Analyzing the Effects of Dynamic Task Allocation on Human-Automation System Performance

By

Aaron William Johnson

B.S.E. Aerospace Engineering
University of Michigan, 2008

S.M. Aeronautics and Astronautics
Massachusetts Institute of Technology, 2010

SUBMITTED TO THE DEPARTMENT OF AERONAUTICS AND ASTRONAUTICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN AERONAUTICS AND ASTRONAUTICS

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

Signature redacted

Signature of Author: _____

Aaron W. Johnson
Department of Aeronautics and Astronautics
October 17, 2014

Signature redacted

Accepted by: _____

Paulo Lozano
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Analyzing the Effects of Dynamic Task Allocation on Human-Automation System Performance

By

Aaron William Johnson

DOCTOR OF PHILOSOPHY IN AERONAUTICS AND ASTRONAUTICS

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Signature redacted

Certified by: _____

Charles M. Oman
Senior Research Engineer, Senior Lecturer
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Thesis Committee Chair

Signature redacted

Certified by: _____

Kevin R. Duda
Senior Member of the Technical Staff
The Charles Stark Draper Laboratory, Inc.
Thesis Committee Member

Signature redacted

Certified by: _____

Thomas B. Sheridan
Professor Emeritus
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Thesis Committee Member

Analyzing the Effects of Dynamic Task Allocation on Human-Automation System Performance

by
Aaron William Johnson

Submitted to the Department of Aeronautics and Astronautics on
October 17, 2014 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics

Abstract

Modern complex aerospace systems employ flight deck automation to increase the efficiency and safety of systems while reducing operator workload. However, too much automation can lead to overtrust, complacency, and a decrease in operator situation awareness. In an attempt to prevent these from occurring, the operator and the automation often share responsibility for performing tasks. The tasks allocated to each agent are rarely fixed; instead, they can be dynamically re-allocated throughout operations based on the state of the operators, system, and environment. This thesis investigates how dynamic task re-allocation has been implemented in operational aerospace systems, and investigates the effect of control mode transitions on operator flying performance, visual attention, mental workload, and situation awareness through experimentation and simulation.

This thesis reviews the dynamic task allocation literature and discusses the ways in which the concept can be implemented. It highlights adaptive automation, in which the dynamic re-allocation of tasks is initiated by the automation in a manner that is adaptive – in response to the state of the operator, system, and environment – and workload-balancing – with the purpose of keeping the operator in control as much as possible while remaining at a moderate level of mental workload. Adaptive automation is enthusiastically supported in the literature; however, for reasons discussed, it has not been deployed in any operational civilian aerospace system.

In the experiment, twelve subjects sat at a fixed-base lunar landing simulator and initiated transitions between automatic and two manual control modes. Visual fixations were recorded with an eye tracker, and subjects' mental workload and situation awareness were measured using the responses to a secondary two-choice response task and a tertiary task of verbal call-outs of the vehicle state, respectively. Subjects were found to re-allocate attention according to the priority of tasks: during mode transitions from autopilot to two-axis manual control the percent of total attention on the attitude indicator (which was required for the primary flying task) increased 14% while attention on instruments required for the secondary and tertiary tasks decreased 5%. Subjects' conception of task priority appeared to be influenced by instructions given during training and top-down and bottom-up properties of the tasks and instrument displays. The attention allocation was also affected by the frequency of control inputs required. The percent of attention on the attitude indicator decreased up to 13% across mode transitions where the flying task was not re-allocated because the pitch guidance rate-of-change decreased from -9 to 0 °/s throughout the trial. Consequently, fewer control inputs and less attention were necessary later in the trial.

An integrated human-vehicle model was developed to simulate how operators allocate attention in the lunar landing task and the effect this has on flying performance, mental workload, and situation awareness. The human performance model describes how operators make estimates of the system states, correct these estimates by attending and perceiving information from the displays, and use these estimates to control the vehicle. A new attention parameter – the uncertainty in operators' estimates of system states between visual fixations – was developed that directly relates attention and situation awareness. The model's attention block was validated against experimental data, demonstrating an average difference in the percent of attention ≤3.6% for all instruments. The model's predictions of flying performance, mental workload, and situation awareness were also qualitatively compared to experimental data.

Thesis supervisor:
Charles M. Oman
Senior Research Engineer, Senior Lecturer
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

Acknowledgements

When I started my Ph.D. three years ago, all I had were a bunch of disparate ideas that didn't really fit together. I knew that I wanted to focus on adaptive automation, that I would do an experiment in the Draper lunar landing simulator, and that I would eventually make a model of... something. But that was it. Many people have helped this thesis to grow from that rough outline, and they have my sincere thanks.

This research would not have been possible without financial support from the National Space Biomedical Research Institute (through NASA NCC 9-58, Project HFP02001), the MIT Department of Aeronautics and Astronautics, and the Charles Stark Draper Laboratory, Inc.

Thank you to my committee members for guiding me along this three-year path. You helped me to define the research questions, pursue answers, and connect everything together into this thesis. To Chuck Oman, thank you for always encouraging me to probe deeper and make sure that I really knew what I was talking about. To Kevin Duda, thank you for constantly making time to hear out my ideas. To Tom Sheridan, thank you for sharing your incredible knowledge of human-automation interaction.

Thank you to my thesis readers for taking the time to read numerous drafts in more- and less-finished states and provide valuable feedback. To Andy Liu, thank you for your comments, all of which were detailed and insightful. To Alan Natapoff, thank you for all of the help you've given me in analyzing my complex experimental dataset and uncovering the important results.

Thank you to everyone at the Charles Stark Draper Laboratory, Inc. for supporting me in numerous ways. To Justin Vican, thank you for always being available to help us troubleshoot the lunar landing simulator. To Justin Kaderka, thank you for modifying the simulator for our experiment and helping to run subjects. To Gail Dourian and Steve Kolitz in the Education Office, thank you for organizing the DLF program and for providing funding for travel and tuition over the past year.

Thank you to everyone at MIT for the opportunities you've given me outside of research. To Leia Stirling, Dava Newman, David Mindell, and Jeff Hoffman, thank you for having me as your TA and allowing me to develop my teaching skills. To Becky Fearing and Chad Galts, thank you for giving me a chance to develop my science communication skills in really fun ways.

Thank you to all of my friends, both inside and outside the MVL, for being awesome. You've provided encouragement and laughter exactly when they were needed. In particular, I would like to thank the friends with whom I have spent my entire tenure in the MVL (not to rub it in, but in order of graduation): Torin Clark, Justin Kaderka, Brad Holschuh, and Allie Anderson. You will all make great astronauts someday. (Also, let's collaborate on projects so we have an excuse to visit each other.)

Thank you to my family, both the Johnsons and the Millers, for sharing the good times and helping me through the tough times. Most importantly, thank you to my amazing wife, Gretchen. Words can't express how thankful I am for all the financial, mental, and emotional support you've given me the past three years. I love you, and I'm incredibly excited for the rest of our lives together. This thesis is for you.

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

- A-SA = Attention-Situation Awareness model
- AOI = area of interest
- APU = average peak uncertainty
- AUC = area under the [receiver operating characteristic] curve
- Auto = automatic control of attitude and rate-of-descent
- CASCaS = Cognitive Architecture for Safety Critical Task Simulation
- DEM = digital elevation map
- FAA = Federal Aviation Administration
- HPM = human performance model
- HSD = horizontal situation display
- HTA = hierarchical task analysis
- HUD = head-up display
- LAP = landing aim point
- LoA = level of automation
- LPR = landing point redesignation
- MIDAS = Man-machine Integration Design and Analysis System
- MSE = mean square error
- MT = mode transition
- NASA = National Aeronautics and Space Administration
- OCM = Optimal Control Model
- PA = Pilot's Associate
- PF = pilot flying
- PFD = primary flight display
- PM = pilot monitoring
- RHC = rotational hand controller
- ROC = receiver operating characteristic
- RoD = rate-of-descent
- RPA = Rotorcraft Pilot's Associate
- SOP = standard operating procedure
- TA = two-axis manual control with automatic rate-of-descent control
- TA+RoD = two-axis (pitch and roll) and rate-of-descent manual control
- VUE = value, uncertainty, and effort attention primitives

This page intentionally left blank.

When the pilot and airplane exchange
Control over tasks, as arranged,
The pilot makes a decision
How to allocate vision
If the level of workload has changed.

1.0 Introduction

Modern complex aerospace systems have flight deck automation capabilities that execute tasks previously performed by human operators. Commercial transport aircraft can operate almost completely automatically from takeoff to landing, and fully-autonomous lunar landing vehicles have been proposed for future uncrewed resupply missions to the Moon (National Aeronautics and Space Administration, 2008). The Apollo Lunar Module had a fully automatic landing capability (Bennett, 1972), but it was never used. The interaction between operators and these autonomous systems often occurs without trouble. However, sometimes this interaction can break down with disastrous consequences. Take for example the 2013 crash of Asiana Flight 214 at San Francisco International Airport (National Transportation Safety Board, 2014). The pilots believed that the autothrottle was engaged when it was actually in a "hold" mode. As a result, neither the pilots nor the automation were controlling the airspeed. The Boeing 777 descended much faster than usual, and the pilots failed to notice the issue until it was too late. The aircraft impacted the seawall at the end of the runway, causing 181 injuries and 3 fatalities. This accident, one of many caused in part by poor human-automation interaction, emphasizes that system designers must work to facilitate effective collaboration between the human pilot and the automated systems as the aerospace field looks to the next generation of these complex vehicles.

