Especially in the case of the Vigilance state, where situational awareness is also low, it is important to have less automation and to increase operator involvement in performing the tasks. Thus, a higher invoking threshold would be expected for these states. In the Challenged and Overload states, the workload is high, and thus these would indicate the need for lower invoking thresholds. This is especially important for the Overload states, in which workload is high and situational awareness maybe diminishing or lost.

CHAPTER 8
USING SIMULATION TO EVALUATE ADAPTIVE AUTOMATION REVOKING STRATEGIES

8.1. Introduction

This chapter describes the simulation-based revoking strategies experiments that are used to evaluate potential revoking strategies. Specifically, these experiments use DES modeling to determine the effects of varying: (1) the automation duration and (2) the revoking threshold on workload and situational awareness.

This study addresses two hypotheses. The **first hypothesis** is that the tradeoff between cognitive workload and situational awareness depends on duration of the automation. As the duration increases, both workload and situational awareness are expected to decrease. However, it is not known whether this tradeoff curve is linear, concave, convex, or none of these.

The **second hypothesis** is that the tradeoff between workload and situational awareness depends on the level of the revoking threshold. Since lower revoking thresholds are expected to result in increased automation, as the threshold decreases, both workload and situational awareness are expected to decrease. However, it is not known whether this tradeoff curve will be linear, concave, convex, or none of these.

8.2. Revoking Strategies Experimental Procedure

Two computational experiments are run to test the revoking strategy hypotheses. In the first experiment, the revoking threshold level when the automation is deactivated is fixed while the duration the automation is active is varied. In the second experiment, the duration the automation is active is fixed, while the revoking threshold level is varied.

8.2.1. Experimental Design: Duration Experiment

The Duration Experiment uses a 3x2x8 design, with three task load levels, two dual-task scenarios, and six adaptive automation duration times. The three levels of task load are low, medium, and high. The two dual-task scenarios are Scenario 2 and Scenario 4. The six durations are 1, 2, 5, 10, 15, 20, 25, and 30 seconds. Each of the 48 variants (3x2x8) is simulated and run using 10 independent replications, with each of the 10 replications using a unique random number seed. Due to the low variability within each simulated variant compared to the variability between variants, it is determined that additional replications are unnecessary.

8.2.2. Adaptive Automation Model Variables: Duration Experiment

The adaptive automation DES model contains four variables of interest for the simulation experiments: the Invoking Threshold, the Revoking Threshold, the Adaptive Automation Check Duration, and the Adaptive Automation Duration.

8.2.2.1. Invoking Threshold

The Invoking Threshold is the level at which the adaptive automation will turn-on if the operator's total workload, including interference, exceeds its value. Based on the analysis and results from the Invoking Thresholds Experiments described and discussed in CHAPTER 7, the Revoking Threshold for the Duration Experiments is held constant at 24.

8.2.2.2. Revoking Threshold

The Revoking Threshold is the level at which the automation is deactivated (i.e., turns off) if the operator's total cognitive workload, including interference, falls below its value. For the purposes of this experiment, the Revoking Threshold is equal to the Invoking Threshold. For the purposes of this experiment, the Revoking Threshold is equal to the Invoking Threshold at 24.

8.2.2.3. Adaptive Automation Check Duration

The Adaptive Automation Check Duration is the frequency at which the adaptive system compares the Invoking Threshold to the operator's total workload when the operator has manual control. For the purposes of this experiment, the Adaptive Automation Check Duration is set to 0.10 seconds. This value is selected as an appropriate duration in order to balance the frequency of monitoring the operator's total workload and the computational expense of the simulation.

8.2.2.4. Adaptive Automation Duration

The Adaptive Automation Duration is the amount of time that the adaptive system is active before checking to see if the automation should be revoked. This is the variable of interest for the Durations Experiments, and is set to values of 1, 2, 5, 10, 15, 20, 25, and 30 seconds.

8.2.3. Simulation Performance Measures: Duration Experiment

The simulation runs produce predicted cognitive workload and two measures of situational awareness, as described in Sections 6.1 and 7.4.3, respectively.

8.3. Revoking Strategies Experiment Results: Duration Experiment

Figure 63 provides the simulation outputs for Scenario 2 of the Duration Experiment. These graphs show the average predicted cognitive workload or situational awareness value by duration for each of the three task load levels. The relationship between duration and workload, as well as duration and situational awareness, is decreasing and non-linear. Many of these Sshaped curves reveal a diminishing return as automation is increased. The curves begin to level-off after 10 seconds, with significant differences in workload values between 1 and 10 seconds, and only slight differences between 10 and 30 seconds.

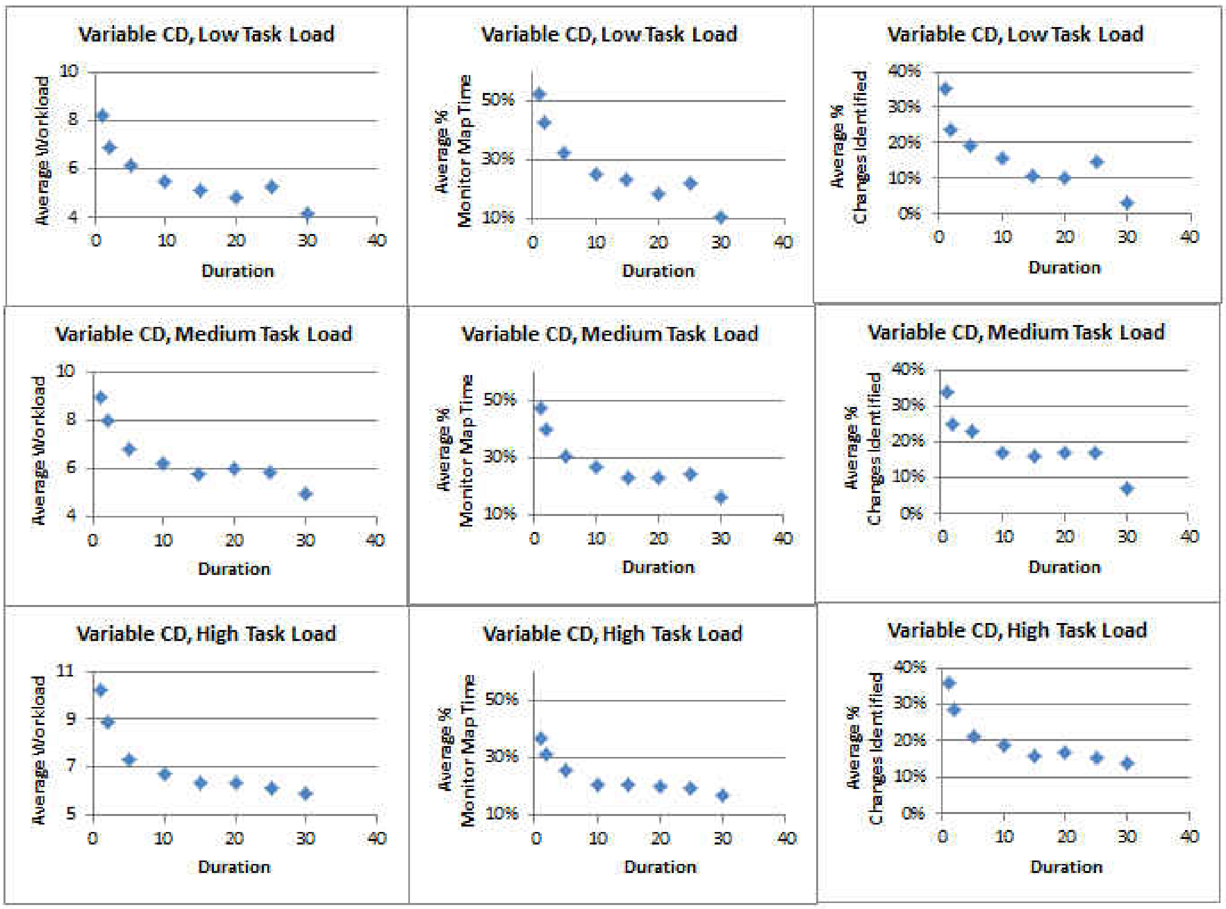

Figure 63: Duration versus Predicted Cognitive Workload and Situational Awareness, Scenario 2

Figure 64 displays the Analysis of Variance (ANOVA) results (assuming a 5% level of significance) that correspond to each of the graphs shown in Figure 63. While there is some overlap in the 95% confidence intervals at the higher durations, these ANOVAs support that the differences in predicted cognitive workload and situational awareness are statistically significant for duration values between 1 and 10 seconds.

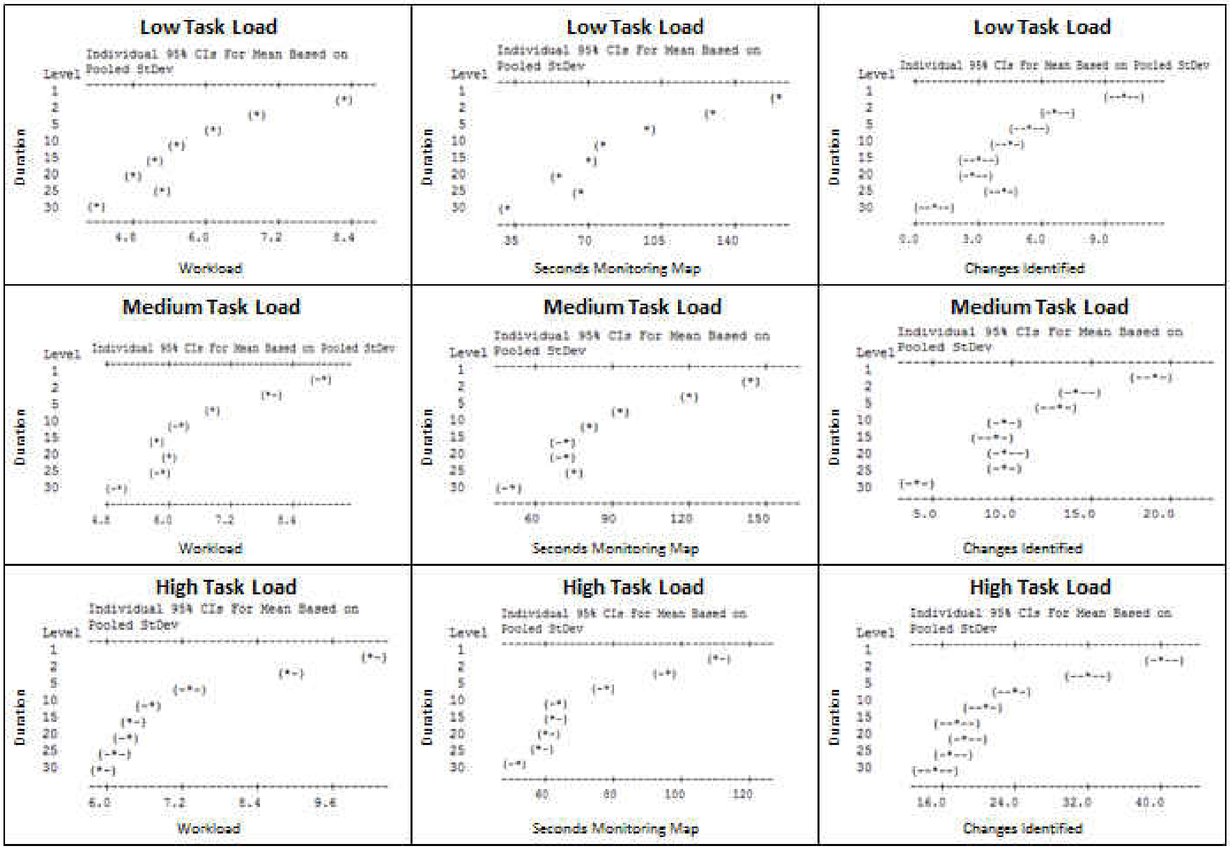

Figure 64: ANOVA Results: Duration versus Predicted Cognitive Workload and Situational Awareness, Scenario 2 (assuming a 5% level of significance)

Figure 65 provides the simulation performance measures for Scenario 4 of the Duration Experiment. The relationship between duration and predicted cognitive workload, as well as duration and situational awareness, is also decreasing and non-linear. As with Scenario 2, Scenario 4 also shows that, at about 10 seconds, workload gains begin to decrease as automation increases. Note that, while the absolute scales of the axes on the graphs vary, the size of the range is held constant. The difference in predicted cognitive workload between a duration of 1 second and a duration of 30 seconds decreases by approximately 6 workload points at the low task load, 4 workload points at medium task load, and 2.5 workload points at high task load. Thus, the predicted cognitive workload graphs reveal a reduced spread as the task load increases, indicating lower returns in the workload-duration tradeoff at higher task loads. A similar trend can be seen with the situational awareness curves.

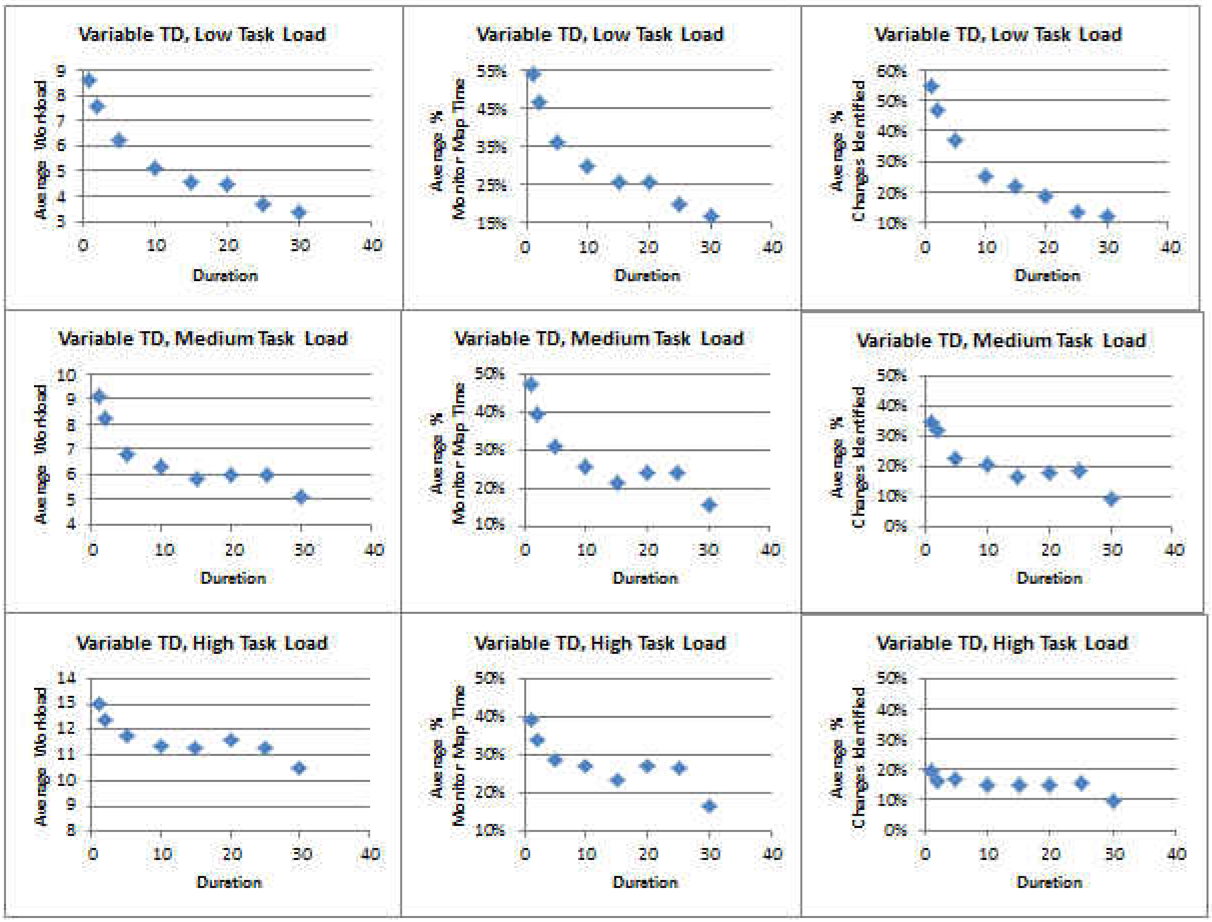

Figure 65: Duration versus Predicted Cognitive Workload and Situational Awareness, Scenario 4

Figure 66 shows the ANOVA results (assuming a 5% level of significance) that corresponds to each of the graphs shown in Figure 65. While there is some overlap in the 95% confidence intervals at the higher durations, these ANOVA results support that the differences in workload and situational awareness are statistically significant for durations between 1 and 10 seconds.

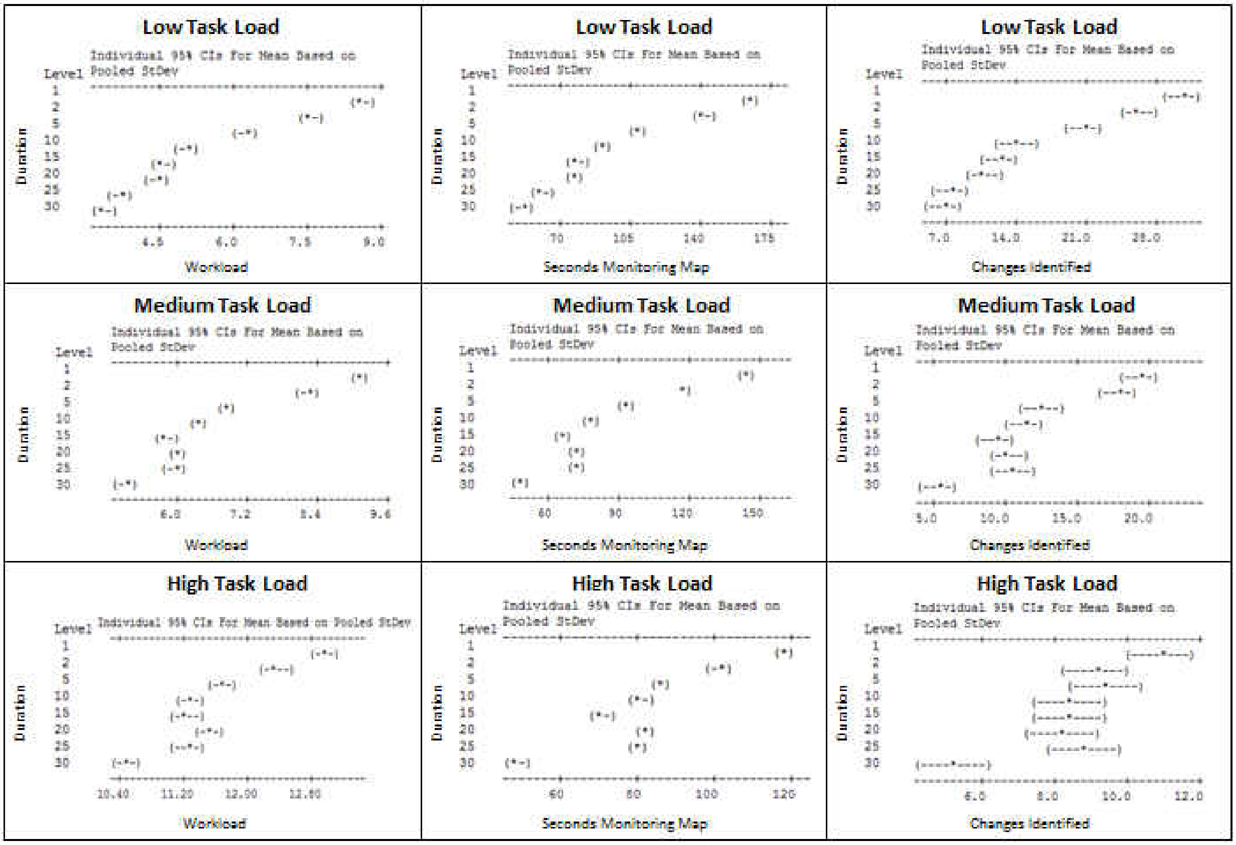

Figure 66: ANOVA Results: Duration versus Predicted Cognitive Workload and Situational Awareness, Scenario 4 (assuming a 5% level of significance)

8.4. Revoking Strategies Experiment Discussion: Duration Experiment

The Duration Experiment confirms the hypothesis that the tradeoff between cognitive workload and situational awareness depends on the duration of the automation. Also, as the duration increases, both workload and situational awareness decrease. The experiment reveals the tradeoff to be non-linear, with diminishing returns to increasing the duration of the automation at low durations, and a leveling off as duration increases. For Scenario 4, increasing the automation duration had less impact on workload and situational awareness at higher task loads. In Scenario 2, the task load varies for the Change Detection task (i.e., the task being automated), while the task load for the Threat Detection task is constant. The automation occurs for the task that is experiencing increased workload; hence, the graphs display declining workload levels with increasing durations across task loads. However, with Scenario 4, the task load varies for the Threat Detection task while the task load for the Change Detection task remains constant. The automation is thus limited in its ability to assist the operator with increasing task load, which explains the lower workload gains to increases in duration for higher task loads in this scenario.

8.5. Revoking Strategies Experiment: Threshold Experiment

This section describes the experimental design, important model variables, and simulation performance measures for the Revoking Threshold Experiment.

8.5.1. Experimental Design: Threshold Experiment

The Threshold Experiment uses a 3x2x6 design, with three task load levels, two dual-task scenarios, and six revoking thresholds. The three levels of task load are low, medium, and high. The two dual-task scenarios are Scenario 2 and Scenario 4. The six revoking thresholds are 5, 10, 15, 20, 22 and 24 workload units. Each of the 36 variants (3x2x6) is simulated and run using 10 independent replications, with each of the 10 replications using a unique random number seed. Based on the low variability within variant compared to the variability between variants, it is determined that additional replications are unnecessary.

8.5.2. Adaptive Automation Model Variables: Threshold Experiment

The adaptive automation DES model contains four variables of interest for the simulation experiments: the Invoking Threshold, the Revoking Threshold, the Adaptive Automation Check Duration, and the Adaptive Automation Duration.

8.5.2.1. Invoking Threshold

The Invoking Threshold is the level at which the adaptive automation is activated if the operator's total workload, including interference, reaches or exceeds its value. Based on the analysis from the Invoking Thresholds Experiments, the Invoking Threshold for the Duration Experiments is held constant at 24.

8.5.2.2. Revoking Threshold

The Revoking Threshold is the level at which the automation is deactivated (i.e., turns off) if the operator's total cognitive workload, including interference, falls below its value. This is the variable of interest for the Revoking Thresholds Experiments and is set to values of 5, 10, 15, 20, 22 and 24.