Automating tasks that were once completed by a human has the potential to increase the efficiency and safety of systems while reducing operator workload (Ephrath & Curry, 1977; Wiener & Curry, 1980; Wickens & Hollands, 2000). However, research has also identified a number of ways in which automation can create new problems for the human operator. Bainbridge collectively termed these problems the "ironies of automation" (1983). Operator workload can be increased if automation is applied at the wrong time (Wiener, 1989) or if the operator does not trust the automation (Lee & Moray, 1992; Lee & Moray, 1994; Parasuraman & Riley, 1997; Lee & See, 2004; Dixon et al., 2007). Over-reliance can result in complacency in the short term (Parasuraman et al., 1993) and the loss of manual flying skills in the long term (Wiener & Curry, 1980), which can both lead to incidents and accidents (Wood, 2004; Gillen, 2010; Lowy, 2011; Stock et al., 2013; National Transportation Safety Board, 2014). If insufficient information is presented about the automation's actions, the operator may lose awareness of the automation state and have a decreased ability to anticipate the automation's behavior (Sarter & Woods, 1995; Sarter et al., 1997; Woods & Sarter, 1998). This may result in the operator performing inappropriate actions without realizing it. The Federal Aviation Administration (FAA) has encouraged airlines to "provide manual flight operations when appropriate" (Federal Aviation Administration, 2013) and issued more than two dozen recommendations relating to automation and training in the report of the Flight Deck Automation Working Group (2013). The National Aeronautics and Space Administration (NASA) Human Research Roadmap also lists poor human-computer, -robot, and -automation interaction as potential risks for future planetary exploration that could result in an increase in crew errors, injuries, and failed mission objectives (National Aeronautics and Space Administration, 2014).

One solution to mitigate the increased complacency, loss of manual flying skills, and decreased awareness caused by over-automating is to allocate some, but not all, tasks to the automation. This removes a portion of the workload burden from the operator but still keeps him in the loop as a way to preserve awareness of the state of the aircraft and the automation (Kaber & Endsley, 1997). These task allocations can be described as intermediate levels of automation (Sheridan & Verplank, 1978; Endsley & Kaber, 1999; Parasuraman et al., 2000). Level of automation (LoA) hierarchies describe a continuum that spans from fully manual single or multiloop control (low LoA) to full automation (high LoA) where the operator sets the highest-level goal and trusts the automation to achieve it without further human monitoring. At intermediate LoAs the operator is involved in "supervisory control" (Sheridan, 1992). Here, the automation is responsible for the lower-level control loops – gathering information and performing actions – and the operator is responsible for the higher-level control loops – processing information, making decisions, and setting intermediate goals. The operator also has the responsibility for supervising the automation and ensuring that it is performing correctly.

For practical and safety reasons, and because of unanticipated events or subsystem failures, the LoA cannot be the same in all situations (Airbus SAS & Flight Safety Foundation, 2006). As a result, modern complex aerospace systems feature ***dynamic task allocation***, where the allocation of tasks between the human and the automation is not fixed during operations. This re-allocation of tasks is often ***adaptive***, meaning it responds to the state of the operator, system, and environment. It is possible for dynamic task allocation to be non-adaptive, as is discussed in Section 2.1.2. This is rare, and in this thesis dynamic task allocation is assumed to be adaptive unless noted otherwise. It is also possible for systems to be adaptive without featuring a re-allocation of tasks, as is discussed in Section 3.2.

Current aerospace systems feature dynamic task allocation in the form of control mode transitions (Boeing, 2003b). Mode transitions during operations are encountered when the vehicle enters a new phase of flight, or they can be forced by an equipment failure. Adaptive dynamic task allocation can also be ***workload balancing***, meaning that it occurs with the intent of keeping the operator in control as much as possible while remaining at a moderate level of mental workload (Rouse et al., 1987). For example, the operator may initiate a change from high to low LoA in order to practice flying manually during low-workload phases of flight, or the automation may decide to remove tasks from an overloaded operator. As Section 3.2 demonstrates, not all adaptive dynamic task allocation is workload balancing.

This thesis is structured around a fundamental, overarching hypothesis concerning how a human operator responds to dynamic task allocation (Figure 1). Each of the chapters in this thesis addresses this hypothesis in a different manner.

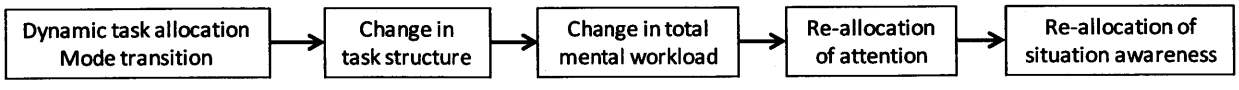


Figure 1. Chapters of this thesis are structured around this proposed hypothesis describing operator response to dynamic task allocation.

Dynamic task allocation causes a change in the ***task structure***. The task structure describes the current and future tasks that fall under the responsibility of the human or the automation. Accomplishing a task at the desired level of performance requires a minimum amount of cognitive and attention resources, or ***mental workload*** (O'Donnell & Eggemeier, 1986). This is different than the physical workload that an operator must expend to complete a task. Humans have a limited capacity of cognitive and visual attention resources that they can allocate among the tasks to be performed (Wickens, 1980; Wickens & Hollands, 2000). When the total cognitive and visual attention demands exceed the limiting capacity, the operator must sacrifice attention on the lower-priority tasks in favor of the higher-priority tasks. If attention on a task is reduced below the minimum requirement there will be a noticeable decrease in performance on this task.

The reduction of attention on the lower-priority tasks also causes a decrease in ***situation awareness*** on these tasks. Situation awareness is defined by Endsley as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (1995, p. 36). Endsley did not define what tasks are part of "the environment," but the most common interpretation is everything beyond the immediate primary task (e.g. flying the vehicle), particularly the set of information needed to make appropriate strategic decisions. Consequently, experimental measures of situation awareness often focus on system states that do not directly relate to the primary task. This interpretation of situation awareness is narrow, and becomes problematic when the task structure changes across a mode transition. System states that were involved in the primary task may become a part of the environment and vice versa. This thesis takes a broader view of situation awareness, making no distinction between system states that belong to the environment and those that belong to the primary task. It is assumed that the operator has an internal model of all key system states and is able to project these estimates into the future based on knowledge of the vehicle dynamics. It is impossible for operators to have perfect knowledge of the vehicle dynamics, and as a result these projections have uncertainty. The rate at which the uncertainty grows is proportional to the accuracy of the operator's internal model and the rate at which the state is changing. By making observations of the actual state, the operator can "correct" these estimates, reducing the uncertainty. However, the uncertainty begins to grow again when the operator looks away.

When an operator reduces his attention on a lower-priority instrument across a mode transition, he has less frequent opportunities to correct his estimate of the system state. This, in turn, increases the average uncertainty in the estimate and reduces the operator's situation awareness of this system state. This decrease in situation awareness is greatest for the system states that have a high rate-of-change, or for which the operator has an inaccurate mental model. Like mental workload and attention, the total amount of operator situation awareness does not decrease. Instead, it is allocated among the tasks in proportion to the attention that they each receive. As a result, the operator usually has high situation awareness of higher-priority tasks and low situation awareness of lower-priority tasks.

This hypothesized operator response to dynamic task allocation (Figure 1) explains how an operator can continue to satisfactorily perform the primary task even as his workload increases (Yerkes & Dodson, 1908; Hebb, 1955). He is sacrificing attention, performance, and situation awareness on the other tasks. Performance on the primary task only drops whenever the workload demands of the primary task itself exceed the total attention capacity, or if the operator fails to re-allocate attention in the proper manner.

Operator response to dynamic task allocation has been investigated in part, but never as a complete five-stage process as this thesis proposes. Previous research has found that pilots with better situation awareness, measured as the response to one query, had lower perceived workload and pilots with poor situation awareness had higher perceived workload (Yu et al., 2014). Other research has shown that situation awareness of a vehicle state, measured by task performance (Ratwani et al., 2010) or by verbal situation awareness queries (Moore & Gugerty, 2010), increased as subjects allocated more attention to the instrument displaying that state. These studies collectively probe the relationship between attention, mental workload, and situation awareness, but not during dynamic task allocation. Previous work done on the same project that supported this thesis found that reversion to manual control increased subjects' mental workload and decreased the measure of situation awareness in proportion to the number of manually-controlled vehicle axes during simulated lunar landings (Hainley et al., 2013). However, this prior study did not measure visual attention. Therefore, this thesis investigates the complete operator response to dynamic task allocation hypothesized in Figure 1 and develops evidence that supports the connections between changes in the task structure, visual attention, mental workload, and situation awareness.

1.1 Research Aims and Thesis Organization

The first aim of this thesis is to clearly define dynamic task allocation and explain the ways in which it can occur. **Chapter 2** examines the ways in which adaptive dynamic task allocation can be implemented in the context of the existing literature. Its primary focus is on the question of *decision authority* – the extent to which each agent (the human and the automation) has authority to re-allocate tasks. **Chapter 3** investigates the extent to which dynamic task allocation occurs in current aerospace systems. It begins with a hierarchical task analysis of approach and landing in the Boeing 767. Then, it discusses adaptive systems in aircraft of more recent design and considers whether or not they are examples of workload-balancing, adaptive dynamic task allocation. Finally, Chapter 3 discusses "associate systems" in military aircraft and rotorcraft and draws general conclusions about why operational systems do not employ dynamic task allocation as the literature promotes.

The scientific aim of this thesis is to investigate how experimental subjects re-allocated visual attention when performing control mode transitions in a simulated lunar landing task. This is covered in **Chapter 4**. To the author's knowledge, this experiment is the first to concurrently measure mental workload, situation awareness, flying task performance, and visual attention across control mode transitions. This allows for an analysis of how the attention re-allocation across a mode transition affects

other aspects of an operator's performance and cognitive state. These experimental results are used to hypothesize the factors that motivated subjects to re-allocate their attention in the manner that they did.

The engineering aim of this thesis is to develop a closed-loop human operator-vehicle model that simulates operator visual attention, flying performance, mental workload, and situation awareness during control mode transitions. This is the focus of **Chapter 5**. To the author's knowledge, the attention block in this integrated model is the first to predict the attention allocation for both supervisory and manual control tasks, and to capture changes in the attention allocation across mode transitions as tasks vary in priority and shift between supervisory and manual control. The chapter concludes by comparing the model's predictions of attention, mental workload, situation awareness, and flying task performance to experimental data obtained in Chapter 4.

The final aim of this thesis is to provide recommendations for future research. This is one of the foci of **Chapter 6**, along with an executive summary of the entire thesis. These recommendations are intended to help bridge the gap between the findings of the existing research and how aircraft manufacturers currently implement dynamic task allocation.

The questions this chapter expands
Are how, when, and at whose command
There may be some refinement
Of tasks and their assignment
To both automation and man.

2.0 Implementation of Workload-Balancing Dynamic Task Allocation through Adaptive and Adaptable Automation

As described in Chapter 1, dynamic task allocation is a necessity in real-world operational systems. A task re-allocation may be forced by a system failure, or it may be an adaptive response to keep the operator's mental workload at a moderate level. This chapter focuses on the latter and addresses the key questions of how, when, and under whose command the task structure can be dynamically re-allocated. It cites a number of possible answers for these questions from the existing research and, when available, highlights recommendations from the current body of research.

2.1 Model of Dynamic Task Allocation

Figure 2 shows a conceptual model of an aerospace system with a human operator and automation. Tasks in the task structure are allocated to both agents, who make control inputs to the system and receive information about the system state and the environment. The allocation of tasks to the human and the automation can change based on the answers to the following three key questions:

- Who or what decides to dynamically re-allocate tasks?
- What factors trigger the dynamic re-allocation of tasks?
- How should tasks be dynamically allocated to the operator and the automation?

The remainder of this chapter addresses each of these questions in turn.

2.1.1 *The Decision Authority*

The agent – the operator or the automation – that chooses to re-allocate tasks during operations is called the ***decision authority***. When the dynamic task allocation is both adaptive and workload balanced, two specific terms are used in the literature to distinguish between the decision authority: ***adaptable automation*** when the operator is the decision authority and ***adaptive automation***[1] when the automation is the decision authority (Oppermann, 1994). Adaptable automation keeps the authority to allocate tasks with the operator, who is ultimately responsible for system safety. This is the stated goal of Boeing and Airbus's automation philosophies, which guide the design of automation in their aircraft (Abbott, 2001). Although both manufacturers use flight control computers to simplify their aircraft handling qualities and incorporate some degree of flight envelope and load protection, both philosophies leave much of the final authority for automation use to the pilots. As detailed in the Chapter 3, most autoflight automation used in aircraft cockpits today remains adaptable, and not adaptive.

[1] Note that "adaptive systems" are not necessarily examples of "adaptive automation," which specifically refers to adaptive, workload-balanced dynamic task allocation. Section 3.2 provides examples of this distinction. These terms come from prior literature, which is why they are used in this thesis despite the potential for confusion.

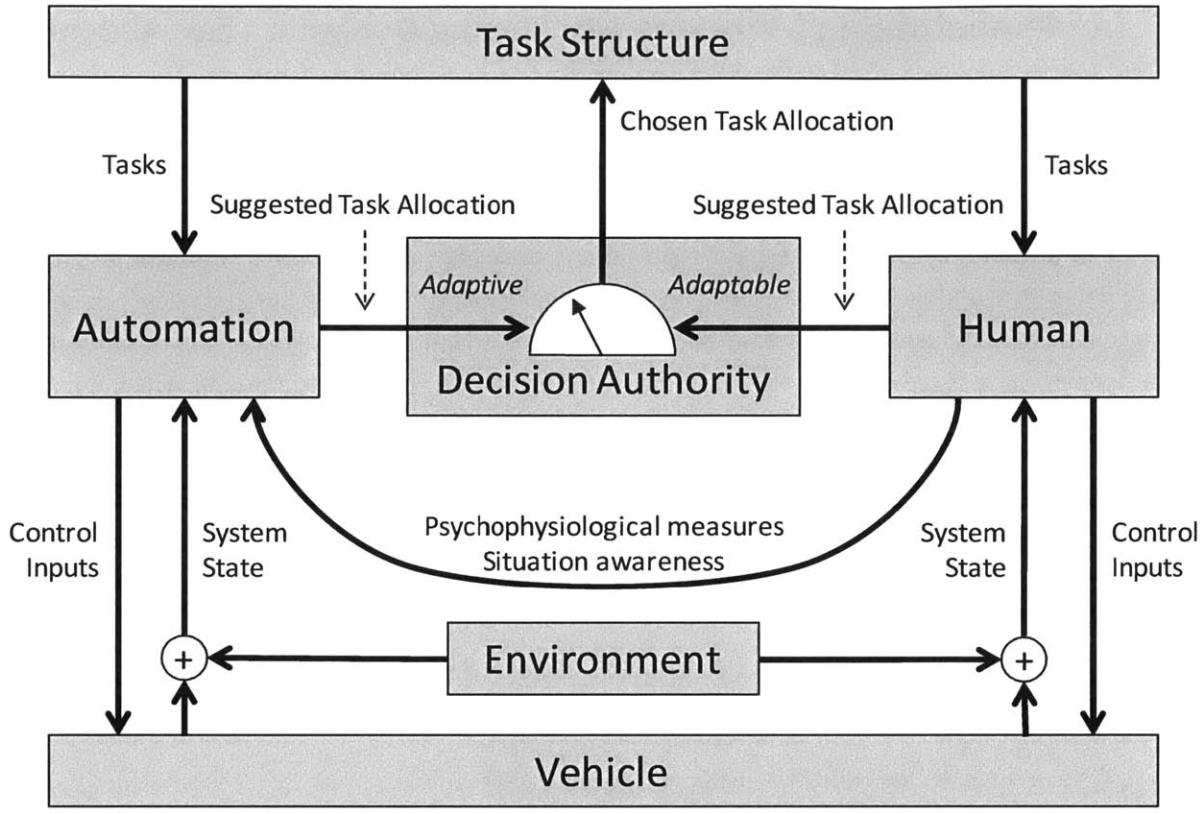


Figure 2. Conceptual model of an aerospace system that responds to inputs made by a human operator and automation. Each agent can each perform tasks and suggest new task allocations with a certain degree of authority.

One concern with adaptable automation is that it loads the operator with an additional task – evaluating his workload and the situation and determining whether or not to re-allocate tasks – at a time when his mental workload may already be high (Bailey et al., 2006). As a result, the operator may make an expedited or uninformed decision and fail to employ automation in the best way to maximize system performance. Furthermore, there are situations when the operator may be unable to re-allocate tasks. He may not have enough time or he may be physically unable to do so (e.g., due to high g-loads) (Scerbo, 1996; Scerbo, 2007). Adaptive automation is not without its limitations either. Automation surprise and mode confusion become a concern when the automation's actions are not transparent and understandable to the operators (Sarter & Woods, 1995; Woods & Sarter, 1998).

Adaptive and adaptable automation are not the only possible decision authorities; rather, they are the two extremes of a continuum with intermediate degrees of operator and automation authority in between (Clamann & Kaber, 2003; Hancock, 2007; Sauer et al., 2011). For example, the automation may suggest a certain task allocation that the user can confirm or choose to ignore. This is known as "pilot command by initiation" (Hancock, 2007) or "management by consent" (Billings, 1996a). Or, the automation may suggest a certain task allocation that will automatically go into effect unless the operator vetoes it. This is

called "pilot command by negation" (Hancock, 2007) or "management by exception" (Billings, 1996a). Another possibility is for the automation to wait for the operator to take an action, warn him if the action has not been performed close to the deadline, and then only perform the action itself when the deadline arrives and the operator has not responded. In these three examples both the pilots and the automation share decision authority to different degrees.

Much of the existing dynamic task allocation literature has focused on enabling adaptive automation and investigating the benefits it provides for the human operator. In particular, researchers have examined the factors that the automation can use to trigger a dynamic re-allocation of tasks (Section 2.1.2) and how the automation should re-allocate tasks between itself and the operators (Section 2.1.3).

2.1.2 *Triggering Dynamic Task Allocation in Adaptive Automation Systems*

If the automation in a complex aerospace system has any degree of authority to re-allocate tasks, it needs some criteria to determine when to do so. The process by which the automation makes this decision is called the ***trigger***. Numerous triggers have been investigated theoretically and experimentally. However, it remains difficult to measure vehicle and/or operator state during real-time operations.

One option for triggering a task re-allocation is to alternate periods of manual and automated control over a certain time interval without any regard to the state of the system, environment, or operator. This has been shown to increase subjects' detection of automation failures (Parasuraman et al., 1993; Parasuraman et al., 1996). In these experiments, transferring control to the subjects prevented them from becoming complacent when control was returned to the automation, and encouraged them to supervise the automation more carefully.

Alternating periods of manual and automated control is an example of non-adaptive dynamic task allocation. As a result, this trigger is only beneficial for operations with continuous, unchanging tasks, like the cruise phase of flight. If this trigger were implemented during a phase of operations in which new tasks were always beginning, like approach and landing, the tasks might be re-allocated at an inappropriate time. For example, the operator could regain control at the very moment that a difficult task begins, causing him to quickly become overloaded.

The most common adaptive triggers described in the literature are various proxy metrics of operator mental workload. These triggers are workload balancing, as described in Chapter 1. The goal is to remove enough tasks from the operator so that his cognitive and attention resources are not exceeded and he will be able to devote the necessary amount of resources to each task for which he is responsible. Prior research in laboratory settings has found that adaptive automation is able to provide a performance benefit to operators when it is matched to their workload – providing more automation assistance during times of high operator workload and less during times of low workload (Parasuraman et al., 1999).

One can define six classes of operator workload triggers, which assess or predict workload in different ways: 1) detection of critical events, 2) use of subjective mental workload assessment scales, 3) measurement of operator primary task performance, 4) measurement of operator secondary task performance, 5) psychophysiological assessments, and 6) use of operator performance models.