8.5.2.3. Adaptive Automation Check Duration

The Adaptive Automation Check Duration is the frequency at which the adaptive system compares the Invoking Threshold to the operator's total workload when the operator has manual control. For the purposes of this experiment, the Adaptive Automation Check Duration is set to 0.10 seconds. This value is selected as an appropriate duration in order to balance the frequency of monitoring the operator's total workload and the computational expense of the simulation.

8.5.2.4. Adaptive Automation Duration

The Adaptive Automation Duration is the amount of time that the adaptive system stays on, before checking to see if the automation should be revoked. The duration used for these experiments is 0.10 seconds. As with the Adaptive Automation Check Duration, this duration is chosen to simulate continuous monitoring.

8.5.3. Simulation Performance Measures: Threshold Experiment

The simulation runs provide predicted cognitive workload values and two measures of situational awareness, as described in Sections 6.1 and 7.4.3, respectively.

8.6. Revoking Strategies Experiment Results: Threshold Experiment

Figure 67 and Figure 68 show the simulation results for Scenarios 2 and 4 of the Threshold Experiment, respectively. These graphs show the average predicated cognitive workload or situational awareness value by revoking threshold for each of the three task load levels. As can be seen from the graphs, the relationships are relatively linear and level with the middle revoking thresholds (thresholds 10 to 22), but display non-linear behavior at the extremes (thresholds 5 and 24). In both scenarios, a revoking threshold of 5 produces a significant lowering in overall workload, but varies in the impact on situational awareness depending on the scenario, task level, and situational awareness measure. For Scenario 2, a revoking threshold of 5 produces measurable declines in average percentage of time monitoring map, but not in average percentage of changes identified. For Scenario 4, a revoking threshold of 5 produces measurable declines in average percentage of time monitoring map for medium and high task loads. These differences are found to be statistically significant (see Figure 69 and Figure 70). While the revoking threshold of 24 produces slight increases in workload, this difference is only statistically significant for high task loads in Scenario 4.

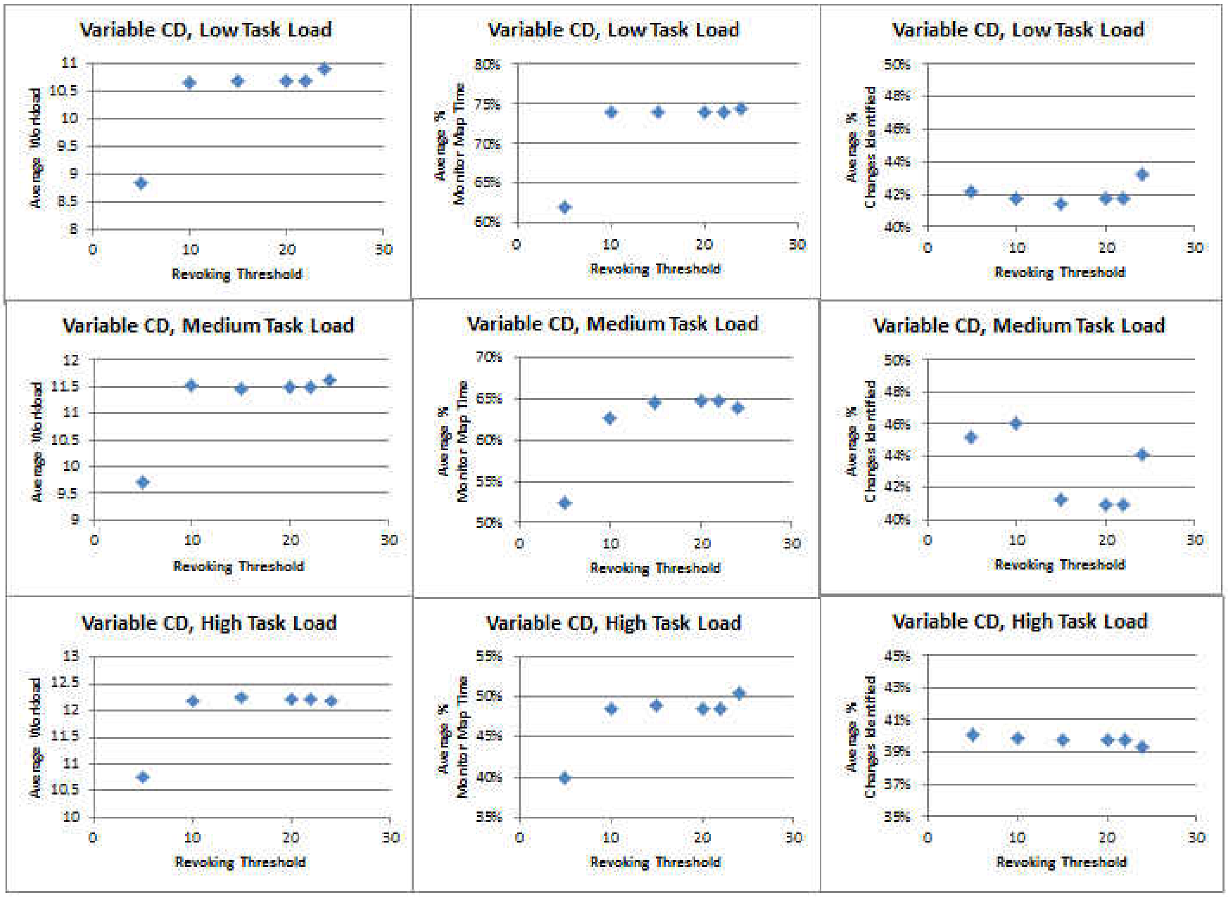

Figure 67: Revoking Threshold versus Predicted Cognitive Workload and Situational Awareness, Scenario 2

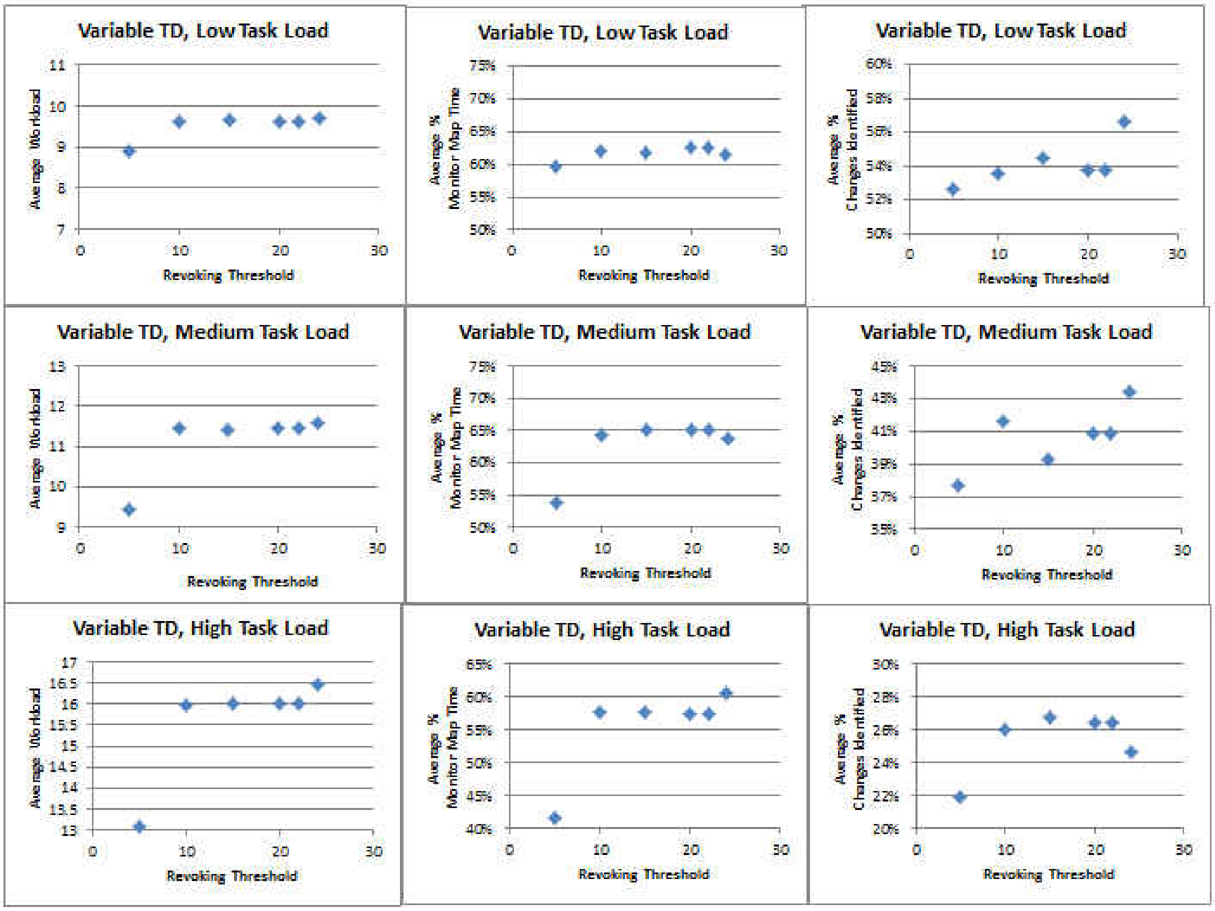

Figure 68: Revoking Threshold versus Predicted Cognitive Workload and Situational Awareness, Scenario 4

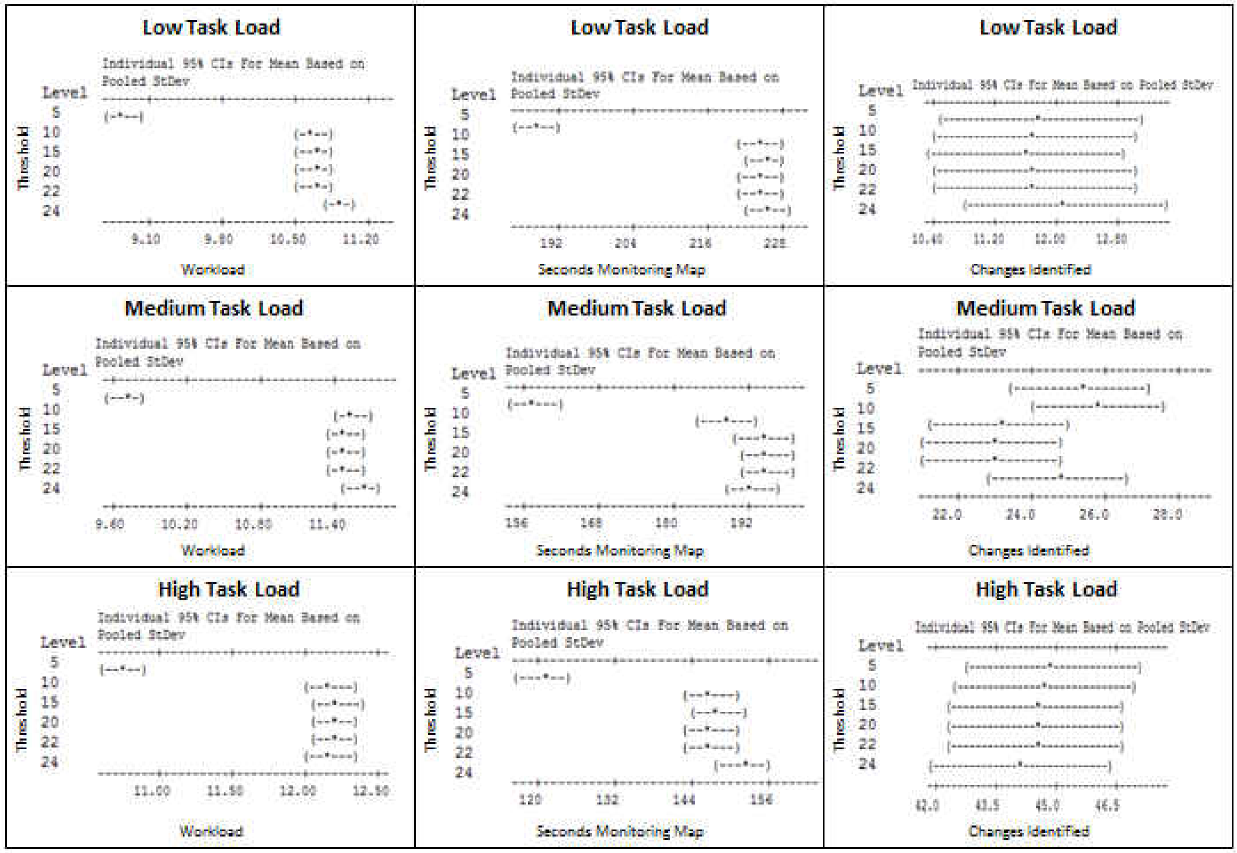

Figure 69: ANOVA Results: Revoking Threshold versus Predicted Cognitive Workload and Situational Awareness, Scenario 2 (assuming a 5% level of significance)

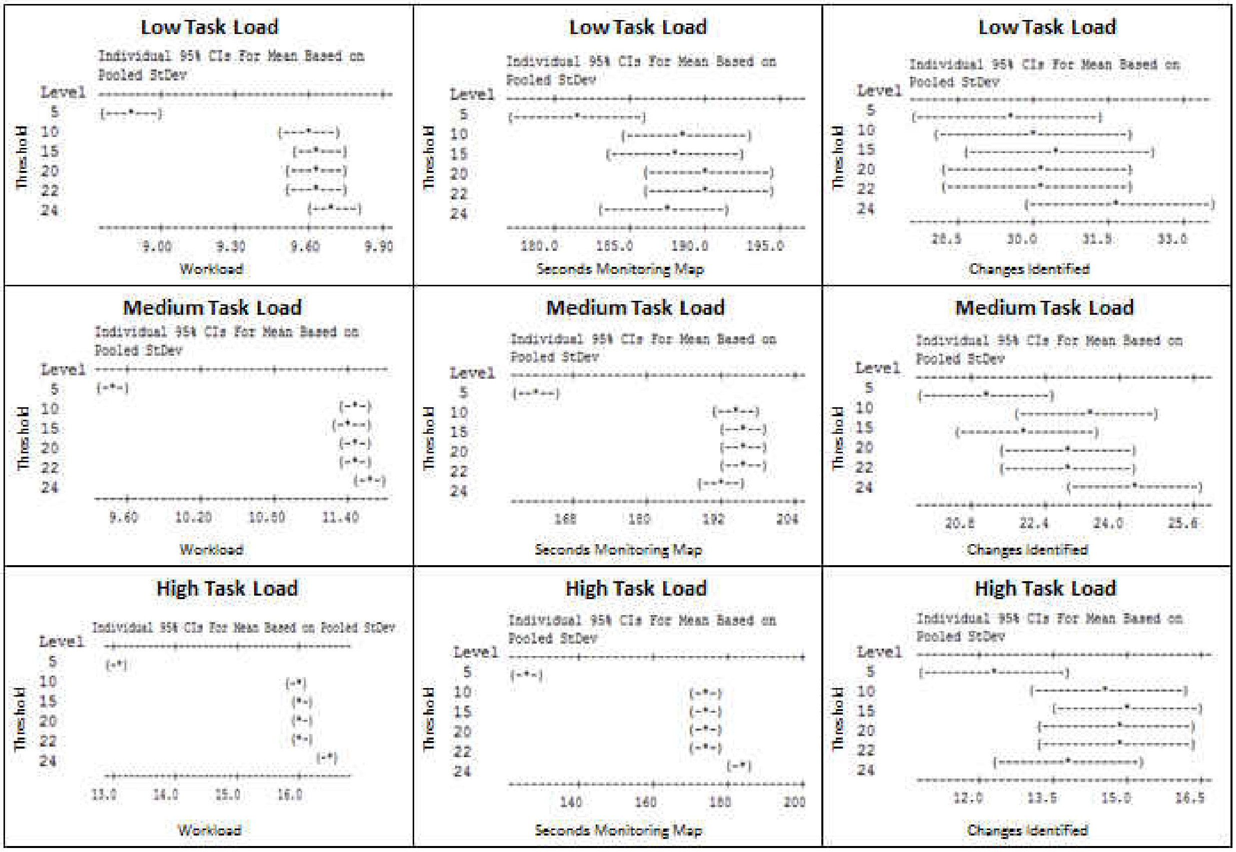

Figure 70: ANOVA Results: Revoking Threshold versus Predicted Cognitive Workload and Situational Awareness, Scenario 4 (assuming a 5% level of significance)

8.7. Revoking Strategies Experiment Discussion: Threshold Experiment

The Revoking Threshold Experiment results partially supports the hypothesis that the tradeoff between workload and situational awareness depends on the threshold for revoking automation, and that, as the revoking threshold decreases, both workload and situational awareness decrease. The results reveal that most revoking thresholds produce similar workload and situational awareness levels, thus these tasks are relatively insensitive to the revoking threshold level. However, statistically significant workload and situational awareness impacts did occur at extremely low revoking threshold levels.

8.8. Revoking Strategies Evaluation

The results from the Revoking Strategies Duration Experiment and the Revoking Strategies Threshold Experiment reveal that an adaptive system's duration and revoking threshold can impact the operator's cognitive workload and situational awareness levels. Thus, systems engineers can adjust the level of adaptive automation and influence the cognitive workload of the operator using these two design characteristics. For the scenarios in this investigation, there is a direct tradeoff between reducing workload and increasing situational awareness. The above analysis focuses on the relative impacts of the various potential levels of duration or revoking threshold. However, an aspect of greater interest is whether a fixed automation duration or a dynamic revoking threshold is the preferred revoking strategy.

Table 34 summarizes the workload scores by duration for each of the variants from the Revoking Strategies Duration Experiment. This table reveals that all of the durations evaluated produce large workload improvements over the baseline system without adaptive automation. From the ANOVA results in Section 8.3, all of these workload scores are significantly different from the baseline. Furthermore, in most variants, a duration of 5 seconds or longer produces workload scores in the target range of 1.5 to 7.5, with the exception being Scenario 4, Segment C.

Table 34: Workload Values by Duration

Scenario - Segment	Baseline WL Score	Durations 1	2	5	10	15	20	25	30
S2_A	13.47	8.23	6.89	6.14	5.51	5.13	4.83	5.25	4.16
S2_B	14.27	8.94	7.98	6.83	6.19	5.76	5.98	5.82	4.98
S2_C	13.88	10.23	8.90	7.32	6.70	6.37	6.34	6.12	5.93
S4_A	10.59	8.58	7.56	6.26	5.07	4.55	4.43	3.69	3.36
S4_B	13.95	9.13	8.24	6.81	6.35	5.80	6.00	5.95	5.11
S4_C	20.55	12.97	12.35	11.68	11.29	11.24	11.52	11.25	10.47

Table 35 summarizes the workload scores by revoking threshold for each variant from the Revoking Strategies Threshold Experiment. This table also shows large, statistically significant improvements over the baseline (refer to the ANOVA results in Section 8.6). However, none of these scores produce workload values within the desired target range. For all scenario-segment combinations, the duration of 1 second produces a lower workload value than a revoking threshold of 5, which has the lowest workload scores of the revoking threshold experiments. Therefore, using duration instead of revoking threshold as the revoking strategy allows for greater reductions in workload, making a fixed duration the preferred revoking strategy for this experiment.

Table 35: Workload Values by Revoking Threshold

Scenario - Segment	Baseline WL Score	Revoking Threshold					
		5	10	15	20	22	24
S2_A	13.47	8.83	10.66	10.68	10.70	10.70	10.92
S2_B	14.27	9.71	11.53	11.47	11.48	11.48	11.63
S2_C	13.88	10.74	12.17	12.22	12.20	12.20	12.17
S4_A	10.59	8.88	9.60	9.64	9.63	9.63	9.70
S4_B	13.95	9.46	11.47	11.42	11.45	11.45	11.58
S4_C	20.55	13.06	15.99	16.01	16.03	16.03	16.47

8.9. Summary

The analysis in this chapter addresses Sub-Question 5, "How can simulation modeling be used to determine a preferred revoking strategy for adaptive automation?" The revoking strategy analysis for the scenarios in this study reveals that a fixed minimum automation duration is preferred for these particular tasks. This determination is based on the average workload values achievable by each of these revoking strategies when compared with the goal provided by the target workload range. In this particular case, all levels of the duration experiments outperformed all levels of the revoking threshold experiments. However, it is possible that in other situations the reverse might be true, or that some levels could be equivalent (e.g. a duration of 1 second could be equivalent to a revoking threshold of 10). The Revoking Strategies Duration Experiment and the Revoking Strategies Threshold Experiment reveal that in this study the workload and situational awareness decrease together, thus the preferred revoking strategy is contingent on the goal. A revoking strategy based on duration produces lower workload, but also produces lower situational awareness. Thus, if maintaining situational awareness is important, a revoking strategy based on revoking threshold may be the better option, since this strategy produces significant workload improvements over the baseline system, while achieving greater situational awareness than the duration strategy. Furthermore, within a particular strategy, there may be non-linear tradeoffs between workload and situational awareness that also need to be considered, such as the diminishing improvements in workload from increased automation duration, as found in the Revoking Strategies Duration Experiment.

CHAPTER 9
SUMMARY AND FUTURE DIRECTIONS

9.1. Introduction

This chapter highlights the key findings regarding simulation of cognitive workload and adaptive automation invoking and revoking strategies by reviewing the research question and sub-questions. The discussion then previews potential extensions to the work.

9.2. Summary of the Research

The goal of this research is to evaluate the effectiveness of automation revoking strategies. Specifically, this research seeks to capture the relative impacts of various automation revoking strategies on cognitive workload. This is the first study to model human behavior in order to simulate the performance of an adaptive system and automation revoking strategies.

The research undertaken is an attempt to bridge the fields of cognitive workload and situational awareness modeling, adaptive automation, and computer simulation modeling in a way not previously explored. In CHAPTER 2 a review of the literature of cognitive workload modeling, adaptive automation, and computer simulation applicable to cognitive workload modeling research is given. It is shown that there has been no other work that joins these fields for the modeling and prediction of cognitive workload using computer simulation, even for a specific class of human-computer integrated systems. In addition, the literature shows no clear performance, workload, or situational awareness benefits for system-based automation invoking and revoking methods using computer simulation models.

CHAPTER 3 presents the five-phase research approach that methodically answers five research sub-questions that help to ultimately address the primary research question of this investigation, which is: Can simulation-based modeling of cognitive workload be used for evaluating adaptive automation invoking and revoking strategies? The answers to the sub-questions reveal that DES is a viable and practical means for evaluating and predicting cognitive workload and situational awareness.