2.1.2.1 Critical Events

When triggering on critical events, the amount of automation increases or decreases when certain pre-defined events occur that are anticipated to raise or lower the operator's workload (Parasuraman et al., 1999; Scallen & Hancock, 2001; de Visser & Parasuraman, 2011). Example critical events include the appearance of an unidentified target on a military aircraft's radar, which initiates a shift from a monitoring phase to an investigation phase, or the beginning of the final approach in commercial aviation. Triggering on critical events is the simplest workload trigger, as it does not require the automation to take any measurements from the operator. However, this trigger may only partially or implicitly reflect the operator's true workload. If the assumptions about the operator's workload after the critical event are incorrect, the automation may be inappropriately applied.

2.1.2.2 Subjective Assessments

Subjective assessments help the operator to rate his own level of mental workload using a structured scale. These techniques ask the operator a series of questions and use the responses to give an estimate of mental workload. The Modified Bedford Workload Scale gives a single overall measure of workload and assesses the amount of spare attention an operator has while performing a flying task (Roscoe & Ellis, 1990). On the other hand, the NASA-TLX technique assesses six different dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration level (Hart & Staveland, 1988). This gives a more detailed picture of a task's workload demands, but requires the operator to answer many more questions. Strengths of subjective measures are their relative ease of implementation and high face validity (Proctor & van Zandt, 2008). However, they have the same disadvantage as adaptable automation in that they give the operator an additional task to complete at a time when his mental workload may already be high. For this reason, subjective assessments are often used in experimental settings where they can be administered after a task is performed. Other disadvantages of subjective assessments are that operators may confound mental and physical workload, or that they may have trouble distinguishing task difficulty from the workload required to deal with these task demands (O'Donnell & Eggemeier, 1986).

2.1.2.3 Primary Task Performance Measurements

This class of triggers uses measurements of the operator's performance on the primary task to infer his workload level. It assumes an inverted-U-shaped relationship between performance and workload, the Yerkes-Dodson Law: performance on the primary task drops when the overall workload becomes either very high or very low and increases when workload returns to a manageable intermediate level (Yerkes &

Dodson, 1908; Hebb, 1955). If the operator's mental workload varies, but stays within this intermediate region, performance on the primary task will not visibly change. This is a strong disadvantage of triggering adaptive automation on primary task performance: it is only sensitive to very high or very low levels of mental workload. While some research has used primary task performance to measure workload, many researchers instead use secondary task performance.

2.1.2.4 Secondary Task Performance Measurements

As described in Chapter 1, attention on the secondary task decreases whenever the attention demands of the primary task increase if both tasks compete for the same sensory and cognitive resources (Wickens, 1980; Wickens & Hollands, 2000). The operator sacrifices performance on the secondary task in order to keep satisfactory performance on the primary task. As a result, secondary task performance is more sensitive to changes in workload than primary task performance. Many different secondary tasks have been used in the literature to measure operator workload: gauge monitoring (Kaber & Riley, 1999; Clamann et al., 2002; Clamann & Kaber, 2003; Kaber et al., 2005), change detection (Parasuraman et al., 2009), and responding to a light illumination (Lowenthal, 2012; Hainley et al., 2013), for example. Experiments have also been conducted using multiple primary and secondary tasks in conjunction (Calhoun et al., 2011).

Unlike critical events, triggering on primary or secondary task performance directly reflects the operators' mental workload. Triggering on task performance also has the advantage of being able to react to unexpected changes in workload. One major disadvantage is that operator performance may vary between individuals in complex, real-world tasks due to experience, fatigue, or other factors. This inherently large variability makes it difficult to define performance criteria that are sufficiently reliable across all operators. The criteria must either be set to constant values, providing poor results for a large portion of the operators, or tuned for each operator, which requires extensive upfront costs in money and time. Another limitation is that primary and secondary task performance measurements are reactive to changes in workload, rather than predictive. It would be more beneficial for the automation to anticipate when subjects will become overworked to the point that performance will decrease and to increase the amount of automation before this occurs.

2.1.2.5 Psychophysiological Assessments

Sometimes grouped together with operator performance measurements, psychophysiological assessments (also called augmented cognition or neuroergonomics) record psychological and physiological measurements that serve as proxies for operator mental workload or engagement. Psychophysiological measurements investigated in the literature include heart rate variability, galvanic skin response, eye movement, and brain activity from fNIR and EEG recordings (Casali & Wierwille, 1984; Wilson, 2002; St. John et al., 2004; Bailey et al., 2006; Fidopiastis et al., 2009; de Greef et al., 2009).

Psychophysiological assessments are an attractive trigger because they provide a continuous real-time estimate of mental workload. They can also be used to estimate mental workload when the operator is supervising the automation and there are no operator performance metrics to assess. Psychophysiological assessments are not mentally intrusive to the tasks being performed, and it is believed that they have less of an effect on operator performance. As with operator performance measurements, psychophysiological assessments are reactive to changes in workload, rather than proactive (Scerbo, 2007). There is also a debate about the sensitivity and diagnosticity of psychophysiological assessments and whether they actually capture changes in workload or engagement (Casali & Wierwille, 1984; Scerbo, 2007; Cummings, 2010). Psychophysiological assessments tend to have high intra- and inter-subject variability, making true changes in workload harder to predict. They also cannot make instantaneous predictions of workload because they require data collected over a period of time. Lastly, the equipment required to make psychophysiological assessments can be physically intrusive to the operator.

2.1.2.6 Operator Performance Modeling

In this class of triggers, the automation uses human performance models in real-time to predict the intent or behavior of the operator and estimate his future workload (Rencken & Durrant-Whyte, 1993; Sharit, 1997; Miller et al., 1999; Goodrich et al., 2004; Arciszewski et al., 2009; de Greef & Arciszewski, 2009; de Greef et al., 2010; Klein & van Lambalgen, 2011). Operator performance modeling uses much of the same information as other triggers (e.g. upcoming critical events or operator task performance), but combines these measurements in specific ways to predict operator workload in the future. It is not reactive like the other triggers discussed. For example, an operator performance model may indicate a slight increase in operator workload that, by itself, does not currently exceed the threshold for a workload-triggered task re-allocation. However, knowing the current operational context and upcoming events, the operator performance model may predict that the operator's workload will increase above the threshold in the near future. In response, tasks can be re-allocated to prevent the operator's workload from ever exceeding this threshold. One limitation of this trigger is that it is only as good as its model, and underlying assumptions and any deficiencies in the model will lead to inappropriate automation. The question – to be considered later in this thesis – is whether such algorithms are sufficiently reliable that they can be deployed in operational systems.

2.1.3 *Allocation of Tasks*

During operations, tasks should be dynamically re-allocated to the agent – the operator or the automation – that can best perform them given the current system state. This will help to assure satisfactory system performance. The fixed and dynamic allocation of tasks has been the focus of many theoretical, analytical, and experimental investigations (see Sherry & Ritter, 2002 and Marsden & Kirby, 2004 for reviews). The most basic task allocation guidelines simply outline which tasks humans can perform better than the automation and vice versa. Fitts' List states that automation is good at performing

repetitive tasks that might bore human operators, tasks that necessitate a high degree of precision, and time-critical tasks that require a faster response than humans can provide (Fitts, 1951). Conversely, the human operators are better at thinking creatively and improvising in unfamiliar situations based on past experience. Billings' concept of "human-centered automation" is another set of guidelines that prescribes how automation can be best implemented as a tool to enhance the human pilot's performance (Billings, 1996). These guidelines state that the automation should be predictable, understandable, and only used in situations where it is necessary and not just because the technology exists to automate a function. Billings' guidelines are more specific than Fitts', discussing specific human-centered automation requirements for aircraft control automation, information automation, and air traffic control management automation.

Fitts' List can serve as a set of general guidelines that provide a starting point for design. However, the capabilities of automation have greatly increased since 1951 and the list of tasks that automation can perform better than humans has grown. Furthermore, the best allocation of tasks is more complicated than Fitts discussed. It is context-specific, depending on a number of factors including the training and ability of the operators, the capabilities of the automation, the operating procedures, and the mission to be completed. Pritchett et al. (2013a, 2013b) have developed a model that aims to address some of these context-specific factors by quantifying the "effectiveness" of particular task allocations in commercial aviation approach and landing. In this model, the researcher uses a cognitive work analysis to describe all of the tasks that must be completed and assigns them to the simulated human operator or the automation. Additional teamwork actions are also created based on the task allocation. Once the task allocation is set, the model quantifies how "effective" the task allocation is on the basis of 8 different metrics: 1) workload, 2) mismatches between responsibility and authority, 3) stability of the human's work environment, 4) coherency of [task] allocation, 5) interruptions, 6) automation boundary conditions, 7) system cost and performance, and 8) human's ability to adapt to context.

More generally, several experimental studies have investigated the effects of allocating tasks to the human or automation based on the information processing stages they involve. Endsley and Kaber (1999) and Parasuraman et al. (2000) both discussed how the level of automation – a continuum of task allocations between the automation and operator that range from fully manual to fully automatic – does not need to be consistent across each stage of information processing. The authors used a simplified model of information processing consisting of four stages: 1) information acquisition, 2) information analysis, 3) decision and action selection, and 4) action implementation (in the terminology of Parasuraman et al., 2000). In the first stage, information is gathered from the environment through a number of different sensory channels and pre-processed. The information capacity of each sensory channel is limited (e.g. an operator's foveal vision only encompasses about 2° visual angle), which makes attention a serial process. In the second stage, this information is consciously perceived, manipulated, and analyzed. The second stage lasts just prior to the point of decision, which is contained within the third stage. Once an action has been decided upon it is carried out in the fourth stage and the process repeats. Parasuraman et al. (2000) discussed ten levels of automation, and how these did not need to be consistent

across the four stages. Instead, tasks involving different information processing stages could employ different levels of automation. Kaber and Endsley (1999) discussed ten similar level of automation, and outlined in detail how the four stages of information processing were allocated to the human or the automation under each level.