CHAPTER 4 describes the construction of the baseline discrete event simulation model, which is based on the general, yet, relevant case scenario of a human supervisory control situation that involves a system operator who receives and interprets intelligence outputs from multiple unmanned assets, and then identifies and reports potential threats and changes in the environment. This scenario is common in the military context, specifically in tactical-level counterinsurgency intelligence, surveillance, and reconnaissance (ISR) operations. However, the findings of this research investigation are not only relevant to ISR tasks, but they also have broader applicability to other human-computer integrated systems. The main tasks of the scenario are change detection tasks and threat detection tasks. The chapter also outlines the modeling assumptions considered in the construction of the discrete event simulation computer simulation model of the case scenario.

CHAPTER 5 explains the successful verification and validation of the baseline DES model of the ISR case scenario. The successful model validation demonstrates that discrete event simulation is a viable alternative to live trials involving human participants. The simulations results are as predictive of time-weighted average cognitive workload as well-established subjective methods such as the ISA and NASA-TLX, and it is more predictive than many surrogate physiological measures. Thus, it addresses Sub-Question 1, “Can simulation modeling predict cognitive workload as well as established measures of cognitive workload?" CHAPTER 5 also demonstrates how DES modeling can be used to evaluate alternative system design configurations. Hence, it answers research Sub-Question 2, "Can computer simulation modeling be used to evaluate system designs based on predicted cognitive workload?" Discrete event simulation can quite effectively be used to compare the relative workload differentials that different design alternatives produce, allowing for the identification of a preferred design solution without the costly and time-consuming burden of prototyping and live field testing.

CHAPTER 6 describes an enhancement to the validated baseline model in which channel interference is considered. It also addresses Sub-Question 3, "How can simulation modeling be used to determine the target level or range of cognitive workload scores for adaptive automation?" When establishing a target workload range, it is important to characterize the workload-performance relationship in order to determine whether the particular task (or set of tasks) in question falls along an increasing, decreasing, or flat portion of the workload-performance continuum. It is also possible for a task (or set of tasks) to span multiple segments of this continuum. Once this relationship is known, the objective to minimize workload, maximize workload, or maintain workload within a target range should be easier to identify. Another important consideration when establishing the target range is to consider whether there are any objective performance goals. These performance goals also influence the identification of the target workload range.

CHAPTER 7 summarizes the incorporation of the adaptive automation feature in the baseline DES model with interference. The analysis in this chapter addresses Sub-Question 4, "How can simulation modeling be used to determine a preferred invoking threshold for adaptive automation?" When selecting an invoking strategy, both the workload and situational awareness impacts of the thresholds must be considered.

Finally, CHAPTER 8 summarizes the empirical analysis of the invoking and revoking threshold values as well as the duration the automation is invoked. The analysis in this chapter addresses Sub-Question 5, "How can simulation modeling be used to determine a preferred revoking strategy for adaptive automation?" The revoking strategy analysis demonstrates how to analyze expected cognitive workload and situational awareness values for various levels of revoking thresholds and automation durations in order to determine a preferred revoking strategy. For the specific case in this study, the analysis reveals that a fixed minimum automation duration is preferred for these particular tasks. Also, this research investigation reveals that the workload and situational awareness decrease together; thus, the preferred revoking strategy is contingent on the desired goal. For example, a revoking strategy based on duration produces lower workload, but also results in decreased situational awareness. Thus, if maintaining situational awareness is important, a revoking strategy based on revoking threshold may be more appropriate, achieving greater situational awareness than the duration strategy. Furthermore, within a particular strategy, there may be non-linear tradeoffs between workload and situational awareness that also need to be considered, such as the diminishing improvements in workload from increased automation duration.

Overall, the results from this research study are quite significant since it demonstrates the effectiveness of using computer simulation, specifically discrete event simulation modeling, to model and predict cognitive workload and situational awareness. In fact, simulation is as effective as some well-established subjective measures, and more effective than numerous physiological measures. This would allow for system designers and developers faster identification of preferred design configurations without the costly and time-consuming burden of prototyping and live field testing, thus using only a fraction of the time and effort.

9.3. Future Research Directions

The research presented in this dissertation and the conclusions drawn has laid sufficient foundation for a number of opportunities for extending the work. The future directions of this research include examining alternative revoking strategies, refining the cognitive workload calculations, and revising the dependent variables.

9.3.1. Examine Alternative Revoking Strategies

This study provides a comparative evaluation of revoking strategies based on duration or threshold. However, there is potential to examine other revoking strategies, such as hybrid strategies that combine a minimum duration with a threshold.

9.3.2. Refine the Cognitive Workload Adjustments

This study uses Bierbaum's (1989) adaptation of the VACP model to generate workload values. The VACP model is high in fidelity, with the ability to assign workload based on both resource channel and demand level. However, the VACP model's fidelity can be increased by incorporating operator skill levels, operator personal stressors, fatigue, and environmental stressors. An adjustment for operator skill levels would account for the same task being higher workload for novices than for experts. Even for a particular operator, the same task may vary in workload based on personal stressors such as emotional stress, lack of sleep, hunger, etc. Fatigue adjustments would adjust the experienced workload in the model over time in order to account for performance degradation and increased workload just from performing the task over an extended period of time. Environmental stressor such as heat, cold, humidity, and lighting can also impact workload and should be considered.

9.3.3. Expand the Set of Dependent Variables

The cognitive workload variable analyzed in this study is calculated over time at each instance that there is a change in the event log for the discrete simulation event log. Thus, the workload output from the simulation includes a workload profile over time, as shown in Section 6.1. However, for the analysis in this research, this workload profile is collapsed into a time-weighted average for each scenario-segment combination. Each scenario-segment combination is designed to have a unique task load, but within a particular scenario-segment combination the task load is relatively constant. This design is the justification for using time-weighted average. However, workload does vary slightly within scenario-segment combinations, and in other cases consisting of uneven workload, the workload profiles could vary considerably, rendering time-weighted averages less useful. Extending this analysis to consider the workload profile could provide a unique set of insights for tasks with variable task loads.

For this study, the primary dependent variable of interest is cognitive workload, with a secondary focus on situational awareness. In this research, both situational awareness variables capture activity that is occurring in the Change Detection task, thus, this analysis could be improved by also capturing the Threat Detection task in the situational awareness variable. The analysis of situational awareness could also be improved by deliberately designing situational awareness variables that measure all three levels of situational awareness as described by Endsley (1995).

Since cognitive workload is this the dependent variable of interest, this study incorporates performance as inputs into the models. For example, the probability of identifying an enemy soldier as a threat or identifying an icon appearance as an appearance are incorporated into the model as probabilities for decision paths. However, performance can also be of interest as a dependent variable. In order to design models to have performance be an output instead of an input, the models could use either workload or situational awareness values to influence the operator's performance results.

APPENDIX A:
DISCRETE EVENT SIMULATION MODEL NODE DESCRIPTIONS

A.1. Change Detection Task Nodes

Each node in Figures 22 and 23 contain a number of properties including task times, release conditions, beginning and ending effects, crew assignments, path exit logic, and workload values. This section discusses these properties for each node in detail. Table 36 summarizes the custom variable used in the Change Detection task.

Table 36: Change Detection Custom Variables

Variable Name	Values	Purpose
CHANGETYPE	0 to 3 0 = no change (start of the scenario) 1 = appearance 2 = disappearance 3 = movement	Identifies the type of change for the current change event
CUMDURATION	Integer	Tracks the cumulative duration for all change events that have occurred including the current change event
HUMAN SYSTEM	Boolean (True, False)	Batching variables
TASKDURATION	Integer	Time until next change event occurs
TIMEBLOCK	Integer	Counter variable to track which change event is currently taking place (all of the change events are numbered sequentially)

A.1.1. Start Node

The Start Node is a system task that begins the segment. For Segments A1, B1, and C1, this node creates two entities: one that follows the human operator’s task route, and one that follows the system’s task route. For segments A2, B2, and C2, this node creates a third entity which follows the ISA route. These entities are batched together in the end node, before proceeding on to the next segment.

A.1.2. Monitor Map Node

This node captures the monitoring task that is performed by the human operator. The task time is calculated by subtracting the `CUMDURATION` variable from the Clock variable. The `CUMDURATION` variable is updated in the Update System Variables node. This dynamic calculation of the Monitor Map task synchronizes the human operator's performance with the system, so that the operator "sees" a change at the same time as the system updates the current icon, which represents a change made to the situation map.

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 6.0 in the Visual dimension, which corresponds to Scan/Search Monitor.

Upon completion of this task, the entity faces a tactical branching logic, with the entity proceeding to the See Change node if the Clock is less than segment end time and proceeding to the End node if the clock is greater than or equal to the segment end time.

This node has an ending effect if the Clock is greater than the segment end time, then it updates a variable used to batch the multiple entities created by the Start Node. Once the batching occurs, this signifies that the operator has finished his or her portion of the current segment.

A.1.3. See Change Node

This node captures the human operator's reaction to a change on the situation map. The task time for this node is 0.24 seconds. This time is taken from the Simple Reaction Time micromodel in IMPRINT corresponding to an On or Off Response (Card, Moran, & Newell, 1983). The human operator is the primary assignee for this task, with no contingency operators.

This task has a VACP value of 5.0 in the Visual dimension, which corresponds to Discriminate (Detect Visual Differences), and a VACP value of 1.2 in the Cognitive Dimension, which corresponds to Alternative Selection.

Upon completion of this task, the entity faces a probabilistic branching logic, with the entity proceeding to the Identify Change node or returning to the Monitor Map node. This branching logic captures the occurrence of the operator not responding to a change that occurs, thus for some reason the operator misses the change. PRIME 2 participants are instructed to guess if they see a change but could not discriminate which type of change it is, thus a "no response" from the PRIME 2 data is assumed to be due to the participant actually missing the change (rather than indecision). Based on the first 30 participants, the likelihood of seeing a change is shown in Table 37.

Table 37: Percent of No Responses for Change Detection Task

Segment	Probability of No Response	Probability of Returning to Monitor Map Task	Probability of Moving On to the Identify Change Task
Segment A	22%	22%	78%
Segment B	23%	23%	77%
Segment C	35%	35%	65%

A.1.4. Identify Change Node

This node captures the identification task performed by the human operator. In this task, the operator uses their cognitive reasoning, memory, and visual perception to determine which type of change he or she has seen. The amount of time it takes to accomplish this task is derived from the PRIME 2 data for the first 30 participants. The specific distribution differs for each segment; see Table 38 for distribution information. For a complete discussion on fitting distributions see Section 4.8.2.

Table 38: Probability Distributions for the Identify Change Node in the Change Detection Task

Segment	Distribution	Parameters	Kolmogorov-Smirnov Test p-value
Segment A	LogNormal	$\mu = 1.64$ $\sigma = 0.61$	0.0977
Segment B	LogLogistic	$\alpha = 6.4634$ $\beta = 1.3856$	0.4678
Segment C	Logistic	$\mu = 1.17$ $\sigma = 0.24$	0.2792

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 5.0 in the Visual dimension, which corresponds to Discriminate/Detect Visual Differences, and a VACP value of 6.8 in the Cognitive dimension, which corresponds to Evaluation/Judgment of Several Aspects.

Upon completion of this task, the entity faces a tactical branching logic, in which the entity proceeds to the system Identify Appear, Identify Disappear, or Identify Movement node that corresponds to the type of change that occurred.

A.1.5. Identify Appear, Identify Disappear, and Identify Movement Nodes

These nodes perform system logic to direct the entity to select the appropriate button. These nodes are necessary because the button that will be selected by the operator depends on the change type that occurred, but is not necessarily the same as the change type that occurred. For example, in Segment A, if the actual change was an appearance, then there is a 92% chance that the operator will select the `APPEAR` button, a 1% chance that the operator will select the

DISAPPEAR button, and a 7% chance that the operator will select the MOVEMENT button. Table 39 displays the probabilities of a response given the actual change type for each segment.

Table 39: Change Identification Responses by Change Type and Segment

Actual, Response	Segment A	Segment B	Segment C
Appear, Appear	92.0%	86.0%	69.0%
Appear, Disappear	1.0%	3.0%	11.0%
Appear, Move	7.0%	11.0%	20.0%
Disappear, Appear	1.0%	2.0%	12.4%
Disappear, Disappear	96.0%	96.0%	77.3%
Disappear, Move	3.0%	3.0%	10.3%
Move, Appear	16.0%	16.0%	23.0%
Move, Disappear	8.0%	11.0%	17.0%
Move, Move	76.0%	73.0%	60.0%

Upon completion of this task, the entity uses probabilistic branching logic from Table 39 to proceed to the appropriate node. Since these nodes are solely to capture internal system logic, they are performed by the computer system and do not have any task time or workload values associated with them.

A.1.6. Select Appeared Button, Select Disappeared Button, and Select Movement Button Nodes

In these nodes the human operator selects the button that corresponds with the response for the type of change identified. This task assumes that the human operator rests the mouse cursor on the DISAPPEAR button, which is centered between the APPEAR and MOVEMENT buttons. Thus, the task time to select the DISAPPEAR button is simply 0.40 seconds, based on the Pushbutton micromodel in IMPRINT (Harris, Lavecchia, & Bittner, 1988). The task times for selecting the APPEAR and MOVEMENT buttons are 1.35 seconds each. These values are calculated using the Cursor Movement with Mouse micromodel in IMPRINT (Card et al., 1983). The cursor movement is equal to

$$1.03 + 0.096 \times \log_2(P_1/P_2 + 0.50) \tag{8}$$

where P_1 is equal to the distance to target (in pixels), and P_2 is equal to the size of target (in pixels). The monitor screen size is 11.75 x 18.75 inches and the resolution is 1680x1050. Thus, there are 89 pixels per inch. The distance from the center of the `DISAPPEAR` button to the center of the other buttons is 3.9375 inches, or approximately 350 pixels (P_1 value) The button sizes are 2.375 x 0.3125 inches, or 211 x 28 pixels, for a P_2 value of approximately 5,908 pixels. Thus, using Eq. 8, the time to move the cursor is 0.95 seconds, to which 0.40 seconds is added to account for the time to press the button.

The human operator is the primary assignee for these tasks, with no contingency operators. These tasks have a VACP value of 4.0 in the Visual dimension, which corresponds to Locate/Align, and a VACP value of 2.20 in the Fine Motor dimension, which corresponds to Discrete Actuation. Upon completion of these tasks, the entity faces a single branching logic, which proceeds back to the Monitor Map task.

A.1.7. System Wait Node

The System Wait node occurs in Segments A2, B1, B2, C1, and C2. The purpose of this node is to delay the start of the first change event, because the first change event does not happen immediately at the start of the scenario. The task time is the delay time and is calculated by subtracting the Clock time from the time that the current segment's first change occurs. This node is performed by the computer system and does not have any workload values associated with it. Upon completion of this task, the entity faces a single branching logic, in which the entity proceeds to the System: Make Changes Node.

This node has a beginning effect of updating the CUMDURATION value, to set the clock to the start time for the segment. This value will be used for the first Monitor Map task. The node also has an ending effect to update the CUMDURATION to adjust for the delay time incurred. An additional ending effect includes an update to the TASKDURATION variable, for use as the initial value in calculation the task time for the System: Make Changes node.

A.1.8. System: Make Changes Node

This node represents the change events that occur during the Change Detection task. See Appendix B for the Change Detection event logs. The task time for this node is the duration until the next change event and is represented by the system variable TASKDURATION. This node is performed by the computer system and does not have any workload values associated with it. Upon completion of this task, the entity faces a single branching logic, in which the entity proceeds to the Update System Variables Node.

This node has a beginning effect of setting the TASKDURATION based on the TIMEBLOCK variable. The TIMEBLOCK variable is a counter that keeps track of which change event is currently taking place (all of the change events are numbered sequentially). The node has an ending effect of updating the CHANGETYPE variable based on the TIMEBLOCK. The CHANGETYPE variable is an integer with values from 0 to 3, where 0 represents no change (start of the scenario), 1 represents an appearance, 2 represents a disappearance, and 3 represents a movement.

A.1.9. Update System Variables Node

The Update System Variables node performs system logic to update the CUMDURATION, TIMEBLOCK, and the batching variables. The CUMDURATION variable captures the cumulative duration of all the change events including the one that is in process. Thus, the difference between the current Clock time and the CUMDURATION is the amount of time until the next change event. This variable is used by the Monitor Map task. The CUMDURATION variable is updated as a beginning effect by adding the current TASKDURATION to the CUMDURATION's current value.

The TIMEBLOCK counter is updated as an ending effect by adding 1 to the current value. The node also has an ending effect that if the Clock is greater than the segment end time, then it updates the batching variable to signify that the System has finished its portion of the current segment.

Since this node is solely to capture internal system logic, it is performed by the computer system and does not have any task time or workload values associated with it. Upon completion of this node, the entity faces a tactical branching logic, in which it proceeds to the End node if the Clock is greater than or equal to the segment end time, otherwise it returns to the System: Make Changes Node to trigger the next change event.

A.1.10. System Wait for ISA Prompt Node

The purpose of the System Wait for ISA Prompt Node is to delay the ISA prompt until the appropriate time in the scenario. The ISA prompts always occur at clock times: 3:50, 8:50, and 13:50, in segments A2, B2, and C2, respectively. The task time is a fixed duration that delays the ISA prompt to its respective time. Since this node is solely to capture internal system logic, it is performed by the computer system and does not have any workload values associated with it. Upon completion of this node, the entity faces a single branching logic leading to the Listen to ISA Audio Prompt node.

A.1.11. Listen to ISA Audio Prompt Node

During the Listen to ISA Audio Prompt node, the human operator listens to the pre-recorded instruction: "Please rate your workload." This prompt lasts 1.7 seconds. This task has a VACP value of 3.0 in the Audio dimension, which corresponds to Interpret Simple Semantic Content (Speech). Upon completion of this node, the entity faces a single branching logic leading to the Decide Workload node.

A.1.12. Decide Workload Node

In the Decide Workload node, the human operator must decide what level to rate their workload. The task time for this node is represented by a probability distribution derived from a sample of audio files from the first 30 participants. The decide time is calculated by measuring the delay between the end of the ISA audio prompt and the beginning of the participant's oral response. Analyses of the data reveal that the probability distribution differs for single- and dual-task scenarios. For the single-task scenarios, the distribution is a LogLogistic, with a shape parameter $\alpha = 4.3682$ and a $\beta = 0.72904$. The Kolmogorov-Smirnov p-value for this distribution is 0.3674. See Section 4.8.1.1 for a detailed discussion of fitting these probability distributions.

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 6.8 in the Cognitive dimension, which corresponds to

Evaluation/Judgment of Several Aspects. Upon completion of this task, the entity faces a single branching logic leading to the Speak ISA Value node.

A.1.13. Speak ISA Value Node

The Speak ISA Value node captures the human operator speaking his or her ISA workload value. This spoken response is a single number between 1 and 5. The task time for this activity varies from 0.1 to 0.4 seconds and is represented by a discrete empirical probability distribution derived from a sample of audio files from the first 30 participants. Analysis of the data reveals that the probability distribution differs for single- and dual-task scenarios. Table 40 provides the probabilities for the ISA verbal response time for the single-task scenarios.

Table 40: Probabilities for Speaking ISA Values in Single-Task Scenarios

0.1 seconds	0.2 seconds	0.3 seconds	0.4 seconds
2%	58%	35%	5%

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 2.0 in the Speech dimension, which corresponds to Simple (1-2 Words). Upon completion of this task, the entity is disposed.

A.1.14. End Node

The End node is a system task that ends the segment. The multiple entities created by the Start node are batched together in this node before proceeding on to the next segment. There is no task time or workload associated with this node.

A.2. Threat Detection Task Nodes

Each node in Figure 19 and 25 contain a number of properties including task times, release conditions, beginning and ending effects, crew assignments, path exit logic, and workload values. This section discusses these properties for each node in detail. Table 41 summarizes the custom variable used in the Threat Detection task.

Table 41: Threat Detection Custom Variables

Variable Name	Values	Purpose
ACTORCHECK	Integer	Uses entity tags to determine if the actor in the Select Threat Queue is still visible on the OCU
ACTORTYPE	1 to 4 1 = Friendly Soldier 2 = Friendly Civilian 3 = Enemy Soldier 4 = Armed Civilian	Identifies the type of actor currently being generated
ARMEDCIVILIAN ENEMYSOLDIER FRIENDLYCIVILIAN FRIENDLYSOLDIER	Integer	Counter variables to track the number of actors visible on the OCU by actor type
TASKDURATIONTD	Non-Negative Number	Time until next change event occurs
THREATBUTTONACTIVATED	Boolean (True, False)	Tracks whether or not the THREAT DETECT button is currently activated
TIMEBLOCKTD	Integer	Counter variable to track which actor is currently being generated (all of the actor appearances are numbered sequentially)
VISIBLETIME	Non-Negative Number	Amount of time an actor is visible on the screen

A.2.1. Start Node

The Start node is a system task that begins the segment. For segments A1, B1, and C1, this node creates a single entity that proceeds to the Generate Actor node. For segments A2, B2, and C2, the Start node also creates a second entity which follows the ISA route.

A.2.2. Generate Actor Node

The Generate Actor node is a system task that begins the process of generating the actors that will appear on the screen for the operator to identify. See Appendix C for the Threat Detection event logs. The task has a beginning effect that assigns a value to the `TASKDURATIONTD` variable based on the `TIMEBLOCKTD` variable. As with the Change Detection variables, the `TIMEBLOCKTD` variable is a counter that keeps track of which actor generation event is currently taking place. The `TASKDURATIONTD` variable stores the amount of time until the next actor is to be generated. The task time for the Generate Actor node is the current value of `TASKDURATIONTD`.

The node has several ending effects. The first is to update the `TIMEBLOCKTD` counter by 1. The second ending effect is to assign the correct value to the `ACTORTYPE` variable based on the `TIMEBLOCKTD` variable. The `ACTORTYPE` variable is an integer with values from 1 to 4, 1 represents a Friendly Soldier, 2 represents a Friendly Civilian, 3 represents an Enemy Soldier, and 4 represents an Armed Civilian. The final ending effect is to update the `VISIBLETIME` based on `TIMEBLOCKTD` variable. The value of the `VISIBLETIME` variable is the amount of time that the current actor being generated will be visible on the screen.

Since this node is solely to capture internal system logic, it does not have any workload values associated with it. Upon completing of this task, the entity faces a tactical branching logic, in which it proceeds to node that corresponds with the current actor generation type: Generate Friendly Soldier, Generate Friendly Civilian, Generate Enemy Soldier, or Generate Armed Civilian.

A.2.3. Generate Friendly Soldier, Generate Friendly Civilian, Generate Enemy Soldier, and Generate Armed Civilian Nodes

The purpose of these nodes is to increase the actor counters and to tag the entities for reference by the Select Threat Queue node. Each actor type has a corresponding variable counter: `FRIENDLYSOLDIER`, `FRIENDLYCIVILIAN`, `ENEMYSOLDIER`, and `ARMEDCIVILIAN`. After the actor of that type is generated, the entity proceeds to the respective Generate node which has a beginning effect of increasing the respective counter by 1. These nodes also have an ending effect of tagging the entities with a unique number.

Since these nodes are solely to capture internal system logic, they are performed by the computer system and do not have any task time or workload values associated with them. Upon updating these variables, the entity encounters a multiple branching logic and is split into three entities (note that each of these three entities have the same “unique” identification number). The first entity proceeds to the respective Decrement node, the second entity proceeds to the Time Check node, and the third entity proceeds to the respective Identify node.

A.2.4. Decrement Friendly Soldier, Decrement Friendly Civilian, Decrement Enemy Soldier, and Decrement Armed Civilian Nodes

The purpose of these nodes is to create a holding place for actors that are visible on the screen. The task time is equal to the `VISIBLETIME` for that entity. These nodes are performed by the computer system and do not have any workload values associated with them. Upon completion of this task, the entity is disposed. These nodes have an ending effect of reducing the respective actor counter (`FRIENDLYSOLDIER`, `FRIENDLYCIVILIAN`, `ENEMYSOLDIER`, and `ARMEDCIVILIAN`) by 1.

A.2.5. Time Check Node

The Time Check node verifies the value of the current `TIMEBLOCKTD` to determine whether the system has completed generating actors for the current segment or if further actors still need to be generated. This is done through tactical branching logic in which the entity will return to the Generate Actor node if the `TIMEBLOCKTD` value is less than the final `TIMEBLOCKTD` value for that segment, and otherwise the entity proceeds to the Wait for Task End node. Since this node is solely to capture internal system logic, it does not have any task time or workload values associated with it.

A.2.6. Wait for Task End Node

The Wait for Task End node delays the final entity from proceeding to the End node until all of the allotted segment time has elapsed. This provides the delay necessary for the actors currently visible on the screen to be identified by the human operator and disposed of by the computer system. This task is performed by the system and does not have any workload values associated with it. Once the allotted time has elapsed, the entity flows to the End node.

A.2.7. Identify Friendly Soldier, Identify Friendly Civilian, Identify Enemy Soldier, and Identify Armed Civilian Nodes

The Identify nodes capture the identification task performed by the human operator. In this task, the operator views each actor on the screen and determines whether or not the actor is a threat. The task time for this identification is 0.45 seconds per actor and is taken from the Simple Reaction Time micromodel in IMPRINT corresponding to a Class Match (Card et al., 1983).

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 5.0 in the Visual dimension, which corresponds to Discriminate/Detect Visual Differences. The task also has a VACP value of 1.2 in the Cognitive dimension, which corresponds to Alternative Selection.

Upon completion of this task, the entity faces a probabilistic branching logic depending on the probability of identifying the current actor as a threat or non-threat. Table 42 presents the probability of detecting an actor as a threat for each actor type by segment. These probabilities are from the first 30 participants in the PRIME 2 study. The model, and these probabilities, assumes that each actor is deliberately identified.

Table 42: Probability of Identifying an Actor as a Threat by Segment

	% Detected as Threats		
	Segment A	Segment B	Segment C
Friendly Soldiers	0.2%	0.3%	0.1%
Friendly Civilians	0.1%	0.2%	0.2%
Enemy Soldiers	96.0%	95.6%	97.0%
Armed Civilians	94.2%	93.3%	93.0%

A.2.8. Non-Threat Node

If the operator identifies the actor as a non-threat, then the entity proceeds to the Non-Threat node. Since the operator does not have to perform any response to non-threats, this node is for collection purposes and does not have any task time or workload values associated with it. Entities that pass through this node are routed through to the Actor Identified disposal node.

A.2.9. Actor Identified Node

The purpose of the Actor Identified node is to collect entities for disposal. This task is for collection purposes and does not have any task time or workload values associated with it.

A.2.10. Threat Node

If the operator identifies the actor as a threat, then the entity proceeds to the Threat node. When the operator identifies a threat, he or she must report it. This node performs internal system logic by collecting all entities identified as threats to route them through the reporting sequence. This task does not have any task time or workload values associated with it. Entities in this task proceed to the Select Threat Queue node.

A.2.11. Select Threat Queue Node

Operators can only report one threat at a time. To report a threat, the operator selects the `THREAT DETECT` button and then clicks on the screen at the threat's current location. Once the `THREAT DETECT` button is selected, it becomes activated. After the operator selects a threat, the `THREAT DETECT` button is deactivated, and thus available to report another threat. The purpose of the Select Threat Queue is to hold entities waiting to be reported. This task prevents the `THREAT DETECT` button from being selected multiple times before a threat is selected. To do this, the Select Threat Queue has a release condition that the `THREATBUTTONACTIVATED` variable must be set to False. The Select Threat Queue then has a beginning effect that sets the `THREATBUTTONACTIVATED` variable to True.

Upon completion of this task, the entity faces a tactical branching logic in order to ensure that the operator can only report actors still visible on the screen. The system retrieves the tag of the entity to be reported and verifies that the match for that tag is still in the system (if it is, it will be in one of the Decrement nodes). If the entity does have a match in the system, then the entity will proceed to the Select Threat Button node, otherwise it will proceed to the Missed

Actors node. The Select Threat Queue captures internal system logic performed by the computer system and thus does not have any task time or workload values associated with it.

A.2.12. Missed Actors Node

The purpose of this node it collect and dispose of entities identified as threats but unable to be reported due to their no longer being visible. This task is performed by the system and does not have any task time or workload values associated with it. Before arriving at this node, the entity will have passed through the Select Threat Queue node, and thus will have set the `THREATBUTTONACTIVATED` variable to True. The next entity in the Select Threat Queue will not proceed until the `THREATBUTTONACTIVATED` variable is False, thus the Missed Actors node has an ending effect of setting the `THREATBUTTONACTIVATED` variable to False, so that subsequence entities may proceed through the reporting sequence.

A.2.13. Select Threat Button Node

In order to report an actor identified as a threat, the operator first uses his mouse to select the `THREAT DETECT` button. The model assumes that the resting position for the mouse cursor is on the `THREAT DETECT` button. Thus, the task time is 0.4 seconds, based on the Pushbutton micromodel in IMPRINT (Harris et al., 1988). However, observations of the operators reveal that the operators frequently "pre-load" the `THREAT DETECT` button. That is, the operators select the `THREAT DETECT` button even though they were not ready to report a threat. This pre-loading aids the operator by allowing him or her to select the threat immediately once a threat appears. To account for this in the model, if an entity arrives at the Threat node and there are no entities in the Select Threat Queue, the model uses a probabilistic expression to determine whether or not the operator had pre-loaded the `THREAT DETECT` button. If the button is pre-loaded the task time is zero instead of 0.4 seconds. The probability of pre-loading is based on the PRIME 2 data for the first 30 participants and is presented by segment in Table 43. For a detailed discussion on obtaining these probabilities, see Section 4.8.

Table 43: Probabilities for Pre-loading `THREAT DETECT` button by Segment

	% Preloaded
Segment A	38%
Segment B	43%
Segment C	30%

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 4.0 in the Visual dimension, which corresponds to Locate/Align, and a VACP value of 2.2 in the Fine Motor dimension, which corresponds to Discrete Actuation.

Upon completion of this task, the entity faces a probabilistic branching logic to determine whether the operator selects an actor as the threat or misaligns his or her mouse cursor, and thus selects a non-actor. Table 44 displays the probability of selecting a non-actor by segment.

Table 44: Probability of Selecting a Non-Actor

	Probability of Selecting Non-Actor
Segment A	3.7%
Segment B	3.3%
Segment C	2.1%

A.2.14. Select Threat Node

The Select Threat node captures the task of re-locating the threat after selecting the `THREAT DETECT` button, tracking the mouse cursor to the threat, and then clicking on the threat. The time to perform this task is a probability distribution derived from the first 30 participants of the PRIME 2 study. The distributions are shown in Table 45; derivations of these distributions are discussed in Section 4.8.

Table 45: Probability Distributions for Selecting Threat in the Threat Detection Task

Segment	Distribution	Parameters	Kolmogorov-Smirnov Test p-value
Segment A	Pearson5	$\alpha = 10.991$ $\beta = 16.37$	0.9759
Segment B	Pearson5	$\alpha = 7.7987$ $\beta = 11.542$	0.9144
Segment C	Pearson5	$\alpha = 7.2163$ $\beta = 9.7574$	0.4593

The human operator is the primary assignee for this task, with no contingency operators. This task has a VACP value of 4.0 in the Visual dimension, which corresponds to Locate/Align, and a VACP value of 4.6 in the Fine Motor dimension, which corresponds to Tracking.

Upon completion of this task, the entity proceeds to the Actor Identified disposal node. The Select Threat node has an ending effect that returns the `THREATBUTTONACTIVATED` variable to False.

A.2.15. Select Non-Actor Node

When reporting a threat, the operator occasionally misses the actor that he or she has identified, for example by accidently clicking next to the actor or between the actor's legs. When this occurs the performance event log registers that the operator has identified Terrain or Sky as a threat. This probability of this occurrence is shown in Table 44 above. The Select Non-Actor node captures there rare occurrences. Since the operator intended to select an actor, this node functions exactly like the Select Threat node, with the same task time probabilities, workload, and ending effects. Upon completion of this task, the entity proceeds to the Actor Identified disposal node.

A.2.16. ISA Task Sequence Nodes

The ISA Task Sequence Nodes in the Threat Detection model operate exactly the same as the ISA nodes in the Change Detection model. For detailed description of the System Wait for ISA Prompt, Listen to ISA Audio Prompt, Decide Workload, and Speak ISA Value nodes, see Section A.1.10 through Section A.1.13.

A.2.17. End Node

The End node is a system task that ends the segment. There is no task time or workload associated with this node. Upon exiting this node, the entity proceeds to the next segment.

A.3. Dual-Task Nodes

The dual-task scenarios do not have any unique nodes. Table 46 summarizes the custom variable used in the dual-task scenarios.

Table 46: Dual-Task Custom Variables

Variable Name	Values	Purpose
`CDBUTTONACTIVATED`	Boolean (True, False)	Tracks whether or not the operator is currently responding to a change event using the mouse; prevents simultaneous responses for change events and threat reporting
`THREATBUTTONACTIVATED`	Boolean (True, False)	Tracks whether or not the operator is currently reporting a threat using the mouse; prevents simultaneous responses for change events and threat reporting

A.3.1. Other Modifications for Dual-Task Scenarios

The dual-tasks scenarios assume that the resting place for the mouse is on the DISAPPEARED button. This location was chosen based on consultation with subject matter experts. The reasoning behind this decision is that the DISAPPEARED button is located approximately in the center of the screen. Furthermore, the tendency to "pre-load" the `THREAT DETECT` button also provides an advantage to maintaining the mouse closer to the change detection response buttons. Thus, for the Threat Detection portions of the dual-task scenarios, the task time for the Select Threat Button node is 1.36 seconds. This value is also calculated using Eq 8. The monitor screen size is 11.75 x 18.75 inches and the resolution is 1680x1050. Thus, there are 89 pixels per inch. The distance from the center of the `DISAPPEAR` button to the center of the `THREAT DETECT` button is 7.625 inches, or approximately 679 pixels (P_1 value) The button size is 1.625 x 0.5 inches, or 145 x 45 pixels, for a P_2 value of approximately 6525 pixels. Thus, the time to move the cursor is 0.96 seconds, to which 0.4 seconds is added to account for the time to press the mouse button.

The Decide Workload and Speak ISA Value nodes differ in their task time probability distributions from the single-task nodes. For the dual-tasks, the distribution of the Decide Workload node is a Pearson5, with a shape parameter α = 3.7144 and a scale parameter β = 2.9176. The Kolmogorov-Smirnov p-value for this distribution is 0.3088. Table 47 provides the probabilities for the ISA verbal response time.

Table 47: Probabilities for Speaking ISA Values in Dual Task Scenarios

0.1 seconds	0.2 seconds	0.3 seconds	0.4 seconds
2%	38%	50%	10%

A.4. Adaptive Automation Nodes

The adaptive automation DES model builds upon the enhanced DES model presented in CHAPTER 6, by incorporating adaptive automation into the model. This section discusses the additional nodes in the adaptive models, not found in the baseline DES model. Table 36 summarizes the custom variable used in the adaptive automation models.

Table 48: Adaptive Automation Custom Variables

Variable Name	Values	Purpose
AACHECKDURATION	0.1 2	Frequency (in seconds) in which the operator's workload value is checked to determine whether or not to invoke adaptive automation
CHANGESEEN	Boolean (True, False)	Tracks whether or not a change event has occurred; used for system logic routing purposes
INVOKEAA	Boolean (True, False)	Indicates whether or not the operator's workload value is above a given threshold
AADURATION	0.1 1 2 5 10 15 20	Amount of time (in seconds) that the automation is active before re-checking the threshold/revoking automation
INVOKINGTHRESHOLD	10 20 23 24 25 30 40	Workload value used to invoke automation
REVOKINGTHRESHOLD	5 10 15 20 22 23 24 25 30 40	Workload value used to revoke automation

A.4.1. Modifications to Monitor Map Node

The Monitor Map task duration expression is updated. Previously, the Monitor Map task duration was equal to the time remaining until the next change occurred. In the adaptive automation DES model, the Monitor Map task duration is equal to established frequency for checking the operator's workload threshold or the time remaining until the next change occurs, whichever is less. This established frequency is represented by the variable, `AACHECKDURATION`. The Monitor Map task also has an update to the ending effects that sets the value of a Boolean variable, `CHANGESEEN`, to indicate whether the task completion is due to a change being seen or due to the need to check the workload level for potentially invoking the adaptive automation.

A.4.2. AA Check Node

The entity representing the operator's task flow then proceeds from the Monitor Map task to the AA Check node. This node has no time associated with it and serves to route the entity appropriately. The node has a beginning effect which compares the operator's current workload to the value of the `INVOKINGTHRESHOLD` variable. The AA Check node has a VACP value of 6.0 in the Visual dimension, which corresponds to Scan/Search Monitor. This workload value is the same as that of the Monitor Map task, and is included in this node so that the current workload calculations include the operator's workload for both monitoring the map and performing the Threat Detection task. Since the AA Check node has no task time associated with it, this workload does not impact the operator's time-averaged workload. If the operator's workload is greater than or equal to the `INVOKINGTHRESHOLD`, then the Boolean variable, `INVOKEAA`, is set to true, otherwise it is set to false. The tactical branching logic for this node will route the entity to the AA node if the `INVOKEAA` variable is true. If the variable is false, the entity will be routed to the See Change node if the `CHANGESEEN` variable is true. Otherwise the entity will be routed back to the Monitor Map task.

A.4.3. AA Node

The AA node represents the system having the automation turned on. While this occurs, the operator does not experience any workload from the Change Detection portion of the task network, rather he or she is only experience workload from the Threat Detection task. The task time for the AA node is determined by the variable, `AADURATION`. When the task concludes the operator's current workload is again compared to the value of the `REVOKINGTHRESHOLD` variable. If the operator's workload is greater than or equal to the `REVOKINGTHRESHOLD`, then the entity returns to the AA node. Otherwise, the entity is routed back to the Monitor Map node.

APPENDIX B:
PRIME 2 STUDY CHANGE DETECTION EVENT LOGS

Table 49: Change Detection Event Log, Variable Event Rate

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
1	None	0	A1	0.06
2	Move	0.06	A1	0.14
3	Appear	0.2	A1	0.06
4	Disappear	0.26	A1	0.14
5	Move	0.4	A1	0.06
6	Appear	0.46	A1	0.06
7	Disappear	0.52	A1	0.14
8	Appear	1.06	A1	0.14
9	Disappear	1.2	A1	0.14
10	Move	1.34	A1	0.06
11	Move	1.4	A1	0.06
12	Appear	1.46	A1	0.14
13	Disappear	2	A1	0.14
14	Move	2.14	A1	0.06
15	Move	2.2	A1	0.16
16	Move	2.36	A2	0.06
17	Disappear	2.42	A2	0.06
18	Appear	2.48	A2	0.06
19	Appear	2.54	A2	0.14
20	Disappear	3.08	A2	0.14
21	Move	3.22	A2	0.06
22	Disappear	3.28	A2	0.14
23	Appear	3.42	A2	0.14
24	Move	3.56	A2	0.06
25	Move	4.02	A2	0.14
26	Appear	4.16	A2	0.14
27	Disappear	4.3	A2	0.06
28	Move	4.36	A2	0.14
29	Move	4.5	A2	0.14
30	Move	5.08	B1	0.07
31	Disappear	5.15	B1	0.03
32	Appear	5.18	B1	0.03
33	Appear	5.21	B1	0.07
34	Disappear	5.28	B1	0.03
35	Move	5.31	B1	0.03
36	Appear	5.34	B1	0.07
37	Move	5.41	B1	0.03
38	Disappear	5.44	B1	0.07

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
39	Appear	5.51	B1	0.03
40	Move	5.54	B1	0.03
41	Disappear	5.57	B1	0.07
42	Move	6.04	B1	0.03
43	Appear	6.07	B1	0.07
44	Disappear	6.14	B1	0.03
45	Disappear	6.17	B1	0.03
46	Move	6.2	B1	0.07
47	Appear	6.27	B1	0.07
48	Disappear	6.34	B1	0.03
49	Move	6.37	B1	0.07
50	Appear	6.44	B1	0.03
51	Disappear	6.47	B1	0.07
52	Appear	6.54	B1	0.07
53	Move	7.01	B1	0.03
54	Appear	7.04	B1	0.03
55	Disappear	7.07	B1	0.07
56	Disappear	7.14	B1	0.07
57	Appear	7.21	B1	0.19
58	Move	7.4	B2	0.03
59	Disappear	7.43	B2	0.07
60	Appear	7.5	B2	0.07
61	Disappear	7.57	B2	0.03
62	Appear	8	B2	0.03
63	Move	8.03	B2	0.03
64	Appear	8.06	B2	0.07
65	Move	8.13	B2	0.03
66	Disappear	8.16	B2	0.03
67	Move	8.19	B2	0.03
68	Disappear	8.22	B2	0.07
69	Appear	8.29	B2	0.07
70	Move	8.36	B2	0.07
71	Disappear	8.43	B2	0.03
72	Appear	8.46	B2	0.03
73	Appear	8.49	B2	0.07
74	Disappear	8.56	B2	0.07
75	Move	9.03	B2	0.07
76	Disappear	9.1	B2	0.07
77	Move	9.17	B2	0.07

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
78	Appear	9.24	B2	0.03
79	Move	9.27	B2	0.07
80	Disappear	9.34	B2	0.07
81	Appear	9.41	B2	0.07
82	Appear	9.48	B2	0.03
83	Disappear	9.51	B2	0.03
84	Disappear	9.54	B2	0.03
85	Appear	9.57	B2	0.08
86	Appear	10.11	C1	0.04
87	Disappear	10.15	C1	0.03
88	Move	10.18	C1	0.02
89	Move	10.2	C1	0.03
90	Appear	10.23	C1	0.01
91	Disappear	10.24	C1	0.02
92	Move	10.26	C1	0.02
93	Disappear	10.28	C1	0.01
94	Appear	10.29	C1	0.03
95	Disappear	10.32	C1	0.02
96	Appear	10.34	C1	0.04
97	Move	10.38	C1	0.03
98	Appear	10.41	C1	0.04
99	Move	10.45	C1	0.03
100	Disappear	10.48	C1	0.04
101	Disappear	10.52	C1	0.03
102	Move	10.55	C1	0.04
103	Appear	10.59	C1	0.03
104	Disappear	11.02	C1	0.02
105	Move	11.04	C1	0.03
106	Appear	11.07	C1	0.04
107	Appear	11.11	C1	0.01
108	Disappear	11.12	C1	0.02
109	Move	11.14	C1	0.01
110	Appear	11.15	C1	0.04
111	Move	11.19	C1	0.01
112	Disappear	11.2	C1	0.02
113	Move	11.22	C1	0.01
114	Appear	11.23	C1	0.01
115	Disappear	11.24	C1	0.04
116	Disappear	11.28	C1	0.02

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
117	Move	11.3	C1	0.01
118	Appear	11.31	C1	0.04
119	Disappear	11.35	C1	0.01
120	Move	11.36	C1	0.02
121	Appear	11.38	C1	0.03
122	Move	11.41	C1	0.02
123	Disappear	11.43	C1	0.01
124	Appear	11.44	C1	0.04
125	Move	11.48	C1	0.03
126	Disappear	11.51	C1	0.04
127	Appear	11.55	C1	0.01
128	Appear	11.56	C1	0.01
129	Disappear	11.57	C1	0.04
130	Move	12.01	C1	0.02
131	Move	12.03	C1	0.03
132	Disappear	12.06	C1	0.04
133	Appear	12.1	C1	0.03
134	Move	12.13	C1	0.03
135	Appear	12.16	C1	0.02
136	Disappear	12.18	C1	0.02
137	Disappear	12.2	C1	0.01
138	Appear	12.21	C1	0.03
139	Move	12.24	C1	0.02
140	Move	12.26	C1	0.02
141	Move	12.28	C1	0.14
142	Disappear	12.42	C2	0.02
143	Move	12.44	C2	0.01
144	Appear	12.45	C2	0.03
145	Disappear	12.48	C2	0.02
146	Move	12.5	C2	0.02
147	Appear	12.52	C2	0.01
148	Move	12.53	C2	0.04
149	Appear	12.57	C2	0.03
150	Disappear	13	C2	0.02
151	Move	13.02	C2	0.03
152	Disappear	13.05	C2	0.02
153	Appear	13.07	C2	0.03
154	Appear	13.1	C2	0.04
155	Disappear	13.14	C2	0.03