Several studies have found that dynamically automating tasks involving information acquisition and action implementation led to better primary and secondary task performance than when information analysis and decision making were automated (Clamann et al., 2002; Clamann & Kaber, 2003; Kaber & Endsley, 2003; Kaber et al., 2005). The authors hypothesized that this occurred because it was easier for subjects to understand what the automation was doing when it was gathering information or carrying out an action the subjects selected. When the automation was analyzing information or making a decision, the subjects had to spend additional time checking that what the automation did was in line with their mental model of the situation.

2.2 Conclusion

This chapter discusses three key questions that, when answered, describe how dynamic task allocation can be implemented in operational systems. The first and most important question asks who or what decides to dynamically re-allocate tasks. When the operator serves as this decision authority, workload-balancing, adaptive dynamic task allocation is called adaptable automation. When the automation is the decision authority, it is called adaptive automation.

The second questions asks what factors should trigger the dynamic re-allocation of tasks when the automation is the decision authority. Most of the research has investigated mental workload-based triggers. The operator can give subjective assessments of his own workload, or the trigger can measure proxies for mental workload including the occurrence of pre-defined events that are assumed to change workload, operator primary and secondary task performance, and psychophysiological assessments. The trigger can also combine these measurements with operator performance models to predict future changes in workload. While these triggers have all been investigated theoretically and in simplified experiments, it still remains difficult to measure or estimate the vehicle and/or operator state during real-time operations.

The final question asks how tasks should be dynamically allocated to the operators and the automation. Theoretical guidelines, like Fitts' List or human-centered automation, give general, high-level starting points for the task allocation. Other research has developed models to simulate task allocations and quantify their "effectiveness", and has experimentally investigated how operators perform when the automation executes tasks that involve the different stages of information processing.

As this chapter has shown, much of the existing literature has been focused on investigating the benefits of adaptive automation and developing the technologies and techniques that enable it. However, the next chapter shows that aircraft manufactures do not appear to be as enthusiastic about adaptive automation as the research community.

This page intentionally left blank.

"Adaptive automation is great!"
Is what earlier research relates.
But we find this domain,
In operational airplanes,
Is not implemented to date.

3.0 Dynamic Task Allocation in Operational Systems

The previous chapter focuses on how, when, and under whose command workload-balancing, adaptive dynamic task allocation can occur in human-automation systems. Specifically, it outlines the two fundamental alternatives: adaptive and adaptable automation. The literature is enthusiastic about adaptive automation, and studies have repeatedly claimed that it could serve as a great benefit to operators of complex systems. Despite this, there may still be gaps between what researchers are advocating and aircraft manufacturers are implementing. This chapter probes these gaps by looking at how, when, and under whose command the task structure *actually* changes in current operational systems. It deconstructs the approach and landing procedure in commercial aviation, specifically the Boeing 767, in order to uncover the ways in which tasks can be dynamically re-allocated between the pilot and the automation. After this in-depth investigation, this chapter discusses adaptive systems in newer commercial aircraft and identifies if they also feature workload-balancing dynamic task allocation, meaning that they can be classified as adaptive automation. These investigations find that modern commercial aviation employs essentially no degree of adaptive automation. This chapter then briefly discusses adaptive automation "associate systems" that were developed and tested in military aircraft and rotorcraft. However, these systems do not appear to have been implemented in operational vehicles. As this chapter discovers, aircraft manufacturers and airlines have made a deliberate decision not to leverage the benefits of adaptive automation as described in the literature. This chapter concludes by hypothesizing reasons why this is the case.

3.1 Dynamic Task Allocation in Approach and Landing in Commercial Aviation

Approach and landing in commercial aviation was selected as the in-depth case study for this chapter because it represents a prominent industry – with over 70,000 flights per day in the U.S. alone (*A Review of FAA's Efforts*, 2012) – that could potentially benefit from the implementation of adaptive automation. Furthermore, there is information publically available about both normal day-to-day operations and the causes of incidents and accidents in commercial aviation. Approach and landing is a high-workload phase of flight where a number of tasks must be properly executed in sequence, even under normal conditions. Any air traffic control mandated changes in the flight plan, inclement weather, or subsystem malfunctions can increase pilots' workload even further.

In this investigation, the normal Category I instrument approach and landing procedures have been detailed through a hierarchical task analysis (Annett, 2004; Stanton, 2006). A hierarchical task analysis (HTA) takes a specific task and its associated high-level goal (e.g. "safely land the aircraft") and decomposes it into subtasks and subgoals (e.g. "complete approach preparations"). These subtasks can be decomposed into their own subtasks, a process which continues until a sufficient level of detail has been reached for the desired analysis. The HTA of normal approach and landing presented in this chapter has been developed based on the manufacturer's recommended procedures, as described in manuals for the

Boeing 767 aircraft (Boeing, 2003a; Boeing, 2003b). The details are specific to this aircraft, but the same general procedure is used across many aircraft types when flying instrument approaches.

The Boeing 767 has a variety of lower-level automation systems that determine handling qualities and provide engine control, speed protection, and load relief. Other lower-level automation systems detect and alert when subsystem states are non-normal. However, these are not normally configured by the operators, and hence are not considered here. When pilots fly manually they normally refer to an autopilot-driven "attitude flight director" display, but their visual attention is largely determined by the manual flying task.

The HTA of approach and landing shows thirteen high-level goals that must be accomplished in order for the aircraft to land safely. Figures 3.1 through 3.3 show this HTA in three parts. The agent in charge of the task (the human or the automation) is noted by the color of each task. Red tasks are always completed by one or both of the two pilots, and blue tasks are always completed by the automation. Purple tasks can be accomplished by either agent, depending on whether or not the automation is engaged. Lastly, green tasks are those high-level goals that the automation can accomplish only with assistance from one of the pilots. The border of each task box notes if the task is always necessary (solid border), or only necessary if the automation is in use (dashed border). The latter are primarily automation setup tasks, in which the pilots set the parameters for an action (e.g. changing altitude) and the automation subsequently undertakes these actions.

This HTA shows that the automation is not required for a safe landing under normal conditions. However, its use certainly makes the entire process easier, more precise, and more consistent, particularly under actual instrument conditions. The automation can be used to intercept the localizer (Task 4) and to descend past a series of waypoints with altitude constraints to the glide slope intercept at the final approach fix (Task 5). To accomplish these tasks with the automation, one of the pilots sets the heading, altitude, and speed as prescribed by the approach procedure or as directed by air traffic control (Tasks 3.1, 4.1, 5.1, and 6.1). To change the heading, the pilot engages the "heading select" or "localizer capture" mode, and to change the altitude he engages the "flight level change" or "approach" (glide slope and localizer tracking) mode. The automation will automatically retard the throttles and initiate a descent while maintaining the speed set by the pilot.

Once established on the glideslope and localizer, the pilots lower the landing gear (Task 9) and deploy the flaps as the aircraft slows towards the final approach speed (Task 12.1). If the pilots wish to land under manual control, they disconnect the autopilot (Task 11.5) at the decision altitude of a few hundred feet (as specified by the approach procedure), flare (Task 12.4), touch down, and roll out (Task 13). At certain runways, if the crew has sufficient training and recency, they may leave the automation engaged and monitor an automatic flare, rollout centerline tracking, and autobraking, even under very reduced visibility. In this case, their only manual tasks are to engage and disengage the thrust reversers (Tasks 13.2 and 13.6) and disengage the autobrakes (Task 13.7).

If the pilots wish to regain manual control at any time, they can switch the automation off and fly manually with or without the flight director and autothrottle engaged, whichever they prefer. When the automation is engaged, it is the pilots' responsibility to ensure that it is functioning correctly and to decide whether or not to assume control by disengaging the automation. Even if the automation is in control of the entire approach down to minimums, there are still tasks that must be accomplished by the pilots – arming the autobrakes, programming in the airspeed and lowering the landing gear and flaps.

Most importantly, the autoflight automation never decides itself to take over a task, nor does it even suggest that the pilots might want to re-allocate tasks. The automation has essentially no part of the decision authority. This is true even in high-workload situations, like when an unanticipated event occurs. One such event is a go-around and transition into a missed approach (Flight Safety Foundation, 1998a; Flight Safety Foundation, 1998b; Boeing, 2003a). The pilots are required to initiate this procedure if the aircraft is not well established on the localizer and glide slope in the final approach configuration and speed by the minimum stable approach altitude, or if the weather obscures the runway when reaching a predetermined decision altitude. In a go-around, the aircraft must be pitched up and the thrust increased in order to gain altitude. The automation can be pre-programmed to complete both of these tasks once a single "go around" button is pushed. However, the automation never decides to initiate a go-around itself. The pilots are required to push the button to allow the automation to switch from approach mode to go-around mode.

3.2 Adaptive Systems in Commercial Aviation

Newer aircraft like the Boeing 777 and 787 or the Airbus A350 and A380 feature ***adaptive systems*** in which the automation can sense and react to the state of the vehicle, environment, and operator. These systems inherently fulfill one part of the three-part definition of adaptive automation (Section 2.1.1). This section discusses whether four of these adaptive systems fulfill the other two parts of the definition: re-allocating tasks between the automation and the pilots in order to balance the operator's workload. The systems investigated are: 1) flap active load relief, 2) flight envelope protection, 3) aircraft monitoring systems and electronic checklists, and 4) adaptive interfaces.

3.2.1 *Flap Active Load Relief*

Aircraft flaps can be extended at a number of positions (e.g. 5-25° in 5° increments). Each position is designed to be used within a certain airspeed range, which is indicated on the primary flight display. If the airspeed limit is exceeded for a particular flap setting, the automation automatically retracts the flaps to an appropriate setting (Boeing, 2011). This prevents damage to the flaps due to excessive aerodynamic loads. When the airspeed decreases to an appropriate level, the flaps are re-extended to the commanded setting. If the airspeed does not decrease, the automation prevents pilots from manually extending the flaps.

In flap active load relief there is a dynamic re-allocation of the flap extension and retraction task from the pilots (who perform the task under normal conditions) to the automation. However, active flap load relief cannot be considered adaptive automation because it is not workload balancing. The aim is to protect the aircraft with a temporary safety precaution, and not to reduce operator workload or increase operator performance. The flap extension/retraction task is re-allocated back to the pilots as soon as the aircraft reaches an appropriate airspeed for the commanded flap setting, even if their mental workload is above the desired level.

3.2.2 *Flight Envelope Protection*

Flight envelope protection systems ensure that an aircraft operates within its performance limits. In the Boeing 777, there are three types of flight envelope protection (Boeing, 2011). Stall protection limits the pitch up at the minimum maneuvering speed to prevent the aircraft from stalling and losing lift. Overspeed protection limits the pitch down at the maximum maneuvering speed to prevent damage to the airframe or control surfaces. Both stall protection and overspeed protection are forms of pitch envelope protection. The third type of flight envelope protection is roll envelope bank angle protection. This limits the amount that an aircraft can roll to prevent stall. All of these flight envelope protection systems are possible because of fly-by-wire systems, in which the pilots control the aircraft through an intermediary flight control computer. The flight control computer can modify the aircraft's handling qualities during flight to prevent it from exceeding its performance limits.

The two major manufacturers of commercial aircraft, Boeing and Airbus, employ flight envelope protection in different ways (Abbott, 2001). Boeing employs "soft" flight envelope protection, where the pilots face increasing resistance from the control column when they approach the aircraft's limits. However, with sufficient force the pilots can exceed the flight envelope, which may be necessary for upset recovery or to avoid an imminent mid-air collision. Boeing flight envelope protection cannot be considered adaptive automation because tasks are not being re-allocated between the pilot and the automation. Even as the aircraft approaches the edge of the flight envelope, the task of controlling the aircraft attitude remains with the pilots. The automation never takes over this task; instead, it changes how the pilots' control inputs translate to control surface deflections. As a result, Boeing flight envelope protection can be considered ***adaptive control***, rather than adaptive automation. Flight envelope protection is only one example of adaptive control; there are a number of other general ways in which this concept can be implemented (Sheridan, 2011). Beyond commercial aviation, adaptive control has also been designed and tested for rotorcraft (Fletcher et al., 2008; Warwick, 2009) and proposed for lunar landing (Duda et al., 2010).

Different from Boeing, Airbus employs "hard" flight envelope protection, which will not allow the pilots to exceed the flight envelope. When the pilots push an Airbus aircraft to the limits of its normal flight envelope, their control inputs are ignored and the automation temporarily takes over the flying task from the pilots. However, as with flap active load relief, this dynamic task allocation is not workload

balancing. The intent is not to reduce pilot workload, but merely to keep the aircraft safe. As a result, the flying task is re-allocated back to the pilots as soon as the aircraft moves away from the edge of the flight envelope. The pilots also have the ability to turn the flight control computer off at any time, after which they can exceed the flight envelope.

3.2.3 *Aircraft Monitoring Systems and Electronic Checklists*

The Airbus Electronic Centralized Aircraft Monitoring (ECAM) system and the Boeing Engine Indicating and Crew Alerting System (EICAS) use a network of sensors to diagnose problems with the aircraft (Boeing, 2011; de Crespigny, 2012). Whenever there is a problem, the aircraft monitoring system can detect the problem, alert the pilots, and display the appropriate non-normal checklists. Since the aircraft monitoring system knows the state of the aircraft and the switches in the cockpit, these electronic checklists are specific to the current state. The aircraft monitoring systems also state the operational consequences of the non-normal condition for the remainder of the flight.

There is an inherent change in the task structure whenever there is a non-normal event. A number of new tasks appear, and all are of high importance. The automation performs a number of these tasks, identifying the problem and determining the response, without being asked by the pilots. This is not adaptive automation, however, because the automation does not take over tasks that the operator used to perform. Since the automation always performs this information acquisition and analysis task when a non-normal event occurs, the task allocation is fixed instead of being dynamically re-allocated. If the pilots turn the flying task over to the autopilot to reduce their overall workload and allow them to focus on the non-normal checklist, this would be an example of adaptable automation. But because the pilots are the ultimate authority – particularly during non-normal events – the automation never decides on its own to take over the flying task.

3.2.4 *Adaptive Interfaces*

The electronic non-normal checklists described in the previous section are an example of an ***adaptive interface***. When the state of the aircraft changes, the display changes in response to present only the most important information to the pilot. These adaptive interfaces are attractive because modern technology allows for the potential of presenting too much information to the pilot (Federal Aviation Administration Flight Deck Automation Working Group, 2013). This information overload has the potential to overwhelm the pilot, particularly during critical phases of flight. Adaptive interfaces can automatically present the most-relevant and -critical information to the pilots, de-cluttering the displays.

Researchers have investigated adaptive interfaces in numerous systems (see Karwowski et al., 2006 for a review). One example is the response of a military aircraft's head-up display (HUD) to an unusual attitude (Newman, 1995). When the pitch and bank limits are exceeded, the HUD automatically deletes all but the minimum amount of information from the display. This adaptive response is designed to help pilots recover from the unusual attitude as quickly as possible.

Many of these adaptive interfaces employ the same triggering methods as adaptive automation to measure the state of the system, environment, and/or operator (Section 2.1.2). As a result, adaptive interfaces are often equated with adaptive automation in the literature. However, this is not necessarily the case. The two are separate concepts that can occur together or separately. For example, it is not adaptive automation if the display changes but the pilot continues to perform the same tasks. Adaptive automation can also occur without the display changing. It is important for future research to distinguish between adaptive interfaces and adaptive automation by describing whether it is the underlying tasks and/or the display that are changing.

3.3 Associate Systems in Military Aviation

The prior analyses have all focused on commercial aviation because the information is publically available. Information regarding the current implementation of adaptive automation in military aircraft is more difficult to obtain. One set of available information details research projects that were undertaken in the late 1980s and 1990s to develop ***associate systems*** for military aircraft and rotorcraft: the Pilot's Associate (PA) and Rotorcraft Pilot's Associate (RPA) (Banks & Lizza, 1991; Miller et al., 1999; Miller & Funk, 2001; Smith & Geddes, 2003; Inagaki, 2003; Miller & Dorneich, 2006; Hammer, 2009). Associate systems are defined as "collections of intelligent aiding systems that, collectively, exhibit the behavior of a capable human" (Miller et al., 1999, p. 1321). The PA and RPA monitored the environment for new tasks and then either allocated these new tasks to the pilot or handled them itself based on pre-approved schemes. These systems also reconfigured the displays based on an estimate of the pilot's intentions. The PA and RPA were developed and tested in simulations and actual flight vehicles. The RPA program, for example, concluded with twelve months of flight testing (Robertson, 2000). However, as far as this author knows, neither associate system has been implemented in operational aircraft (Smith & Geddes, 2003). Miller & Dorneich (2006) suggest that an associate system was incorporated into the F-22 Raptor, but detailed and current information was unavailable.

3.4 Conclusion

Much of the existing literature suggests that employing some degree of adaptive automation in a complex aerospace system could help to keep the operators' workload at a manageable, intermediate level. This would help to preserve situation awareness and manual flying performance. However, this chapter has shown that current aerospace systems do not fully implement adaptive automation. The hierarchical task analysis of 767 approach and landing finds that most of the time pilots must tell the automation what tasks to perform. This is adaptable automation. The analysis of adaptive systems in newer commercial aircraft shows that these systems fulfill only part of the definition of adaptive automation. They fail to be workload balancing and/or do not involve the dynamic re-allocation of tasks. Finally, while adaptive automation has been developed and tested for military aircraft and rotorcraft in the form of the Pilot's Associate and Rotorcraft Pilot's Associate, these associate systems have not been implemented operationally to the best of the author's knowledge.

Why is it, then, that aircraft manufacturers and airlines have made a deliberate decision to not leverage the benefits of adaptive automation as described in the literature? It is believed that this is due to three primary factors: 1) difficulties in fully testing and evaluating adaptive automation, 2) manufacturers' desire to make the pilot the ultimate authority, and 3) the reluctance of operators to trust adaptive automation.

The PA and RPA programs demonstrated the difficulty in fully evaluating adaptive automation systems (Smith & Geddes, 2003). Before adaptive automation is implemented in operational aircraft, it must be evaluated in all possible scenarios to assess that it is performing correctly and benefitting the operator. This requires a high-fidelity environment (either a simulation or a physical test vehicle) and a human test subject or a high-fidelity human performance model. The high-fidelity environment can make it difficult to repeat suspect automation behaviors and define and collect appropriate operator performance metrics. Without fully assessing adaptive automation, it will likely not be approved by the Federal Aviation Administration and other government bodies for implementation in operational systems. Future research should work to develop physical and virtual test platforms that allow for these assessments.

As was described in Section 2.1.1, Airbus and Boeing have automation policies that state the pilot is the ultimate responsibility for system safety (Abbott, 2001). The simplest way to ensure this is to implement adaptable automation in the cockpit, as has been done. This way, the pilot can quickly take over when the aircraft or the automation exhibits an unexpected behavior. As the development of adaptive automation sensors and algorithms advances and they become more predictable, manufacturers can give the automation a greater share of the decision authority. Even if the automation gains more authority, manufacturers can keep the ultimate responsibility for system safety with the pilot by implementing a form of form of shared decision authority such as management by consent or exception (Section 2.1.1).

When the PA and RPA programs were first implemented, the test pilots involved were initially very hesitant to trust the associate systems (Smith & Geddes, 2003). As Smith and Geddes state, "It was widely felt by the pilots that no machine could produce anything other than weak or outright silly advice" (p. 669). With time, the pilots' trust increased and they found the associate systems to be helpful. It is expected that these issues with user trust will be compounded if adaptive automation is proposed for commercial aviation. Convincing orders of magnitude more pilots from different cultures and backgrounds to accept adaptive automation is likely to be difficult. One area of future research that will help to facilitate pilot acceptance is increasing the sensitivity and specificity of adaptive automation triggers (Section 6.2.1). As sensitivity and specificity increase, so will the predictability of the adaptive automation and pilot trust in it.