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
156	Move	13.17	C2	0.02
157	Move	13.19	C2	0.01
158	Appear	13.2	C2	0.02
159	Disappear	13.22	C2	0.01
160	Move	13.23	C2	0.01
161	Disappear	13.24	C2	0.04
162	Appear	13.28	C2	0.02
163	Disappear	13.3	C2	0.03
164	Appear	13.33	C2	0.02
165	Move	13.35	C2	0.03
166	Move	13.38	C2	0.03
167	Appear	13.41	C2	0.04
168	Disappear	13.45	C2	0.02
169	Disappear	13.47	C2	0.03
170	Appear	13.5	C2	0.04
171	Move	13.54	C2	0.01
172	Appear	13.55	C2	0.04
173	Disappear	13.59	C2	0.03
174	Move	14.02	C2	0.04
175	Appear	14.06	C2	0.01
176	Move	14.07	C2	0.02
177	Disappear	14.09	C2	0.01
178	Appear	14.1	C2	0.04
179	Disappear	14.14	C2	0.01
180	Move	14.15	C2	0.01
181	Disappear	14.16	C2	0.04
182	Move	14.2	C2	0.02
183	Appear	14.22	C2	0.03
184	Move	14.25	C2	0.04
185	Appear	14.29	C2	0.01
186	Disappear	14.3	C2	0.02
187	Disappear	14.32	C2	0.01
188	Appear	14.33	C2	0.04
189	Move	14.37	C2	0.01
190	Appear	14.38	C2	0.03
191	Disappear	14.41	C2	0.04
192	Move	14.45	C2	0.03
193	Disappear	14.48	C2	0.02
194	Move	14.5	C2	0.04

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
195	Appear	14.54	C2	0.03
196	Move	14.57	C2	0.04
197	Move	15.01	C2	0.06

Table 50: Change Detection Event Log, Constant Event Rate

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
1	None	0	A1	0.04
2	Move	0.04	A1	0.03
3	Appear	0.07	A1	0.07
4	Disappear	0.14	A1	0.03
5	Disappear	0.17	A1	0.07
6	Move	0.24	A1	0.07
7	Appear	0.31	A1	0.03
8	Move	0.34	A1	0.03
9	Disappear	0.37	A1	0.07
10	Appear	0.44	A1	0.07
11	Move	0.51	A1	0.03
12	Appear	0.54	A1	0.03
13	Disappear	0.57	A1	0.07
14	Appear	1.04	A1	0.07
15	Disappear	1.11	A1	0.03
16	Move	1.14	A1	0.03
17	Disappear	1.17	A1	0.07
18	Appear	1.24	A1	0.07
19	Move	1.31	A1	0.07
20	Disappear	1.38	A1	0.07
21	Move	1.45	A1	0.03
22	Appear	1.48	A1	0.03
23	Move	1.51	A1	0.07
24	Appear	1.58	A1	0.07
25	Disappear	2.05	A1	0.07
26	Disappear	2.12	A1	0.03
27	Appear	2.15	A1	0.03
28	Appear	2.18	A1	0.03
29	Disappear	2.21	A1	0.09
30	Disappear	2.36	A2	0.07
31	Appear	2.43	A2	0.07

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
32	Move	2.5	A2	0.07
33	Appear	2.57	A2	0.03
34	Disappear	3	A2	0.03
35	Move	3.03	A2	0.03
36	Disappear	3.06	A2	0.07
37	Move	3.13	A2	0.03
38	Appear	3.16	A2	0.07
39	Disappear	3.23	A2	0.03
40	Move	3.26	A2	0.03
41	Appear	3.29	A2	0.07
42	Disappear	3.36	A2	0.03
43	Appear	3.39	A2	0.07
44	Move	3.46	A2	0.03
45	Move	3.49	A2	0.07
46	Disappear	3.56	A2	0.07
47	Appear	4.03	A2	0.07
48	Appear	4.1	A2	0.07
49	Move	4.17	A2	0.03
50	Disappear	4.2	A2	0.03
51	Appear	4.23	A2	0.03
52	Disappear	4.26	A2	0.03
53	Move	4.29	A2	0.07
54	Appear	4.36	A2	0.03
55	Disappear	4.39	A2	0.07
56	Disappear	4.46	A2	0.03
57	Appear	4.49	A2	0.15
58	Move	5.08	B1	0.03
59	Appear	5.11	B1	0.03
60	Disappear	5.14	B1	0.03
61	Move	5.17	B1	0.07
62	Disappear	5.24	B1	0.07
63	Appear	5.31	B1	0.07
64	Appear	5.38	B1	0.03
65	Disappear	5.41	B1	0.07
66	Move	5.48	B1	0.03
67	Disappear	5.51	B1	0.03
68	Move	5.54	B1	0.03
69	Appear	5.57	B1	0.07
70	Disappear	6.04	B1	0.07

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
71	Move	6.11	B1	0.03
72	Appear	6.14	B1	0.03
73	Move	6.17	B1	0.07
74	Disappear	6.24	B1	0.07
75	Appear	6.31	B1	0.07
76	Disappear	6.38	B1	0.03
77	Appear	6.41	B1	0.07
78	Move	6.48	B1	0.03
79	Disappear	6.51	B1	0.07
80	Appear	6.58	B1	0.07
81	Move	7.05	B1	0.07
82	Disappear	7.12	B1	0.07
83	Appear	7.19	B1	0.03
84	Appear	7.22	B1	0.03
85	Disappear	7.25	B1	0.06
86	Move	7.4	B2	0.03
87	Appear	7.43	B2	0.03
88	Disappear	7.46	B2	0.07
89	Move	7.53	B2	0.03
90	Appear	7.56	B2	0.07
91	Disappear	8.03	B2	0.07
92	Appear	8.1	B2	0.07
93	Disappear	8.17	B2	0.07
94	Move	8.24	B2	0.07
95	Disappear	8.31	B2	0.07
96	Appear	8.38	B2	0.03
97	Move	8.41	B2	0.03
98	Disappear	8.44	B2	0.03
99	Appear	8.47	B2	0.03
100	Move	8.5	B2	0.03
101	Disappear	8.53	B2	0.03
102	Move	8.56	B2	0.03
103	Appear	8.59	B2	0.07
104	Move	9.06	B2	0.07
105	Appear	9.13	B2	0.03
106	Disappear	9.16	B2	0.07
107	Disappear	9.23	B2	0.03
108	Move	9.26	B2	0.07
109	Appear	9.33	B2	0.07

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
110	Appear	9.4	B2	0.03
111	Disappear	9.43	B2	0.07
112	Disappear	9.5	B2	0.03
113	Appear	9.53	B2	0.12
114	Disappear	10.11	C1	0.07
115	Move	10.18	C1	0.07
116	Appear	10.25	C1	0.07
117	Appear	10.32	C1	0.03
118	Disappear	10.35	C1	0.07
119	Move	10.42	C1	0.03
120	Move	10.45	C1	0.07
121	Disappear	10.52	C1	0.03
122	Appear	10.55	C1	0.07
123	Disappear	11.02	C1	0.03
124	Appear	11.05	C1	0.07
125	Move	11.12	C1	0.03
126	Move	11.15	C1	0.03
127	Appear	11.18	C1	0.07
128	Disappear	11.25	C1	0.03
129	Move	11.28	C1	0.07
130	Disappear	11.35	C1	0.07
131	Appear	11.42	C1	0.03
132	Move	11.45	C1	0.03
133	Appear	11.48	C1	0.03
134	Disappear	11.51	C1	0.07
135	Appear	11.58	C1	0.07
136	Disappear	12.05	C1	0.07
137	Move	12.12	C1	0.03
138	Disappear	12.15	C1	0.03
139	Appear	12.18	C1	0.07
140	Appear	12.25	C1	0.03
141	Disappear	12.28	C1	0.10
142	Disappear	12.42	C2	0.03
143	Appear	12.45	C2	0.07
144	Move	12.52	C2	0.03
145	Appear	12.55	C2	0.07
146	Move	13.02	C2	0.07
147	Disappear	13.09	C2	0.03
148	Appear	13.12	C2	0.03

Event Number	Event Type	Simulation Time (min.second)	Segment	Task Duration (min.second)
149	Disappear	13.15	C2	0.03
150	Move	13.18	C2	0.03
151	Appear	13.21	C2	0.03
152	Move	13.24	C2	0.03
153	Disappear	13.27	C2	0.07
154	Appear	13.34	C2	0.07
155	Move	13.41	C2	0.03
156	Disappear	13.44	C2	0.03
157	Move	13.47	C2	0.07
158	Appear	13.54	C2	0.07
159	Disappear	14.01	C2	0.07
160	Appear	14.08	C2	0.07
161	Disappear	14.15	C2	0.03
162	Move	14.18	C2	0.03
163	Disappear	14.21	C2	0.07
164	Move	14.28	C2	0.07
165	Appear	14.35	C2	0.07
166	Appear	14.42	C2	0.03
167	Appear	14.45	C2	0.03
168	Disappear	14.48	C2	0.07
169	Disappear	14.55	C2	0.12

APPENDIX C:
PRIME 2 STUDY THREAT DETECTION EVENT LOGS

Table 51: Threat Detection Event Log, Variable Event Rate

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
None	0	0	0	A1
US Military 08	148	4.0	6.216	A1
US Military 10	149	4.0	11.504	A1
Arab Female 03	177	4.0	15.788	A1
US Military 11	139	4.0	20.077	A1
US Military 04	144	4.0	25.372	A1
Arab Male 05	174	4.0	27.386	A1
Foreign Military 03	183	4.0	30.157	A1
US Military 01	141	4.0	35.957	A1
US Military 03	143	9.745	36.757	A1
Arab Male 04	171	15.536	35.749	A1
Arab Male 02	168	19.067	37.259	A1
US Military 05	134	22.091	36.512	A1
Insurgent 08	188	25.865	35.509	A1
Arab Female 02	173	29.121	36.517	A1
Arab Male 03	169	34.663	38.04	A1
US Military 07	136	36.927	36.533	A1
Arab Male 01	166	41.475	36.004	A1
US Military 01	119	44.491	37.016	A1
Arab Female 03	164	48.27	36.762	A1
Insurgent 07	31	52.035	35.771	A1
Arab Female 01	154	53.801	36.525	A1
Arab Male 03	152	56.072	37.024	A1
Arab Male 05	150	62.126	37.514	A1
US Military 03	121	66.399	36.758	A1
US Military 06	124	71.443	37.774	A1
US Military 03	132	74.213	36.523	A1
Insurgent 03	186	76.47	37.037	A1
US Military 08	126	81.758	36.267	A1
Arab Male 02	153	84.526	36.512	A1
Arab Female 01	159	90.073	35.984	A1
US Military 11	128	92.591	36.491	A1
Foreign Military 01	181	96.378	37.233	A1
Arab Male 02	156	99.136	37.494	A1
US Military 12	129	101.901	36.745	A1
Arab Female 02	160	104.922	37.253	A1
US Military 07	77	156.0	7.573	A2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Male 02	107	156.0	12.334	A2
Arab Female 01	108	156.0	16.386	A2
US Military 01	67	156.0	18.411	A2
US Military 11	65	156.0	24.448	A2
Foreign Military 02	111	156.0	27.457	A2
Arab Male 04	94	157.281	28.934	A2
US Military 08	36	157.3	34.459	A2
US Military 06	57	158.034	37.253	A2
Arab Male 01	98	163.824	35.744	A2
Insurgent 01	115	166.333	37.529	A2
US Military 10	37	171.121	38.032	A2
Arab Female 03	81	174.156	36.251	A2
US Military 07	35	177.182	38.017	A2
US Military 05	116	183.708	36.513	A2
US Military 02	6	188.227	37.275	A2
US Military 01	5	192.517	37.257	A2
US Military 08	78	199.568	37.002	A2
Arab Male 01	102	203.109	36.485	A2
Arab Male 02	106	206.386	37.738	A2
Foreign Military 04	33	210.407	37.235	A2
Arab Female 01	8	212.932	36.724	A2
US Military 11	38	217.71	35.714	A2
Arab Male 04	85	218.969	36.725	A2
Arab Male 04	100	222.236	36.725	A2
US Military 05	52	226.754	37.74	A2
Arab Female 02	88	231.284	35.489	A2
Foreign Military 03	112	236.07	36.976	A2
US Military 07	58	236.07	36.476	A2
Arab Female 01	87	241.605	36.96	A2
US Military 02	68	244.378	35.7	A2
Arab Male 05	103	245.134	37.963	A2
Arab Male 04	7	250.411	37.716	A2
US Military 04	74	253.174	35.958	A2
Arab Male 02	91	255.944	37.216	A2
US Military 01	189	308.0	6.083	B1
US Military 02	190	308.0	11.114	B1
Arab Male 04	247	308.0	12.370	B1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 03	191	308.0	16.665	B1
Arab Female 03	248	308.0	17.169	B1
Foreign Military 04	250	308.0	20.441	B1
US Military 04	192	308.0	22.201	B1
Arab Male 03	249	308.0	22.957	B1
US Military 05	193	308.0	23.962	B1
Arab Female 02	246	308.0	27.246	B1
US Military 06	194	308.0	29.251	B1
Insurgent 09	257	308.0	30.264	B1
Arab Male 05	245	308.0	33.557	B1
US Military 07	195	308.0	34.059	B1
Arab Female 01	244	308.794	36.791	B1
US Military 08	196	309.546	38.052	B1
Arab Male 02	243	312.825	37.29	B1
US Military 10	197	314.333	35.532	B1
US Military 11	198	316.095	36.79	B1
Arab Male 01	242	317.353	36.534	B1
Arab Male 03	241	319.114	36.793	B1
Insurgent 01	258	320.627	37.545	B1
US Military 12	199	322.389	37.045	B1
Arab Female 03	240	323.907	38.297	B1
US Military 02	201	326.175	35.528	B1
US Military 05	204	327.432	37.292	B1
Arab Male 04	239	329.45	36.026	B1
Arab Female 02	238	329.951	36.286	B1
US Military 03	202	331.711	37.04	B1
US Military 04	203	333.482	37.291	B1
Foreign Military 01	251	334.99	36.285	B1
Arab Male 05	237	336.248	38.041	B1
Arab Female 01	236	338.006	37.037	B1
US Military 01	200	340.031	36.018	B1
Arab Male 02	235	341.047	38.287	B1
US Military 08	207	344.079	37.264	B1
Arab Male 03	233	345.585	37.262	B1
US Military 03	213	347.848	36.506	B1
Arab Female 03	232	349.359	37.768	B1
US Military 06	205	351.632	36.753	B1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Male 04	231	352.885	37.774	B1
Foreign Military 02	252	355.651	37.52	B1
US Military 10	208	357.163	37.009	B1
Arab Female 02	230	358.927	38.024	B1
Insurgent 05	256	360.946	36.505	B1
US Military 08	218	361.453	37.517	B1
US Military 07	206	364.474	35.501	B1
Arab Male 05	229	365.224	37.015	B1
Arab Female 01	228	366.487	37.02	B1
US Military 07	217	367.999	37.017	B1
Foreign Military 04	254	370.268	36.515	B1
US Military 11	209	371.525	37.78	B1
Arab Male 02	227	373.787	36.019	B1
US Military 06	216	376.049	37.554	B1
Arab Male 01	226	377.316	36.545	B1
US Military 05	215	380.341	37.805	B1
Arab Male 03	225	382.597	37.308	B1
Insurgent 02	255	384.354	36.805	B1
US Military 10	219	386.877	36.302	B1
Arab Female 03	224	388.135	37.567	B1
Arab Male 04	223	389.9	37.319	B1
US Military 04	214	393.171	36.569	B1
Arab Female 02	222	395.437	36.322	B1
Foreign Military 03	253	395.943	37.082	B1
Arab Male 05	221	398.463	37.338	B1
US Military 02	212	400.225	37.839	B1
US Military 01	211	401.985	36.582	B1
US Military 12	210	403.257	36.319	B1
Arab Male 01	234	403.761	37.322	B1
Arab Female 01	220	406.783	37.069	B1
US Military 03	272	460.0	2.481	B2
Arab Male 02	289	460.0	5.005	B2
US Military 04	273	460.0	6.518	B2
Arab Female 01	290	460.0	8.285	B2
Foreign Military 02	320	460.0	8.788	B2
Arab Male 05	291	460.0	13.078	B2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Female 02	292	460.0	13.583	B2
US Military 02	271	460.0	17.617	B2
US Military 08	288	460.0	19.386	B2
Arab Male 04	293	460.0	22.407	B2
Insurgent 06	326	460.0	23.418	B2
Arab Female 03	294	460.0	23.418	B2
Arab Male 03	295	460.0	25.427	B2
US Military 01	270	460.0	26.938	B2
US Military 12	269	460.0	27.948	B2
Foreign Military 03	321	460.0	29.460	B2
Arab Male 01	296	460.0	31.719	B2
Arab Male 02	297	460.0	32.977	B2
Arab Female 01	298	460.0	35.500	B2
US Military 11	268	460.0	34.492	B2
US Military 10	267	460.467	37.547	B2
Arab Male 05	299	463.236	37.534	B2
US Military 08	266	464.252	36.268	B2
Arab Female 02	300	466.771	37.026	B2
Arab Male 03	315	468.032	37.025	B2
US Military 07	265	470.052	37.023	B2
Insurgent 08	327	471.566	37.518	B2
US Military 06	264	473.583	37.765	B2
Arab Male 04	301	475.347	37.258	B2
US Military 05	263	477.874	36.995	B2
US Military 04	262	479.888	35.991	B2
Arab Female 03	302	481.399	37.5	B2
US Military 03	261	484.419	37.753	B2
Arab Male 03	303	487.439	37.751	B2
Insurgent 07	328	490.466	36.982	B2
US Military 02	260	491.969	37.994	B2
US Military 01	259	493.731	35.224	B2
Arab Male 04	304	495.75	36.216	B2
US Military 05	274	498.77	38.215	B2
Arab Male 02	305	500.27	35.962	B2
Insurgent 03	325	502.033	37.46	B2
US Military 06	275	505.308	37.457	B2
US Military 07	276	507.576	37.462	B2
Arab Female 01	306	510.346	37.715	B2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Male 05	307	513.108	35.705	B2
US Military 08	277	514.869	35.954	B2
Arab Female 02	308	516.629	36.956	B2
Foreign Military 04	322	517.891	36.451	B2
US Military 10	278	519.653	35.189	B2
Arab Male 04	309	521.162	37.459	B2
US Military 11	279	523.932	36.45	B2
Arab Female 03	310	526.948	34.944	B2
Insurgent 04	324	529.713	36.221	B2
US Military 12	280	530.213	37.992	B2
US Military 01	281	532.975	36.489	B2
Arab Male 03	311	534.228	37.004	B2
US Military 02	282	536.735	36.262	B2
Arab Male 01	312	537.235	37.528	B2
US Military 03	283	539.744	37.524	B2
Arab Male 02	313	541.509	36.265	B2
Arab Female 01	314	543.269	36.765	B2
US Military 04	284	544.529	37.263	B2
Foreign Military 05	323	547.304	37.019	B2
Arab Female 02	316	549.314	36.267	B2
US Military 05	285	550.323	36.519	B2
Arab Male 04	317	552.082	37.026	B2
US Military 06	286	553.585	37.292	B2
US Military 07	287	555.854	36.277	B2
Arab Male 03	319	556.861	37.782	B2
Arab Female 03	318	559.877	35.775	B2
US Military 10	359	610.299	6.052	C1
Arab Male 01	374	610.299	8.061	C1
US Military 06	66	610.299	8.315	C1
Arab Female 03	170	610.299	9.067	C1
Arab Male 03	372	610.299	9.067	C1
US Military 11	96	610.299	11.583	C1
US Military 02	352	610.299	11.332	C1
Foreign Military 05	394	610.299	13.338	C1
US Military 05	60	610.299	13.843	C1
US Military 07	357	610.299	16.097	C1
Arab Female 01	113	610.299	15.093	C1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 08	358	610.299	17.101	C1
Arab Female 03	373	610.299	17.101	C1
Arab Male 05	114	610.299	17.351	C1
Arab Male 03	172	610.299	20.617	C1
Foreign Military 02	175	610.299	19.111	C1
US Military 06	356	610.299	21.122	C1
US Military 12	97	610.299	23.125	C1
US Military 08	75	610.299	22.626	C1
Arab Male 04	371	610.299	24.635	C1
US Military 07	73	610.299	24.635	C1
Arab Female 03	388	610.299	26.648	C1
US Military 07	105	610.299	27.152	C1
Insurgent 05	469	610.299	26.143	C1
Arab Female 02	370	610.299	26.898	C1
US Military 01	351	610.299	28.404	C1
Arab Male 04	118	610.299	29.657	C1
Arab Male 05	369	610.299	31.165	C1
Arab Male 01	382	610.299	31.415	C1
Insurgent 07	180	610.299	32.67	C1
Arab Female 02	117	610.299	31.665	C1
US Military 06	104	610.299	35.182	C1
US Military 10	348	610.299	34.181	C1
Arab Female 01	368	610.299	35.182	C1
Arab Female 03	120	610.299	36.188	C1
US Military 08	109	610.299	37.445	C1
US Military 03	51	611.306	35.684	C1
Arab Male 04	167	611.557	36.943	C1
Arab Male 01	123	612.566	36.186	C1
Insurgent 03	182	613.321	37.448	C1
US Military 05	344	614.081	36.184	C1
Arab Female 02	165	614.331	36.942	C1
US Military 01	99	615.341	37.69	C1
US Military 04	343	616.601	37.937	C1
Arab Male 02	367	617.355	35.676	C1
Arab Male 01	366	618.614	37.434	C1
Insurgent 02	395	619.116	36.177	C1
Arab Male 04	387	620.878	36.93	C1
Arab Male 03	122	621.13	35.673	C1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 02	341	622.132	37.181	C1
US Military 05	101	623.387	38.19	C1
Foreign Military 03	392	624.642	36.43	C1
Insurgent 05	396	625.642	37.948	C1
Arab Male 02	125	626.896	36.948	C1
Arab Male 05	163	627.4	36.95	C1
Arab Female 03	364	628.407	37.452	C1
Arab Female 01	162	628.908	37.711	C1
US Military 12	339	630.413	38.219	C1
US Military 04	59	631.421	36.449	C1
US Military 10	337	631.671	37.211	C1
US Military 03	331	633.179	37.216	C1
Insurgent 08	184	634.683	37.72	C1
US Military 07	335	635.689	37.722	C1
US Military 02	110	636.695	37.718	C1
Arab Female 02	362	637.951	36.212	C1
US Military 05	333	638.953	38.23	C1
Arab Female 01	127	639.706	36.972	C1
Arab Male 02	161	640.46	37.474	C1
US Military 10	76	641.214	35.969	C1
Foreign Military 03	176	641.964	37.225	C1
Arab Female 01	360	643.475	37.478	C1
Arab Male 05	130	643.475	36.721	C1
US Military 01	329	644.73	37.231	C1
US Military 08	93	645.731	36.988	C1
Insurgent 09	397	646.99	36.733	C1
Arab Male 01	158	647.744	38.25	C1
Arab Female 02	378	649.503	36.491	C1
US Military 10	95	650.009	37.75	C1
US Military 06	334	651.525	36.234	C1
Foreign Military 02	391	653.282	37.752	C1
Arab Male 05	361	653.532	36.752	C1
Arab Male 03	389	655.293	37.503	C1
Arab Male 04	363	655.545	35.489	C1
Arab Male 01	138	656.803	36.744	C1
US Military 04	332	658.06	37.51	C1
Insurgent 04	187	658.81	36.76	C1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Male 04	151	659.313	38.019	C1
US Military 02	83	661.072	35.251	C1
US Military 01	82	661.828	36.509	C1
US Military 08	336	662.331	38.277	C1
Arab Male 03	365	664.097	37.513	C1
Arab Male 05	376	664.097	35.756	C1
Foreign Military 04	393	665.351	36.509	C1
US Military 02	330	666.363	38.01	C1
Arab Female 03	137	667.369	35.998	C1
Arab Male 02	375	668.127	38.264	C1
US Military 11	338	668.377	36.248	C1
Foreign Military 01	179	670.645	38.007	C1
US Military 01	340	670.645	36.755	C1
US Military 07	92	672.403	37.503	C1
US Military 02	50	673.411	35.491	C1
Arab Female 03	155	673.661	38.251	C1
Arab Male 03	157	675.167	37.508	C1
US Military 11	79	676.172	36.247	C1
US Military 06	90	677.183	37.252	C1
US Military 03	342	678.436	38.272	C1
Foreign Military 01	390	679.44	36.256	C1
Arab Female 01	377	680.449	38.268	C1
Arab Male 03	135	681.204	35.754	C1
US Military 07	346	681.711	37.261	C1
Arab Male 04	379	682.969	35.248	C1
Arab Female 02	147	683.473	37.759	C1
US Military 12	80	683.723	36.004	C1
Insurgent 06	185	684.985	36.247	C1
US Military 08	347	685.74	38.524	C1
Arab Female 01	146	687.25	38.018	C1
US Military 11	349	687.25	36.009	C1
US Military 01	49	689.017	36.753	C1
Arab Female 03	380	689.777	38.009	C1
Arab Male 03	381	691.034	35.994	C1
Insurgent 01	398	692.043	37.754	C1
Arab Male 04	133	692.796	37.001	C1
Arab Female 02	131	694.047	35.75	C1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 12	350	694.306	37.751	C1
Arab Male 05	145	696.07	38.261	C1
US Military 12	46	696.07	35.987	C1
US Military 03	353	697.332	37.251	C1
Arab Male 02	383	698.087	36.997	C1
US Military 05	89	698.849	37.743	C1
Foreign Military 05	178	701.108	37.752	C1
US Military 06	345	702.112	36.498	C1
Arab Female 01	384	702.864	38	C1
US Military 04	354	703.87	35.744	C1
US Military 03	84	704.122	37.498	C1
US Military 11	45	705.385	36.235	C1
Arab Male 02	142	705.637	37.988	C1
Arab Male 01	140	707.15	36.475	C1
US Military 05	355	707.652	35.722	C1
US Military 04	86	708.652	37.488	C1
Arab Female 02	386	710.156	34.725	C1
Arab Male 05	385	710.406	36.997	C1
US Military 01	399	762.0	2.795	C2
US Military 11	498	762.0	6.308	C2
Arab Male 01	515	762.0	5.558	C2
Arab Male 02	500	762.0	9.064	C2
US Military 02	400	762.0	7.561	C2
Arab Male 04	457	762.0	12.603	C2
Arab Female 03	458	762.0	10.070	C2
Arab Female 02	456	762.0	13.862	C2
Arab Male 02	516	762.0	12.094	C2
US Military 02	401	762.0	15.119	C2
Insurgent 05	538	762.0	13.612	C2
Arab Female 01	517	762.0	14.616	C2
US Military 12	499	762.0	17.133	C2
US Military 10	481	762.0	16.630	C2
Insurgent 08	467	762.0	18.643	C2
US Military 04	402	762.0	17.635	C2
US Military 05	403	762.0	20.889	C2
Arab Male 05	518	762.0	20.147	C2
Arab Female 01	502	762.0	21.654	C2
US Military 06	404	762.0	23.918	C2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Male 03	459	762.0	22.410	C2
US Military 11	482	762.0	23.412	C2
Foreign Military 01	460	762.0	25.429	C2
Arab Female 02	519	762.0	25.932	C2
Arab Male 05	503	762.0	27.187	C2
US Military 07	405	762.0	26.436	C2
Arab Male 04	520	762.0	28.696	C2
Arab Female 03	501	762.0	30.959	C2
Foreign Military 03	533	762.0	29.703	C2
US Military 10	407	762.0	30.709	C2
US Military 08	406	762.0	33.717	C2
Arab Female 03	455	762.0	33.217	C2
Arab Male 03	522	762.0	33.717	C2
US Military 10	497	762.0	36.738	C2
US Military 12	483	762.0	35.988	C2
Arab Female 01	454	763.031	37.728	C2
US Military 11	408	763.785	36.216	C2
Arab Male 04	514	764.541	38.48	C2
US Military 01	484	765.797	36.223	C2
US Military 08	496	766.302	37.22	C2
Arab Male 01	452	767.558	36.465	C2
Foreign Military 05	463	768.308	37.727	C2
US Military 07	479	769.561	35.72	C2
US Military 12	409	770.312	37.228	C2
Arab Female 03	521	771.314	36.476	C2
Arab Male 02	453	771.818	38.236	C2
Arab Male 03	451	772.831	36.721	C2
US Military 01	410	773.083	37.725	C2
US Military 08	480	774.603	35.952	C2
Insurgent 09	537	774.853	37.72	C2
US Military 02	411	775.862	36.965	C2
Arab Male 04	449	776.112	37.217	C2
US Military 07	495	777.619	37.473	C2
Arab Male 01	523	777.619	35.96	C2
Arab Male 03	506	779.133	37.975	C2
Arab Female 03	450	779.383	36.717	C2
Foreign Military	532	780.893	37.726	C2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
04				
US Military 06	478	781.897	36.976	C2
US Military 03	412	782.649	38.237	C2
Foreign Military 05	534	783.654	36.731	C2
Arab Male 01	507	784.66	37.489	C2
US Military 05	414	785.412	38.502	C2
Arab Female 02	448	785.412	36.987	C2
Arab Male 02	524	786.926	35.977	C2
Arab Male 05	447	787.179	38.253	C2
US Military 04	413	788.687	36.745	C2
Arab Female 01	446	789.187	38.757	C2
Insurgent 03	465	790.95	36.743	C2
Arab Female 02	504	792.206	38.247	C2
Arab Male 01	444	793.209	37.494	C2
US Military 02	485	793.962	38.252	C2
Arab Male 02	445	794.965	35.989	C2
US Military 06	415	795.467	37.749	C2
US Military 05	477	796.478	36.738	C2
Foreign Military 02	531	796.982	37.75	C2
US Military 04	476	798.488	36.244	C2
US Military 12	494	798.992	37.245	C2
Arab Male 03	443	799.749	36.238	C2
Arab Male 04	505	800.759	37.992	C2
US Military 07	416	801.516	36.985	C2
US Military 11	493	802.27	37.233	C2
US Military 08	417	803.522	38.244	C2
Insurgent 06	466	804.523	36.993	C2
Arab Female 03	442	805.784	37.233	C2
Arab Female 02	511	806.787	37.737	C2
US Military 10	418	808.294	36.982	C2
Arab Male 04	441	808.544	38.247	C2
US Military 03	475	810.555	36.236	C2
US Military 12	420	811.313	37.491	C2
Foreign Military 03	462	812.322	37.24	C2
US Military 11	419	812.827	36.481	C2
Arab Female 03	513	814.341	37.742	C2
Arab Female 02	440	815.346	36.237	C2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Insurgent 01	536	816.1	37.74	C2
Arab Male 05	439	817.613	36.73	C2
US Military 10	492	817.613	35.977	C2
Arab Male 04	512	819.123	37.481	C2
US Military 01	421	819.123	35.72	C2
US Military 08	491	820.636	36.731	C2
Insurgent 07	468	820.636	36.731	C2
Arab Female 03	529	822.399	36.472	C2
Arab Female 01	438	823.158	37.471	C2
Arab Female 02	527	824.165	35.708	C2
US Military 02	422	825.179	37.709	C2
US Military 04	424	825.682	36.7	C2
Arab Male 01	436	826.687	36.456	C2
Foreign Military 01	535	827.693	36.201	C2
US Military 07	490	828.194	37.71	C2
US Military 03	423	829.448	36.205	C2
Arab Male 03	435	830.703	37.466	C2
Arab Male 04	528	830.703	34.95	C2
Insurgent 04	464	832.214	37.968	C2
US Military 01	473	832.464	35.705	C2
Arab Male 05	510	833.723	36.711	C2
US Military 02	474	834.227	36.462	C2
Arab Male 02	437	835.737	35.708	C2
Arab Female 03	430	835.737	37.975	C2
Arab Female 01	525	837.241	36.471	C2
US Military 06	489	837.241	37.725	C2
Insurgent 02	539	839.253	35.967	C2
US Military 05	425	839.253	36.47	C2
Arab Male 04	434	841.011	36.986	C2
Arab Female 01	509	841.262	36.988	C2
US Military 12	472	842.517	36.488	C2
Foreign Military 01	530	843.269	37.495	C2
US Military 08	428	844.776	36.742	C2
US Military 06	426	844.776	35.988	C2
Arab Female 03	526	846.286	36.244	C2
Arab Male 02	508	846.791	37.759	C2
Arab Male 05	432	848.05	35.494	C2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 03	486	848.301	37.506	C2
Foreign Military 04	461	849.813	36.251	C2
US Military 04	487	850.817	37.007	C2
Arab Female 01	431	851.583	36.997	C2
US Military 05	488	852.835	38.265	C2
US Military 11	471	855.093	35.251	C2
Arab Male 02	429	855.343	37.265	C2
US Military 10	470	857.111	36.754	C2
Arab Female 02	433	857.869	37.513	C2
US Military 07	427	858.621	36.255	C2

Table 52: Threat Detection Event Log, Constant Event Rate

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
None	0	0	0	A1
US Military 07	147	3.137	6.191	A1
Arab Female 01	179	3.137	9.960	A1
US Military 10	149	3.137	9.960	A1
US Military 08	148	3.137	14.748	A1
Arab Male 02	178	3.137	15.505	A1
Foreign Military 04	184	3.137	18.267	A1
Arab Female 02	176	3.137	19.774	A1
Arab Male 01	175	3.137	24.311	A1
US Military 04	144	3.137	25.072	A1
Insurgent 01	185	3.137	29.097	A1
Arab Female 03	177	3.137	28.592	A1
Arab Male 05	174	3.137	31.858	A1
US Military 05	145	3.137	33.366	A1
US Military 06	146	3.137	34.623	A1
Arab Female 02	173	4.145	38.136	A1
US Military 03	143	6.668	36.629	A1
Arab Female 01	172	9.69	38.376	A1
US Military 02	142	10.698	35.607	A1
Arab Male 04	171	13.23	37.097	A1
US Military 01	141	14.991	37.099	A1
Foreign Military 03	183	17.013	37.335	A1
Arab Male 02	168	18.023	37.585	A1
US Military 12	140	19.287	36.066	A1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 11	139	21.563	37.064	A1
US Military 10	138	25.103	37.066	A1
Arab Female 02	170	26.87	36.554	A1
Insurgent 08	188	28.393	37.065	A1
US Military 08	137	28.9	36.304	A1
Arab Male 03	169	31.94	38.07	A1
US Military 07	136	34.219	36.795	A1
Arab Female 03	167	35.732	37.797	A1
US Military 06	135	38.258	36.781	A1
US Military 05	134	40.758	36.297	A1
US Military 04	133	42.281	36.291	A1
Arab Male 01	166	43.046	35.778	A1
Foreign Military 05	180	45.053	36.297	A1
Arab Female 01	165	46.305	37.075	A1
US Military 01	119	47.56	35.56	A1
Arab Female 03	164	49.071	35.821	A1
Arab Male 03	162	50.83	35.833	A1
US Military 02	120	50.83	35.577	A1
Arab Male 05	150	52.592	37.368	A1
Insurgent 07	187	53.843	36.877	A1
US Military 03	121	55.353	35.367	A1
Arab Male 04	151	56.359	35.881	A1
Arab Male 01	163	57.118	36.906	A1
US Military 04	122	60.145	36.405	A1
Arab Male 03	152	61.156	37.402	A1
Arab Female 02	161	63.424	37.896	A1
US Military 05	123	65.458	37.364	A1
Arab Female 01	154	68.993	36.621	A1
US Military 06	124	71.272	37.37	A1
Foreign Military 01	181	72.524	36.623	A1
Arab Female 02	155	74.789	36.874	A1
US Military 07	125	76.039	37.902	A1
US Military 08	126	79.328	37.136	A1
Arab Male 02	153	80.835	36.892	A1
US Military 10	127	82.111	37.134	A1
Arab Male 05	158	83.63	36.879	A1
US Military 11	128	87.173	37.11	A1
Foreign Military 02	182	89.96	36.584	A1
US Military 12	129	92.498	37.587	A1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Female 02	160	94.526	36.57	A1
Arab Female 01	159	96.55	36.064	A1
Insurgent 03	186	97.302	37.079	A1
US Military 01	130	99.308	37.101	A1
Arab Male 02	156	101.32	37.104	A1
US Military 02	131	102.07	37.112	A1
Arab Female 03	157	104.087	37.607	A1
US Military 03	132	105.354	37.863	A1
US Military 07	77	121.521	41.921	A1
Arab Female 02	95	122.772	42.93	A1
Arab Female 01	108	123.272	43.437	A1
US Military 05	75	157.37	11.861	A2
Insurgent 02	114	157.37	12.871	A2
US Military 03	73	157.37	15.908	A2
Arab Male 02	107	157.37	15.908	A2
US Military 01	67	157.37	19.956	A2
Arab Male 05	97	157.37	21.724	A2
US Military 11	65	157.37	23.753	A2
Foreign Military 01	110	157.37	26.783	A2
US Military 08	59	157.37	29.078	A2
Arab Male 03	84	157.37	28.818	A2
Arab Male 01	109	157.37	32.11	A2
US Military 06	57	157.37	34.872	A2
US Military 04	51	157.37	35.623	A2
Insurgent 01	115	158.628	37.654	A2
US Military 02	49	161.163	37.13	A2
Arab Female 03	96	162.681	36.62	A2
US Military 12	45	165.452	35.111	A2
Arab Male 03	99	166.209	37.645	A2
Foreign Military 02	111	168.219	36.649	A2
Arab Male 01	98	170.745	36.387	A2
US Military 10	37	171.51	37.894	A2
Arab Male 04	94	173.278	35.874	A2
Arab Male 03	92	176.065	36.89	A2
US Military 07	35	177.326	36.637	A2
Arab Male 04	93	178.587	37.396	A2
US Military 06	33	180.616	36.634	A2
Arab Male 02	83	181.373	37.142	A2
Arab Female 03	81	183.397	35.368	A2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 05	116	185.177	36.351	A2
US Military 04	8	185.937	37.623	A2
Arab Female 01	79	188.213	36.353	A2
US Military 02	6	190.234	38.115	A2
Arab Male 02	106	192.242	36.872	A2
US Military 01	5	193.753	37.124	A2
Arab Male 01	102	198.293	38.373	A2
US Military 08	78	200.814	37.113	A2
US Military 03	7	203.098	37.357	A2
Insurgent 09	118	204.105	37.357	A2
Arab Female 02	80	206.126	37.603	A2
Arab Male 04	105	207.633	37.861	A2
US Military 05	31	209.655	36.342	A2
Arab Male 01	82	211.43	36.083	A2
US Military 08	36	213.205	37.335	A2
Foreign Military 05	113	215.733	37.323	A2
US Military 11	38	217.758	35.805	A2
Arab Male 04	85	218.765	37.568	A2
Arab Male 04	100	222.031	36.562	A2
Insurgent 08	117	224.315	36.548	A2
US Military 01	46	225.073	38.547	A2
Arab Female 03	104	227.343	36.779	A2
US Military 03	50	229.114	36.278	A2
US Military 05	52	230.123	37.545	A2
Arab Female 02	88	232.137	37.547	A2
US Military 07	58	234.649	34.53	A2
Arab Male 05	86	235.66	36.292	A2
Arab Female 02	101	237.668	37.31	A2
US Military 10	60	239.692	36.542	A2
Foreign Military 03	112	240.707	37.786	A2
Arab Male 05	103	242.219	37.289	A2
US Military 12	66	243.229	37.297	A2
Arab Female 03	89	245.744	37.813	A2
Arab Female 01	87	245.997	37.56	A2
US Military 02	68	249.276	37.295	A2
Arab Male 01	90	251.798	34.773	A2
US Military 06	76	253.813	35.77	A2
Arab Male 02	91	255.075	36.78	A2
US Military 04	74	256.584	37.044	A2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 01	189	302.458	11.679	B1
US Military 02	190	302.458	17.742	B1
Arab Male 04	247	302.458	18.503	B1
US Military 03	191	302.458	22.541	B1
Arab Female 03	248	302.458	23.301	B1
Foreign Military 04	250	302.458	25.567	B1
US Military 04	192	302.458	25.818	B1
Arab Male 03	249	302.458	28.097	B1
US Military 05	193	302.458	29.358	B1
Arab Female 02	246	302.458	33.375	B1
US Military 06	194	302.458	34.389	B1
Insurgent 09	257	302.458	36.151	B1
Arab Male 05	245	303.972	37.156	B1
US Military 07	195	306.052	36.333	B1
US Military 08	196	309.078	38.103	B1
Arab Female 01	244	309.078	36.584	B1
Arab Male 02	243	312.88	37.323	B1
US Military 10	197	314.388	36.068	B1
US Military 11	198	316.165	36.809	B1
Arab Male 01	242	317.434	37.044	B1
Arab Male 03	241	318.939	37.3	B1
Insurgent 01	258	320.705	37.8	B1
US Military 12	199	322.475	36.79	B1
Arab Female 03	240	323.986	37.813	B1
US Military 02	201	326.01	35.789	B1
US Military 05	204	327.525	36.286	B1
Arab Female 02	238	329.543	37.295	B1
Arab Male 04	239	329.543	36.038	B1
US Military 03	202	331.816	37.043	B1
US Military 04	203	333.825	36.803	B1
Foreign Military 01	251	334.578	36.304	B1
Arab Male 05	237	336.589	37.818	B1
Arab Female 01	236	338.109	36.798	B1
US Military 01	200	339.871	36.794	B1
Arab Male 02	235	341.379	37.816	B1
US Military 08	207	344.153	37.575	B1
Arab Male 03	233	345.662	37.328	B1
US Military 03	213	348.694	36.068	B1
Arab Female 03	232	349.448	37.845	B1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 06	205	351.461	36.839	B1
Arab Male 04	231	353.474	37.857	B1
Foreign Military 02	252	355.734	37.367	B1
US Military 10	208	357.499	37.115	B1
Arab Female 02	230	359.009	37.874	B1
Insurgent 05	256	361.29	37.118	B1
US Military 08	218	361.549	37.361	B1
US Military 07	206	363.811	35.603	B1
Arab Male 05	229	365.331	36.857	B1
Arab Female 01	228	366.583	37.112	B1
US Military 07	217	368.102	37.12	B1
Foreign Military 04	254	370.121	36.603	B1
US Military 11	209	371.639	37.612	B1
Arab Male 02	227	373.647	36.604	B1
US Military 06	216	376.159	37.359	B1
Arab Male 01	226	377.173	36.595	B1
US Military 05	215	380.462	37.598	B1
Arab Male 03	225	382.74	37.334	B1
Insurgent 02	255	384.504	37.083	B1
US Military 10	219	386.788	36.805	B1
Arab Female 03	224	388.3	37.557	B1
Arab Male 04	223	390.068	37.303	B1
US Military 04	214	393.101	36.79	B1
Arab Female 02	222	395.371	36.541	B1
Foreign Military 03	253	396.129	37.043	B1
Arab Male 05	221	398.658	37.287	B1
US Military 02	212	400.421	37.802	B1
US Military 01	211	401.935	36.788	B1
Arab Male 01	234	403.951	36.789	B1
US Military 12	210	404.967	36.274	B1
Arab Female 01	220	406.974	37.041	B1
US Military 03	272	456.6	6.053	B2
Arab Male 02	289	456.6	8.563	B2
US Military 04	273	456.6	10.071	B2
Arab Female 01	290	456.6	11.843	B2
Foreign Military 02	320	456.6	13.1	B2
Arab Male 05	291	456.6	15.873	B2
Arab Female 02	292	456.6	18.161	B2
US Military 02	271	456.6	21.175	B2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 08	288	456.6	22.95	B2
Arab Male 04	293	456.6	26.217	B2
Insurgent 06	326	456.6	27.486	B2
Arab Female 03	294	456.6	27.986	B2
Arab Male 03	295	456.6	28.988	B2
US Military 01	270	456.6	30.488	B2
US Military 12	269	456.6	31.488	B2
Foreign Military 03	321	456.6	33.255	B2
Arab Male 01	296	456.6	35.534	B2
Arab Male 02	297	456.6	36.545	B2
Arab Female 01	298	457.358	38.304	B2
US Military 11	268	458.875	35.53	B2
US Military 10	267	460.637	37.542	B2
Arab Male 05	299	463.154	37.791	B2
US Military 08	266	464.657	36.032	B2
Arab Female 02	300	466.922	37.032	B2
Arab Male 03	315	468.19	37.02	B2
US Military 07	265	470.207	37.021	B2
Insurgent 08	327	471.721	37.525	B2
US Military 06	264	473.745	37.768	B2
Arab Male 04	301	475.766	36.499	B2
US Military 05	263	478.282	36.756	B2
US Military 04	262	480.05	36.516	B2
Arab Female 03	302	481.81	37.784	B2
US Military 03	261	484.586	38.275	B2
Arab Male 03	303	487.838	37.539	B2
Insurgent 07	328	490.61	36.785	B2
US Military 02	260	492.134	38.302	B2
US Military 01	259	493.645	35.521	B2
Arab Male 04	304	495.662	36.539	B2
US Military 05	274	498.683	38.302	B2
Arab Male 02	305	500.689	35.541	B2
US Military 06	275	502.203	37.566	B2
Arab Female 03	318	503.954	35.815	B2
Insurgent 03	325	505.462	37.329	B2
US Military 07	276	507.478	37.58	B2
Arab Female 01	306	510.498	37.84	B2
Arab Male 05	307	513.27	35.823	B2
US Military 08	277	515.038	36.32	B2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Female 02	308	516.566	37.066	B2
Foreign Military 04	322	517.828	36.817	B2
US Military 10	278	519.849	35.049	B2
Arab Male 04	309	521.351	37.339	B2
US Military 11	279	524.118	36.598	B2
Arab Male 03	319	525.881	37.104	B2
Arab Female 03	310	527.653	35.582	B2
Insurgent 04	324	529.919	36.098	B2
US Military 12	280	530.176	38.356	B2
US Military 01	281	533.217	36.573	B2
Arab Male 03	311	534.217	37.084	B2
US Military 02	282	536.985	36.093	B2
Arab Male 01	312	537.493	37.355	B2
US Military 03	283	540.019	37.33	B2
Arab Male 02	313	541.779	36.836	B2
Arab Female 01	314	543.546	37.099	B2
US Military 04	284	544.551	37.366	B2
Foreign Military 05	323	547.582	36.85	B2
Arab Female 02	316	549.595	36.094	B2
US Military 05	285	550.602	36.351	B2
Arab Male 04	317	552.115	37.36	B2
US Military 06	286	553.632	37.366	B2
US Military 07	287	556.923	36.095	B2
US Military 10	359	610.579	5.797	C1
Arab Male 01	374	610.579	8.825	C1
Arab Male 03	372	610.579	10.832	C1
US Military 08	358	610.579	15.362	C1
Arab Female 03	373	610.579	16.126	C1
US Military 07	357	610.579	17.133	C1
US Military 06	356	610.579	19.151	C1
Foreign Military 05	394	610.579	21.923	C1
US Military 02	352	610.579	22.682	C1
Arab Male 04	371	610.579	24.685	C1
Arab Female 03	388	610.579	26.213	C1
Arab Female 02	370	610.579	27.985	C1
US Military 01	351	610.579	29.988	C1
Arab Male 01	382	610.579	30.499	C1
Arab Male 05	369	610.579	32.521	C1
Insurgent 02	395	610.579	35.299	C1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 10	348	610.579	35.048	C1
Arab Female 01	368	612.088	38.078	C1
US Military 05	344	613.098	35.557	C1
US Military 04	343	615.876	36.81	C1
Arab Male 02	367	617.638	35.812	C1
Arab Male 04	387	619.151	37.818	C1
US Military 02	341	621.916	36.812	C1
Arab Male 01	366	623.418	36.325	C1
Foreign Military 03	392	625.435	37.071	C1
US Military 12	339	626.193	37.326	C1
Arab Female 03	364	628.47	37.573	C1
US Military 10	337	630.491	37.575	C1
Insurgent 05	396	633.512	38.089	C1
US Military 07	335	634.512	37.339	C1
Arab Female 02	362	637.047	37.312	C1
US Military 05	333	639.314	37.825	C1
US Military 03	331	642.338	37.569	C1
Arab Female 01	360	644.364	37.31	C1
US Military 01	329	647.142	36.558	C1
Foreign Military 02	391	647.898	37.31	C1
Arab Female 02	378	649.662	36.55	C1
US Military 02	330	650.673	37.815	C1
Arab Male 05	361	652.185	37.309	C1
US Military 04	332	654.452	37.299	C1
Arab Male 04	363	656.211	36.799	C1
Arab Male 03	389	658.228	37.811	C1
US Military 06	334	659.488	37.056	C1
US Military 08	336	661.754	37.316	C1
Foreign Military 04	393	662.256	36.814	C1
Arab Male 05	376	664.527	37.561	C1
US Military 11	338	665.538	36.3	C1
Arab Male 03	365	667.814	35.781	C1
Arab Male 02	375	669.08	38.293	C1
US Military 01	340	670.595	37.03	C1
Insurgent 09	397	673.854	36.031	C1
US Military 03	342	674.359	37.789	C1
US Military 06	345	677.641	36.01	C1
Arab Female 01	377	679.405	38.536	C1
US Military 07	346	682.18	37.28	C1