The remaining chapters of this thesis address one of these barriers to the implementation of adaptive automation: assessing the impact that dynamic task allocation has on human operators. Chapter 4 experimentally investigates how control mode transitions during simulated lunar landing affect subjects'

visual attention, primary flying task performance, mental workload, and situation awareness. Chapter 5 develops an integrated human-vehicle model that can be used to simulate these behaviors across control mode transitions. Future models with a similar structure can allow researchers to perform longitudinal studies of system performance and discover unexpected emergent behaviors that warrant further human subject testing in physical test platforms.

3.4.1 *Limitations and Future Work*

The primary limitation of this chapter is its scope, in that it primarily focuses on commercial aviation. There may be other aerospace systems – lunar landing, the control of one and multiple UAVs, air traffic control, and military command and control, for example– that may currently employ some degree of adaptive automation. Future work should investigate operational procedures and systems in these domains, as well as adaptive automation-focused research projects similar to the Pilots' Associate and Rotorcraft Pilots' Associate. If any of these operational systems employ adaptive automation, future research should seek to clarify the technical and social factors that fostered this implementation. It is also important for future research to consider the domains that do not feature adaptive automation. In particular, analyzing adaptive automation research projects that have not moved to implementation can help to further highlight the challenges and research gaps that still need to be addressed.

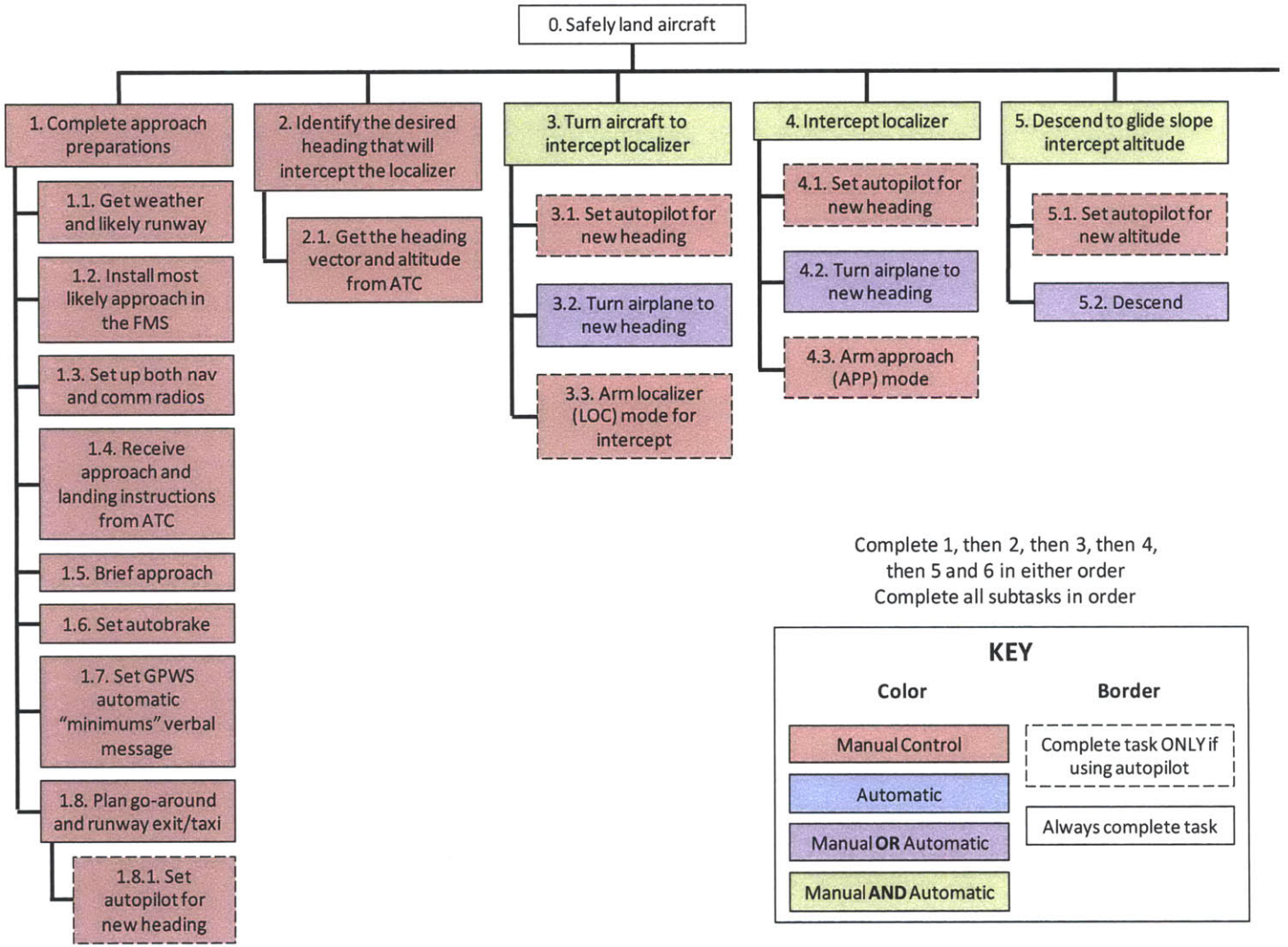


Figure 3.1. Hierarchical Task Analysis of Normal 767 Approach and Landing, Part 1

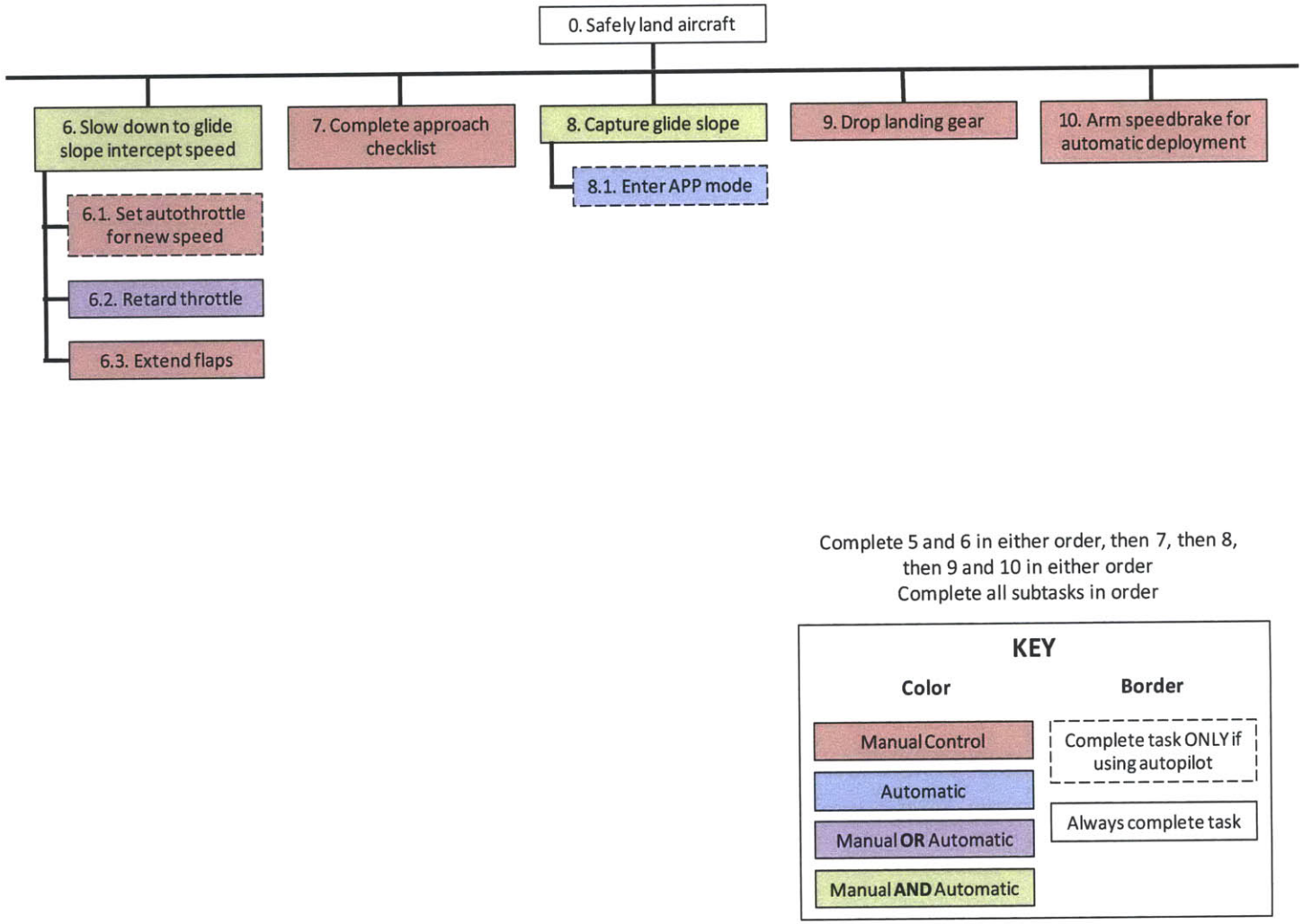


Figure 3.2. Hierarchical Task Analysis of Normal 767 Approach and Landing, Part 2