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Male 04	379	682.945	36.515	C1
US Military 08	347	685.459	38.539	C1
US Military 11	349	686.212	36.014	C1
Insurgent 01	398	689.998	36.012	C1
Arab Female 03	380	690.998	38.29	C1
Arab Male 03	381	693.26	36.278	C1
US Military 12	350	694.513	37.281	C1
US Military 03	353	697.304	37.762	C1
Arab Male 02	383	698.57	36.753	C1
Foreign Military 01	390	700.079	38.017	C1
US Military 04	354	702.088	36.508	C1
Arab Female 01	384	704.6	37.286	C1
US Military 05	355	706.365	35.778	C1
Arab Female 02	386	708.882	35.27	C1
Arab Male 05	385	710.393	36.801	C1
US Military 01	399	764.086	1.009	C2
US Military 02	400	764.086	5.787	C2
Arab Male 04	457	764.086	10.578	C2
Arab Female 03	458	764.086	11.079	C2
US Military 02	401	764.086	13.839	C2
Foreign Military 01	460	764.086	13.839	C2
US Military 04	402	764.086	15.617	C2
US Military 05	403	764.086	18.88	C2
Arab Male 03	459	764.086	20.39	C2
US Military 06	404	764.086	22.911	C2
Insurgent 08	467	764.086	24.422	C2
US Military 07	405	764.086	26.193	C2
Arab Female 02	456	764.086	28.467	C2
US Military 08	406	764.086	30.735	C2
Arab Female 03	455	764.086	31.237	C2
US Military 10	407	764.086	33.513	C2
Arab Female 01	454	764.086	38.053	C2
US Military 11	408	764.086	36.547	C2
Arab Male 02	453	766.608	38.043	C2
Arab Male 01	452	767.863	36.288	C2
Foreign Military 04	461	769.623	37.295	C2
US Military 12	409	772.377	37.062	C2
Arab Male 03	451	773.147	36.545	C2
US Military 01	410	774.915	37.795	C2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
Arab Female 03	450	776.417	35.791	C2
US Military 02	411	778.936	36.787	C2
Arab Male 04	449	779.953	38.285	C2
Insurgent 06	466	782.216	37.032	C2
US Military 03	412	783.216	37.534	C2
Arab Female 02	448	784.727	36.53	C2
Arab Male 05	447	786.741	38.294	C2
US Military 04	413	788.765	36.772	C2
Arab Female 01	446	790.028	38.279	C2
US Military 05	414	792.301	38.287	C2
Arab Male 02	445	793.563	36.519	C2
Arab Male 01	444	795.578	37.272	C2
US Military 06	415	796.84	37.531	C2
Arab Male 03	443	799.369	37.017	C2
Insurgent 03	465	800.38	37.777	C2
US Military 07	416	802.389	36.521	C2
US Military 08	417	803.396	38.281	C2
Arab Female 03	442	805.907	37.036	C2
US Military 10	418	808.427	36.28	C2
Arab Male 04	441	808.677	38.291	C2
Foreign Military 03	462	812.71	37.025	C2
US Military 11	419	813.461	36.524	C2
Arab Female 02	440	816.224	36.276	C2
US Military 12	420	817.235	38.025	C2
Arab Male 05	439	818.994	37.017	C2
Insurgent 07	468	820.5	37.527	C2
US Military 01	421	822.007	36.02	C2
Arab Female 01	438	823.264	38.298	C2
US Military 02	422	825.035	37.789	C2
Arab Male 01	436	827.802	36.785	C2
US Military 03	423	828.815	37.029	C2
Arab Male 03	435	830.336	38.028	C2
Insurgent 04	464	832.35	36.014	C2
US Military 04	424	834.621	36.26	C2
Arab Female 03	430	836.136	37.765	C2
Arab Male 02	437	837.14	36.761	C2
US Military 05	425	838.41	38.012	C2
Arab Male 04	434	841.168	37.28	C2
Arab Female 02	433	843.698	37.271	C2

Object Type	Object Name	First Time Seen (min.second)	Visible Time Duration (seconds)	Segment
US Military 06	426	844.957	36.518	C2
US Military 08	428	846.214	38.295	C2
Arab Male 05	432	848.226	35.779	C2
Foreign Military 05	463	850.243	37.299	C2
Arab Female 01	431	853.25	37.065	C2
Arab Male 02	429	855.761	37.82	C2
US Military 07	427	858.533	36.049	C2

APPENDIX D:
ANALYSIS OF VARIANCE, INVOKING EXERIMENTS, SCENARIO 2

```
One-way ANOVA: S2A WL versus S2A InvTh

Source      DF       SS      MS       F      P
S2A InvTh    6  58.0166  9.6694  142.90  0.000
Error       63   4.2629  0.0677
Total       69  62.2795

S = 0.2601   R-Sq = 93.16%   R-Sq(adj) = 92.50%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N    Mean  StDev  -------+---------+---------+---------+--
10     10  10.883  0.160  (-*--)
20     10  10.937  0.155   (-*--)
23     10  11.567  0.302            (-*--)
24     10  11.580  0.284            (-*--)
25     10  12.972  0.267                                (-*--)
30     10  13.020  0.306                                 (-*-)
40     10  13.037  0.295                                 (-*--)
                          -------+---------+---------+---------+--
                             11.20     11.90     12.60     13.30

Pooled StDev = 0.260
```

Figure 71: ANOVA: Invoking Threshold vs. Workload, Scenario 2, Segment A

```
One-way ANOVA: S2B WL versus S2B InvTh

Source      DF       SS       MS       F      P
S2B InvTh    6  71.8403  11.9734  139.86  0.000
Error       63   5.3934   0.0856
Total       69  77.2337

S = 0.2926   R-Sq = 93.02%   R-Sq(adj) = 92.35%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N    Mean  StDev  -------+---------+---------+---------+--
10     10  11.591  0.292  (--*-)
20     10  11.656  0.267   (--*-)
23     10  11.703  0.322    (-*--)
24     10  11.739  0.354    (--*-)
25     10  13.717  0.287                              (--*--)
30     10  13.745  0.281                               (-*--)
40     10  13.690  0.228                              (--*-)
                          -------+---------+---------+---------+--
                             11.90     12.60     13.30     14.00

Pooled StDev = 0.293
```

Figure 72: ANOVA: Invoking Threshold vs. Workload, Scenario 2, Segment B

```
One-way ANOVA: S2C WL versus S2C InvTh

Source      DF       SS       MS       F      P
S2C InvTh    6  65.8953  10.9826  141.44  0.000
Error       63   4.8918   0.0776
Total       69  70.7871

S = 0.2787   R-Sq = 93.09%   R-Sq(adj) = 92.43%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N    Mean  StDev  --------+---------+---------+---------+-
10     10  11.481  0.256  (-*--)
20     10  11.656  0.233    (--*-)
23     10  12.159  0.300          (--*-)
24     10  11.995  0.234        (-*--)
25     10  13.679  0.272                              (-*--)
30     10  13.775  0.326                               (--*-)
40     10  13.764  0.315                               (--*-)
                          --------+---------+---------+---------+-
                              11.90     12.60     13.30     14.00

Pooled StDev = 0.279
```

Figure 73: ANOVA: Invoking Threshold vs. Workload, Scenario 2, Segment C

```
One-way ANOVA: S2A MM versus S2A InvTh

Source     DF       SS       MS      F      P
S2A InvTh   6  0.08927  0.01488  14.86  0.000
Error      63  0.06308  0.00100
Total      69  0.15235

S = 0.03164   R-Sq = 58.60%   R-Sq(adj) = 54.65%


                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level   N     Mean    StDev   --+---------+---------+---------+-------
10     10  0.71230  0.02298   (-----*----)
20     10  0.72018  0.02678     (-----*----)
23     10  0.73792  0.01734          (-----*-----)
24     10  0.73537  0.01887         (-----*-----)
25     10  0.79627  0.03859                           (-----*----)
30     10  0.79629  0.04277                           (-----*----)
40     10  0.79770  0.04228                           (-----*-----)
                              --+---------+---------+---------+-------
                              0.700     0.735     0.770     0.805

Pooled StDev = 0.03164
```

Figure 74: ANOVA: Invoking Threshold vs. % Time Monitoring Map, Scenario 2, Segment A

```
One-way ANOVA: S2B MM versus S2B InvTh

Source     DF        SS        MS      F      P
S2B InvTh   6  0.076114  0.012686  18.54  0.000
Error      63  0.043114  0.000684
Total      69  0.119227

S = 0.02616   R-Sq = 63.84%   R-Sq(adj) = 60.40%


                                Individual 95% CIs For Mean Based on Pooled StDev
Level   N     Mean    StDev   +---------+---------+---------+---------
10     10  0.59614  0.01604      (-----*----)
20     10  0.58772  0.02078   (-----*----)
23     10  0.60495  0.02832         (-----*----)
24     10  0.60497  0.02642         (-----*----)
25     10  0.65728  0.03317                           (----*-----)
30     10  0.66356  0.02973                             (----*-----)
40     10  0.67050  0.02485                               (-----*----)
                              +---------+---------+---------+---------
                            0.570     0.600     0.630     0.660

Pooled StDev = 0.02616
```

Figure 75: ANOVA: Invoking Threshold vs. % Time Monitoring Map, Scenario 2, Segment B

```
One-way ANOVA: S2C MM versus S2C InvTh

Source     DF        SS        MS     F      P
S2C InvTh   6  0.000830  0.000138  0.25  0.958
Error      63  0.034950  0.000555
Total      69  0.035779

S = 0.02355   R-Sq = 2.32%   R-Sq(adj) = 0.00%


                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level   N     Mean    StDev   -----+---------+---------+---------+----
10     10  0.48826  0.02477            (------------*-----------)
20     10  0.48025  0.02097      (-----------*------------)
23     10  0.47689  0.02365   (-----------*------------)
24     10  0.48007  0.01852      (-----------*-----------)
25     10  0.47997  0.01879      (-----------*-----------)
30     10  0.48009  0.02566      (-----------*-----------)
40     10  0.48436  0.03028         (------------*-----------)
                              -----+---------+---------+---------+----
                                 0.468     0.480     0.492     0.504

Pooled StDev = 0.02355
```

Figure 76: ANOVA: Invoking Threshold vs. % Time Monitoring Map, Scenario 2, Segment C

```
One-way ANOVA: S2A CI versus S2A InvTh

Source     DF       SS       MS      F      P
S2A InvTh   6  0.62026  0.10338  11.36  0.000
Error      63  0.57334  0.00910
Total      69  1.19360

S = 0.09540   R-Sq = 51.97%   R-Sq(adj) = 47.39%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N     Mean    StDev  -+---------+---------+---------+--------
10     10  0.35000  0.06477  (-----*-----)
20     10  0.35357  0.05940  (-----*-----)
23     10  0.40000  0.05533       (-----*-----)
24     10  0.40357  0.05337       (-----*-----)
25     10  0.56429  0.10754                       (-----*-----)
30     10  0.56071  0.13888                       (-----*-----)
40     10  0.56429  0.13863                       (-----*-----)
                             -+---------+---------+---------+--------
                             0.30      0.40      0.50      0.60

Pooled StDev = 0.09540
```

Figure 77: ANOVA: Invoking Threshold vs. % Changes Identified, Scenario 2, Segment A

```
One-way ANOVA: S2B CI versus S2B InvTh

Source     DF       SS       MS      F      P
S2B InvTh   6  0.37311  0.06219  16.18  0.000
Error      63  0.24216  0.00384
Total      69  0.61527

S = 0.06200   R-Sq = 60.64%   R-Sq(adj) = 56.89%


                             Individual 95% CIs For Mean Based on Pooled StDev
Level   N     Mean    StDev    +---------+---------+---------+---------
10     10  0.38929  0.05507    (-----*----)
20     10  0.42143  0.06477         (----*-----)
23     10  0.43571  0.06422           (----*-----)
24     10  0.44107  0.05193           (-----*-----)
25     10  0.56964  0.07066                             (----*-----)
30     10  0.56786  0.07083                             (----*-----)
40     10  0.56071  0.05337                            (----*-----)
                               +---------+---------+---------+---------
                             0.350     0.420     0.490     0.560

Pooled StDev = 0.06200
```

Figure 78: ANOVA: Invoking Threshold vs. % Changes Identified, Scenario 2, Segment B

```
One-way ANOVA: S2C CI versus S2C InvTh

Source     DF        SS        MS      F      P
S2C InvTh   6  0.290015  0.048336  49.87  0.000
Error      63  0.061065  0.000969
Total      69  0.351080

S = 0.03113   R-Sq = 82.61%   R-Sq(adj) = 80.95%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N     Mean    StDev  --+---------+---------+---------+-------
10     10  0.31071  0.04350  (---*---)
20     10  0.32857  0.02910     (---*---)
23     10  0.36518  0.01654            (---*---)
24     10  0.35804  0.02038           (---*---)
25     10  0.46875  0.02862                                (---*---)
30     10  0.47054  0.03421                                (---*---)
40     10  0.45893  0.03698                              (---*---)
                             --+---------+---------+---------+-------
                             0.300     0.350     0.400     0.450

Pooled StDev = 0.03113
```

Figure 79: ANOVA: Invoking Threshold vs. % Changes Identified, Scenario 2, Segment C

APPENDIX E:
ANALYSIS OF VARIANCE, INVOKING EXERIMENTS, SCENARIO 4

```
One-way ANOVA: S4A WL versus S4A InvTh

Source     DF      SS      MS      F      P
S4A InvTh   6  6.6646  1.1108  34.25  0.000
Error      63  2.0429  0.0324
Total      69  8.7075

S = 0.1801   R-Sq = 76.54%   R-Sq(adj) = 74.30%


                           Individual 95% CIs For Mean Based on
                           Pooled StDev
Level   N    Mean  StDev  --+---------+---------+---------+-------
10     10   9.810  0.240  (---*----)
20     10   9.853  0.242    (---*----)
23     10   9.863  0.175    (----*---)
24     10   9.863  0.175    (----*---)
25     10  10.467  0.132                          (----*---)
30     10  10.471  0.128                          (----*---)
40     10  10.471  0.128                          (----*---)
```

Figure 80: ANOVA: Invoking Threshold vs. Workload, Scenario 4, Segment A

```
One-way ANOVA: S4B WL versus S4B InvTh

Source     DF       SS       MS       F      P
S4B InvTh   6  82.8512  13.8085  169.18  0.000
Error      63   5.1420   0.0816
Total      69  87.9932

S = 0.2857   R-Sq = 94.16%   R-Sq(adj) = 93.60%


                           Individual 95% CIs For Mean Based on
                           Pooled StDev
Level   N    Mean  StDev  ----+---------+---------+---------+-----
10     10  11.048  0.333  (-*-)
20     10  11.286  0.286     (-*-)
23     10  11.633  0.355         (-*--)
24     10  11.594  0.309         (-*-)
25     10  13.523  0.200                                 (-*-)
30     10  13.576  0.241                                 (--*-)
40     10  13.576  0.241                                 (--*-)
                          ----+---------+---------+---------+-----
                            11.20     12.00     12.80     13.60

Pooled StDev = 0.286
```

Figure 81: ANOVA: Invoking Threshold vs. Workload, Scenario 4, Segment B

```
One-way ANOVA: S4C WL versus S4C InvTh

Source     DF       SS      MS       F      P
S4C InvTh   6  134.769  22.461  154.59  0.000
Error      63    9.154   0.145
Total      69  143.922

S = 0.3812   R-Sq = 93.64%   R-Sq(adj) = 93.03%


                           Individual 95% CIs For Mean Based on
                           Pooled StDev
Level   N    Mean  StDev  ---+---------+---------+---------+------
10     10  15.964  0.422  (--*-)
20     10  15.968  0.435  (--*-)
23     10  16.533  0.463        (-*--)
24     10  16.634  0.307         (-*--)
25     10  18.748  0.261                            (-*--)
30     10  19.164  0.370                                (--*-)
40     10  19.164  0.370                                (--*-)
                          ---+---------+---------+---------+------
                          16.0      17.0      18.0      19.0

Pooled StDev = 0.381
```

Figure 82: ANOVA: Invoking Threshold vs. Workload, Scenario 4, Segment C

```
One-way ANOVA: S4A MM versus S4A InvTh

Source     DF        SS        MS      F      P
S4A InvTh   6  0.037837  0.006306  15.39  0.000
Error      63  0.025812  0.000410
Total      69  0.063650

S = 0.02024   R-Sq = 59.45%   R-Sq(adj) = 55.58%


                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level   N     Mean    StDev   ---+---------+---------+---------+------
10     10  0.60602  0.01886   (-----*-----)
20     10  0.60984  0.01874     (-----*-----)
23     10  0.61399  0.02055       (-----*-----)
24     10  0.61399  0.02055       (-----*-----)
25     10  0.65735  0.02147                         (------*-----)
30     10  0.65782  0.02068                          (-----*-----)
40     10  0.65782  0.02068                          (-----*-----)
                              ---+---------+---------+---------+------
                               0.600     0.620     0.640     0.660
```

Figure 83: ANOVA: Invoking Threshold vs. % Time Monitoring Map, Scenario 4, Segment A

```
One-way ANOVA: S4B MM versus S4B InvTh

Source     DF        SS        MS     F      P
S4B InvTh   6  0.028110  0.004685  8.01  0.000
Error      63  0.036853  0.000585
Total      69  0.064964

S = 0.02419   R-Sq = 43.27%   R-Sq(adj) = 37.87%


                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level   N     Mean    StDev   ----+---------+---------+---------+-----
10     10  0.61053  0.01450     (-----*-----)
20     10  0.60523  0.01643   (-----*-----)
23     10  0.63105  0.02133             (-----*------)
24     10  0.63031  0.02157             (-----*-----)
25     10  0.65725  0.03176                        (-----*-----)
30     10  0.65492  0.02903                       (-----*-----)
40     10  0.65492  0.02903                       (-----*-----)
                              ----+---------+---------+---------+-----
                                0.600     0.625     0.650     0.675

Pooled StDev = 0.02419
```

Figure 84: ANOVA: Invoking Threshold vs. % Time Monitoring Map, Scenario 4, Segment B

```
One-way ANOVA: S4C MM versus S4C InvTh

Source     DF        SS        MS      F      P
S4C InvTh   6  0.105606  0.017601  37.52  0.000
Error      63  0.029556  0.000469
Total      69  0.135162

S = 0.02166   R-Sq = 78.13%   R-Sq(adj) = 76.05%


                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level   N     Mean    StDev   -+---------+---------+---------+--------
10     10  0.53410  0.02008   (---*---)
20     10  0.53722  0.01766    (--*---)
23     10  0.54974  0.02044       (---*---)
24     10  0.56221  0.02021           (---*---)
25     10  0.60629  0.01436                       (---*---)
30     10  0.62861  0.02772                              (---*---)
40     10  0.62861  0.02772                              (---*---)
                              -+---------+---------+---------+--------
                              0.525     0.560     0.595     0.630

Pooled StDev = 0.02166
```

Figure 85: ANOVA: Invoking Threshold vs. % Time Monitoring Map, Scenario 4, Segment C

```
One-way ANOVA: S4A CI versus S4A InvTh

Source    DF       SS       MS     F      P
S4A InvTh  6  0.00815  0.00136  0.84  0.547
Error     63  0.10246  0.00163
Total     69  0.11061

S = 0.04033   R-Sq = 7.37%   R-Sq(adj) = 0.00%


                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level   N     Mean    StDev   --+---------+---------+---------+-------
10     10  0.56429  0.04994              (------------*------------)
20     10  0.56071  0.05203            (------------*------------)
23     10  0.54107  0.04128   (------------*-----------)
24     10  0.54107  0.04128   (------------*-----------)
25     10  0.56607  0.03041               (------------*------------)
30     10  0.56607  0.03041               (------------*------------)
40     10  0.56607  0.03041               (------------*------------)
                              --+---------+---------+---------+-------
                              0.520     0.540     0.560     0.580

Pooled StDev = 0.04033
```

Figure 86: ANOVA: Invoking Threshold vs. % Changes Identified, Scenario 4, Segment A

```
One-way ANOVA: S4B CI versus S4B InvTh

Source    DF       SS       MS      F      P
S4B InvTh  6  0.58735  0.09789  46.90  0.000
Error     63  0.13151  0.00209
Total     69  0.71886

S = 0.04569   R-Sq = 81.71%   R-Sq(adj) = 79.96%


                              Individual 95% CIs For Mean Based on
                              Pooled StDev
Level   N     Mean    StDev  ------+---------+---------+---------+---
10     10  0.33750  0.03414  (---*---)
20     10  0.35714  0.03948     (---*---)
23     10  0.36250  0.04128      (---*---)
24     10  0.35714  0.03765     (---*---)
25     10  0.53929  0.05507                           (---*---)
30     10  0.53750  0.05354                           (---*---)
40     10  0.53750  0.05354                           (---*---)
                             ------+---------+---------+---------+---
                                 0.350     0.420     0.490     0.560

Pooled StDev = 0.04569
```

Figure 87: ANOVA: Invoking Threshold vs. % Changes Identified, Scenario 4, Segment B

```
One-way ANOVA: S4C CI versus S4C InvTh

Source    DF       SS       MS      F      P
S4C InvTh  6  0.38429  0.06405  34.74  0.000
Error     63  0.11617  0.00184
Total     69  0.50046

S = 0.04294   R-Sq = 76.79%   R-Sq(adj) = 74.58%


                              Individual 95% CIs For Mean Based on
                              Pooled StDev
Level   N     Mean    StDev  -----+---------+---------+---------+----
10     10  0.23393  0.03989       (--*---)
20     10  0.20536  0.03496  (---*---)
23     10  0.23393  0.04410       (--*---)
24     10  0.25179  0.06325         (---*---)
25     10  0.39464  0.02721                            (--*---)
30     10  0.36964  0.04128                        (---*---)
40     10  0.36964  0.04128                        (---*---)
                             -----+---------+---------+---------+----
                                0.210     0.280     0.350     0.420

Pooled StDev = 0.04294
```

Figure 88: ANOVA: Invoking Threshold vs. % Changes Identified, Scenario 4, Segment C

REFERENCES

AccuVal Associates. (2012). *Unmanned aerial vehicles: US exploring tech with caution.* Retrieved from http://www.accuval.net/insights/industryinsights/detail.php?ID=171

Adams, C. M., & Biers, D. W. (2000). ModSWAT and TLX: Effect of paired comparison weighting and weighting context. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 609-612.

Alion Science and Technology Corporation (2013). Improved Performance Research Integration Tool Version 3.5.29.89 (Software).

Allender, L. (2000a). Modeling human performance: Impacting system design, performance, and cost. Paper presented at the *Proceedings of the Military, Government and Aerospace Simulation,* pp. 139-44.

Allender, L. (2000b). Tools for modeling human performance in systems through green-colored glasses: An army perspective. Paper presented at the *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennnium, July 29, 2000 - August 4,* pp. 717-720.

Applied Cognition & Training in Immersive Virtual Environments (ACTIVE) Lab. (2011). *ACTIVE lab at UCF.* Retrieved July 6, 2012, from http://active.ist.ucf.edu/LinkClick.aspx?fileticket=BiXl6kVX6JQ%3d&tabid=430

Arciszewski, H. F. R., de Greef, T. E., & van Delft, J. H. (2009). Adaptive automation in a naval combat management system. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 39*(6), 1188-1199.

Audacity (2013). Audacity 2.0 (Software). Available from http://audacity.sourceforge.net/

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science, 8*(4), 321-348.

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors, 48*(4), 693-709.

Baldauf, D., Burgard, E., & Wittmann, M. (2009). Time perception as a workload measure in simulated car driving. *Applied Ergonomics, 40*(5), 929-935.

Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B. (1989). *Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model* No. AD-A210763. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Biers, D. W., & Anthony, C. (2000). Context effects in the measurement of subjective workload: Prolonged exposure to the same or different contexts. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 613-615.

Boles, D. B., & Adair, L. P. (2001a). The multiple resources questionnaire (MRQ). *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 1790-1794.

Boles, D. B., & Adair, L. P. (2001b). Validity of the multiple resources questionnaire (MRQ). *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 1795-1799.

Boles, D. B., Bursk, J. H., Phillips, J. B., & Perdelwitz, J. R. (2007). Predicting dual-task performance with the multiple resources questionnaire (MRQ). *Human Factors, 49*(1), 32-45.

Boles, D. B., Phillips, J. B., Perdelwitz, J. R., & Bursk, J. H. (2004). Application of the multiple resources questionnaire (MRQ) to a complex gaming environment. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 1968-1972.

Card, S. K., Moran, T. P., and Newell, A. (1983). The Psychology of Human-Computer Interaction. Erlbaum Associates Pub.

Cassenti, D. N., Kelley, T. D., & Carlson, R. A. (2010). Modeling the workload-performance relationship. Paper presented at the *54th Human Factors and Ergonomics Society Annual Meeting 2010, HFES 2010, September 27, 2010 - October 1,* pp. 1684-1688.

Cassenti, D. N., & Kelley, T. D. (2006). Towards the shape of mental workload. Paper presented at the *50th Annual Meeting of the Human Factors and Ergonomics Society, HFES 2006, October 16, 2006 - October 20,* pp. 1147-1151.

Cassenti, D. N., Kelley, T. D., Colle, H. A., & McGregor, E. A. (2011). Modeling performance measures and self-ratings of workload in a visual scanning task. Paper presented at the *55th Human Factors and Ergonomics Society Annual Meeting, HFES 2011, September 19, 2011 - September 23,* pp. 870-874.

Clamann, M. P., Wright, M. C., & Kaber, D. B. (2002). Comparison of performance effects of adaptive automation applied to various stages of human-machine system information processing. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 46.* pp. 342-346.

Cooper, G. E., & Harper, R. P. (1969). *The use of pilot ratings in the evaluation of aircraft handling qualities* No. NASA Ames Technical Report NASA TN-D-5153). Moffett Field; CA: NASA Ames Research Center.

Cooper, G. E., & Harper, R. P. (1997). *Cooper-harper handling qualities rating scale* (SP-3300 Flight Research at Ames, 1940-1997 ed.) National Aeronautics and Space Administration History Office.

Cosenzo, K., Chen, J., Reinerman-Jones, L., Barnes, M., & Nicholson, D. (2010). Adaptive automation effects on operator performance during a reconnaissance mission with an unmanned ground vehicle. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 2135-2139.

Cui, K., Sun, L. Y., & Yu, M. (2008). The difference of subjective mental workload between type A and type B men. *Chinese Mental Health Journal, 22*(4), 298-301.

David, H., Mollard, R., Cabon, P., & Farbos, B. (2000). Point-of-gaze evaluation in simulated ATC. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 665-668.

De Visser, E. J., Legoullon, M., Freedy, A., Freedy, E., Weltman, G., & Parasuraman, R. (2008). Designing an adaptive automation system for human supervision of unmanned vehicles: A bridge from theory to practice. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 221-225.

De Visser, E. J., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making, 5*(2), 209-231.

Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2012). Cognitive conflict in human–automation interactions: A psychophysiological study. *Applied Ergonomics, 43*(3), 588-595.

Dey, A., & Mann, D. D. (2010). Sensitivity and diagnosticity of NASA-TLX and simplified SWAT to assess the mental workload associated with operating an agricultural sprayer. *Ergonomics, 53*(07), 848-857.

Di Nocera, ,F., Camilli, M., & Terenzi, M. (2006). Using the distribution of eye fixations to assess pilots' mental workload. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 63-65.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48*(3), 474-486.

Dixon, S. R., Wickens, C. D., & Chang, D. (2004). Unmanned aerial vehicle flight control: False alarms versus misses. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.*

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors, 49*(4), 564-572.

Dorneich, M. C., Mathan, S., Ververs, P. M., & Whitlow, S. D. (2007). An evaluation of real-time cognitive state classification in a harsh operational environment. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 146-150.

Dorneich, M. C., Ververs, P. M., Whitlow, S. D., Mathan, S., Carciofini, J., & Reusser, T. (2006). Neurophysiologically driven adaptive automation to improve decision making under stress. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 410-414.

Dussault, C., Jouanin, J. C., & Guezennec, C. Y. (2004). EEG and ECG changes during selected flight sequences. *Aviation, Space, and Environmental Medicine, 75*(10), 889-897.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697-718.

Eggmeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. *Multiple task performance* (pp. 207-216).

Ellis, S. R., Dorighi, N. S., Menges, B. M., Adelstein, B. D., & Jacoby, R. H. (1997). In search of equivalence classes in subjective scales of reality. *Design of Computing Systems, Proceedings of the Seventh International Conference on Human Computer Interaction.*

Endsley, M.R. (1993). Situation awareness and workload: Flip sides of the same coin. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology,* 906–911.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Eshel, D. (2011, December). Guardian angels: Airborne platforms and sensors are bolstering ocean patrol. *Defense Technology International*, 36-38.

Finomore, V. S. (2006). Measuring the workload of sustained attention. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 1614-1618.

Fox News. (2012, January 26). New pentagon plan would cut ground forces by 100,000. *FoxNews.Com.* Retrieved from http://www.foxnews.com/politics/2012/01/26/panetta-to-outline-new-defense-spending-cuts-at-pentagon/

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology, 50*(1), 61-76.

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., Prinzel, L. J., & Clouatre, K. (2000). Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology & Biofeedback, 25*(2), 103-115.

Funk, H. B., & Miller, C. A. (2001). *User acceptance and plan recognition: Why even perfect intent inferencing might not be good enough.* (Symposium Technical Report No. FS-01-05). Association for the Advancement of Artificial Intelligence.

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

Greenemeier, L. (2011, March 24). Robots arrive at fukushima nuclear site with unclear mission. *Scientific American.*

Haarmann, A., Boucsein, W., & Schaefer, F. (2009). Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Applied Ergonomics, 40*(6), 1026-1040.

Harriott, C. E., Zhang, T., & Adams, J. A. (2011). Evaluating the applicability of current models of workload to peer-based human-robot teams. Paper presented at the *6th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2011, March 6, 2011 - March 9,* pp. 45-52.

Harris, R. M., Iavecchia, H. P., and Bittner, A. C., Jr. (1988) The HOS Micromodels. *Proceedings of the Human Factors Society 32nd Annual Meeting*, pp. 1051-1055.

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

Hart, S. G. (2006). NASA-task load index (NASA-TLX): 20 tears later. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 904-908.

Hjortskov, N., Rissen, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Sogaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology, 92*(1-2), 84-89.

Horner, R. D., Szaflarski, J. P., Jacobson, C. J., Elder, N., Bolon, S., Matthews, G., et al. (2011). Clinical work intensity among physician specialties: How might we assess it? what do we find? *Medical Care, 49*(1), 108-113.

Hunn, B. P., & Heuckeroth, O. H. (2006). *A shadow unmanned aerial vehicle (UAV) improved performance research integration tool (IMPRINT) model supporting future combat systems* No. ARL-TR-3731. Aberdeen Proving Ground, MD: Army Research Laboratory.

Inagaki, T. (2003a). Adaptive automation for comfort and safety. *International Journal of ITS Research, 1*(1), 3-12.

Inagaki, T. (2003b). Adaptive automation: Sharing and trading of control. *Handbook of cognitive task design* (pp. 147-169).

Jordan and Brennen (1992). *Instantaneous self-assessment of workload technique (ISA)*. Retrieved from http://www.skybrary.aero/bookshelf/books/1963.pdf

Jou, Y. T., Yenn, T. C., Yang, L. C. (2011) Investigation of automation deployment in the main control room of nuclear power plants by using adaptive automation. H*uman Factors and Ergonomics in Manufacturing & Service Industries, 21*(4), 350-360.

Jung, H. S. (2001). Establishment of overall workload assessment technique for various tasks and workplaces. *International Journal of Industrial Ergonomics, 28*(6), 341-353.

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

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kaber, D. B., Perry, C. M., Segall, N., Mcclernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics, 36*(5), 447-462.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors, 47*(4), 730-741.

Kaber, D. B., Wright, M. C., & Sheik-Nainer, M. (2006). Investigation of multi-modal interface features for adaptive automation of a human-robot system. *International Journal of Human-Computer Studies, 64*(6), 527-540.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Kantowitz, B. (1987). Mental workload. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 81-121). North-Holland: Elsevier Science Publishers.

Keller, J. (2002). Human performance modeling for discrete-event simulation: Workload. Paper presented at the *Proceedings of the 2002 Winter Simulation Conference, 1.* pp. 157-62.

Klein, M. I., Warm, J. S., Riley, M. A., Matthews, G., Gaitonde, K., Donovan, J. F., et al. (2008). Performance, stress, workload, and coping profiles in 1st year medical students' interaction with the Endoscopic/Laparoscopic and robot-assisted surgical techniques. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 885-889.

Laughery, R. (1999). Using discrete-event simulation to model human performance in complex systems. *Proceedings of the 1999 Winter Simulation Conference,* pp. 815-820.

Laux, L., & Plott, C. (2007). Using operator workload data to inform human reliability analyses. Paper presented at the *Joint 8th IEEE HFPP Conference on Human Factor and Power Plants and 13th HPRCT Annual Workshop on Human Performance, Root Cause, Trending, Operating Experience, Self-Assessment, August 26 -31,* pp. 309-313.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*, 50-80.

Litaker, H. L., Thompson, S., & Archer, R. D. (2008). An evaluation of the long duration use of a head mounted display. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 1640-1644.

Liu, D., Wasson, R., & Vincenzi, D. A. (2009). Effects of system automation management strategies and multi-mission operator-to-vehicle ratio on operator performance in UAV systems. *Journal of Intelligent & Robotic Systems, 54*(5), 795-810.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics, 44*(3), 229-243.

Martin, A. (2012, April). U.S. expands use of underwater unmanned vehicles. *National Defense, NDIA's Business and Technology Magazine.*

Martinez, L. (2012, January 26). Pentagon: Fewer soldiers, more drones will save money. *ABC News.* Retrieved from http://abcnews.go.com/Blotter/pentagon-fewer-soldiers-drones-save-money/story?id=15448631#.UW7VarXU_vt

Mathwave Technologies (2010). EasyFit Version 5.5 (Software). Available from http://www.mathwave.com/help/easyfit/index.html

McCracken, J. H., and Aldrich, T. B. (1984) Analyses of selected LHX mission functions: implications for operator workload and system automation goals. Technical Note AS1479-024-04, Fort Rucker, AL: Army Research Institute Aviation Research and Development Activity.

McGrogan, J., Schneider, M., Wirthlin, R., Coloumbi, J., & Miller, M. (2011). Predicting the impact of multi-aircraft control on UAS operations. *Proceedings of the 2011 Industrial Engineering Research Conference,* pp. 1-8.

MiniTab, Inc. (2013). Minitab Version 16.2.4 (Software).

Mitchell, D. K., & McDowell, K. (2008). Using modeling as a lens to focus testing. Paper presented at the *2008 International Symposium on Collaborative Technologies and Systems, CTS'08, May 19, 2008 - May 23,* pp. 477-482.

Mitchell, D. K. (2008). United we stand: Using multiple tools to solve a multi-dimensional problem. Paper presented at the *52nd Human Factors and Ergonomics Society Annual Meeting, HFES 2008, September 22, 2008 - September 26, 3*. pp. 1939-1943.

Naatanen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

Ng, E., & So, R. H. Y. (2003). The use of cooper-harper technique to enhance the measurement of rated sense-of-presence in virtual and real environments. *Contemporary Ergonomics*, pp. 265-270.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 1*, 44-64.

North, R. A., & Riley, V. A. (1989). A predictive model of operator workload. *Applications of Human Performance Models to System Design*, pp. 81-90.

Or, C. K. L., & Duffy, V. (2005). *A methodology for non-intrusive mental workload measurement in simulated and on-road driving.*

Parasuraman, R., Cosenzo, K. A., & De Visser, E. J. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21*(2), 270-297.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*(4), 665-679.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions, 30*(3), 286-297.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors, 50*(3), 511-520.

Park, J., & Park, S. H. (2007). The effects of various visual enhancements during continuous pursuit tracking tasks. *Human-Computer Interaction: Interaction Platforms and Techniques.* Lecture Notes in Computer Science, vol. 4551, pp. 125-132.

Pepe, A. A., Zachary, W., Read, S., Miller, L., Iordanov, V., Miller, D., & Chopra, G. (2008). Motive and affect based control for uninhabited vehicles. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 226-230.

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

Pickup, L., Wilson, J. R., Norris, B. J., Mitchell, L., & Morrisroe, G. (2005). An integrated workload scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics, 36*(6), 681-693.

Random.org (2012). Random Integer Set Generator. Available from http://www.random.org/integer-sets/

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology.Applied, 9*(2), 119-137.

Recarte, M. A., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology, 11*(2), 374-385.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 185-218). Amsterdam: Elsevier.

Rockwell Automation, Inc. (2010). Input Analyzer Version 13.5 (Software).

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight; A decade of practical use.* No. TR 90019. Farnborough: Royal Aerospace Establishment.

Rottger, S., Bali, K., & Manzey, D. (2009). Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics, 52*(5), 512-523.

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

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics, 35*(11), 991-1009.

Sarno, K. J., & Wickens, C. D. (1992). Predictive workload models and multiple-task performance. Paper presented at the *Innovations for Interactions,* pp. 12-16.

Sauer, J., Kao, C. S., Wastell, D., & Nickel, P. (2011). Explicit control of adaptive automation under different levels of environmental stress. *Ergonomics, 54*(8), 755-766.

Scerbo, M. W. (2001). Adaptive automation. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1077–1079). London: Taylor & Francis.

Sheridan, T. B., & Parasuraman, R. (2006). Human-automation interaction. *Reviews of human factors and ergonomics* (pp. 89–129).

Sheridan, T. B., & Verplank, W. (1978). *Human and computer control of undersea teleoperators*. Cambridge: Massachusetts Institute of Technology, Man-Machine Systems Laboratory.

Steinhauser, N. B., Pavlas, D., & Hancock, P. A. (2009). Design principles for adaptive automation and aiding. *Ergonomics in Design, 17*(2), 6-10.

Tattersall, A. J., & Fairclough, S. H. (2003). Adaptive automation and modes of control. In G. R. Hockey, A. W. K. Gaillard & O. Burov (Eds.), *Operator functional state: The assessment and prediction of human performance degradation in complex tasks.* (pp. 238-248). Amsterdam: ISO Press.

Taylor, G., Reinerman-Jones, L., Cosenzo, K., & Nicholson, D. (2011). Comparison of multiple physiological sensors to classify operator state in adaptive automation systems. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 195-199.

Thompson, S., Litaker, H. L., Archer, R. D., & Howard, R. (2008). Hatch size evaluation for the altair lunar sortie habitat. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 1374-1378.

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

United States Army Research Laboratory. (2010). *Improved performance research integration tool.* Retrieved June 13, 2012, from http://www.arl.army.mil/www/default.cfm?page=445

Urquhart, R. L. (2003). The effects of noise on speech intelligibility and complex cognitive performance. ProQuest Information & Learning. *Dissertation Abstracts International: Section B: The Sciences and Engineering, vol. 63*.

Warm, J. S., & Parasuraman, R. (2007). Cerebral hemodynamics and vigilance. In R. Parasuraman, & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 146-158). New York: Oxford University Press.

Waugh, J. D., Glumm, M. M., Kilduff, P. W., Tauson, R. A., Smyth, C. C., & Pillalamarri, R. S. (2000). Cognitive workload while driving and talking on a cellular phone or to a passenger. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 276-279.

Welford, A. T. (1968). Fundamentals of Skill. London: Methuen.

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

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

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman, & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-101). San Diego, CA: Academic Press.

Wickens, C. D., & Hollands, J. G. (2000). Attention, time-sharing, and workload. *Engineering psychology and human performance* (Third Edition ed., pp. 439). Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Wilson, G. F., & Russell, C. A. (2004). Psychophysiologically determined adaptive aiding in a simulated UCAV task. *Human performance, situation awareness, and automation: Current research and trends* (pp. 200-204). Mahwah, NJ: Erlbaum Associates.

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., Caldwell, J. A., & Russell, C. A. (2007). Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. *International Journal of Aviation Psychology, 17*(2), 219-247.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors, 49*(6), 1005-1018.

Wojciechowski, J. Q. (2007). Modeling human performance with environmental stressors: A case study of the effect of vehicle motion. Paper presented at the *51st Annual Meeting of the Human Factors and Ergonomics Society, HFES 2007, October 1, 2007 - October 5, 2.* pp. 780-783.

Wojcik, T. M. (2002). Human performance modeling and decision analysis in a concept crew workstation. Paper presented at the *Human Factors and Ergonomics Society Conference,* pp. 1910-14.

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