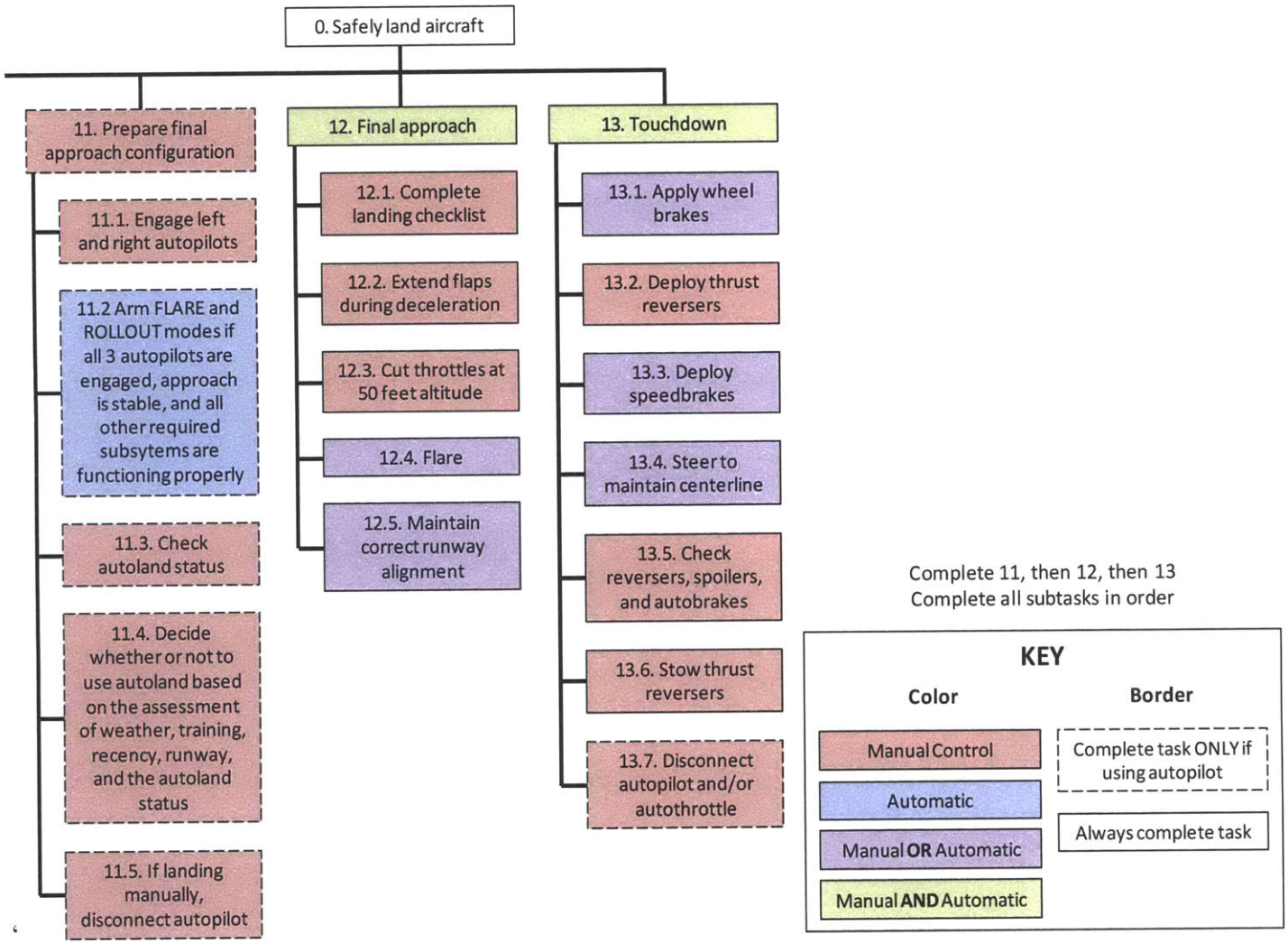


Figure 3.3. Hierarchical Task Analysis of Normal 767 Approach and Landing, Part 3

The subjects were told of their mission:
To fly through control mode transitions.
The data desired,
An eye tracker required;
Attention was the acquisition.

4.0 Pilot Visual Fixations on Flight Displays across Lunar Landing Mode Transitions

The previous two chapters discuss workload-balancing dynamic task allocation in the form of adaptive and adaptable automation. While much of the existing literature has focused on investigating and enabling adaptive automation, operational aerospace systems do not fully implement this concept. This chapter helps to fill one of the gaps between research and implementation: understanding and evaluating how operators respond to the re-allocation of tasks. Specifically, this chapter presents the results of an experimental investigation into how subjects' visual attention, flying performance, mental workload, and situation awareness changed across control mode transitions in simulated lunar landing. Understanding how operators respond to dynamic task re-allocations can help to improve the design of automation, displays, and cockpit procedures so that operators can transition "gracefully" – preserving performance and situation awareness while minimizing workload (Hainley et al., 2013).

4.1 Background

An operator's visual attention in the cockpit is driven by both bottom-up and top-down processes (Wickens et al., 2003). From the bottom-up, the likelihood of fixating on a particular instrument increases as its salience (e.g. color and luminance relative to the surrounding instruments) increases and the distance between it and other instruments decreases. Top-down factors include the expected frequency of events involving a particular instrument (called the "expectancy"), and the value of that information to the current task structure. The expectancy of information is analogous to the instrument's bandwidth – how frequently it needs to be sampled – which is a function of how quickly it is changing. Experience is another top-down factor that influences the operator's strategy for visually scanning the instruments (Bellenkes et al., 1997; Hayashi, 2003; Huemer et al., 2005). Bellenkes et al. (1997) hypothesized that experts have different scan patterns than novices because they are better at extracting information from displays, constructing a mental model of the vehicle behavior, and changing their scan pattern when the situation dictates.

A change in the phase of flight, a failure, or a control mode transition may all modify the tasks performed by the operator. In response, operators must re-allocate their limited capacity of attention resources (Wickens, 1980; Wickens & Hollands, 2000). Visual attention becomes the major constraint when the level of automation decreases across a mode transition and the operator gains more tasks perform. If the cumulative task workload increases to a level where there are insufficient attention resources for all tasks, operators must shed attention on lower-priority tasks in favor of higher-priority ones. If the sampling frequency of an instrument decreases below the minimum required for satisfactory performance, given the process bandwidth and the duration of the task, performance will obviously decrease. However, there will be other effects. As the time between fixations increases, the uncertainty in

the operator's estimate of these vehicle states also increases. This is noted as a decrease in the operator's situation awareness of these states. In order to retain good situation awareness of the high-priority tasks, the operator has been forced to reduce attention, and therefore situation awareness, on the lower-priority tasks.

Mental workload is the amount of cognitive resources required to perform a set of particular tasks (O'Donnell & Eggemeier, 1986). It can be measured by asking the operator to give subjective assessments using rating scales, by taking physiological measurements that correspond to cognitive effort, or by measuring the decrement in performance on a primary or secondary task (see Section 2.1.2 and Cain, 2007 for reviews). Secondary task measures are more sensitive to workload changes when both the primary and secondary tasks compete for the same sensory resources (Wickens, 1980; Wickens & Hollands, 2000). When the primary task requires more attention than is currently unused, attention must be taken from the secondary task. If attention on the secondary task is reduced below the minimum required for satisfactory performance, there will be a measurable effect. The experiment described in this chapter employed an embedded two-choice secondary task – requiring the subject to extinguish a cockpit light (the "comm light") as the acknowledgement of receipt of a data communication – to infer cognitive workload while preserving face validity (Lowenthal, 2012; Hainley et al., 2013; Kaderka, 2014).

Situation awareness may be measured with subjective rating scales employed by the operator (e.g., Situation Awareness Rating Technique, SART (Taylor, 1990)) or by experts (e.g., SALSA (Hauss & Eyferth, 2003)). More often, objective techniques are used that assess the operator's answer to questions about the system state while the simulation continues (e.g., Situation Present Assessment Method, SPAM (Durso et al., 1995)) or is frozen and blanked (e.g., Situation Awareness Global Assessment Technique, SAGAT (Endsley, 1988)). There is concern that these techniques are intrusive and distracting, and that they test memory rather than situation awareness. The experiment described in this chapter assesses situation awareness using a tertiary task developed by Hainley et al. (2013) – making verbal call-outs of the vehicle state at pre-determined intervals. Situation awareness is assumed to increase as the number of missed call-outs decreases, and vice versa. This technique provides an unintrusive real-time measure of situation awareness with high face validity, and has been used in other prior experiments (Kaderka, 2014).

Hainley et al. (2013), which was supported by the same project as this thesis, performed a similar experiment and found that manual control reversions increased subjects' mental workload and decreased the measure of situation awareness in proportion to the number of manually-controlled vehicle axes during simulated lunar landings. However, this experiment did not measure visual fixations or ask subjects to perform control mode transitions from low→high levels of automation. This experiment closes both of these research gaps.

4.2 Methodology

4.2.1 *Lunar Landing Simulator Displays*

The lunar landing simulator display was located across two computer screens separated by a 1.4 in (3.6 cm) gap (Figure 4). The primary flight display (PFD) was shown on the left screen and the horizontal situation display (HSD) was shown on the right screen. The center of the displays was located approximately 2 ft (61 cm) from the subjects' eyes.

The PFD depicted the vehicle roll and pitch in multiple ways: the brown/blue terrain/sky horizon, the pitch ladder, and the roll indicator. The yellow "waterline" indicator represented the simulated vehicle's current pitch and roll state, and the magenta needles served as an attitude flight director indicating the attitude required to converge with the optimal reference trajectory. This trajectory brought the vehicle to a hover 150 ft (46 m) above the selected landing aim point (LAP) (Bilimoria, 2009; Stimpson, 2011; Hainley et al., 2013). The PFD displayed the vehicle's remaining fuel, altitude, and rate-of-descent (RoD) through analog and digital displays, as shown in Figure 4. A RoD reference trajectory was indicated with a magenta fly-to triangle on that instrument. The pre- and post-transition control modes were displayed above the attitude and altitude indicators, with the current mode highlighted.

The vehicle position was represented on the HSD by the stationary white crosshairs over the moving digital elevation map (DEM). Pictures of the DEMs used in the experiment can be seen in Appendix A. The shades of grey and contour lines on the DEM indicated lunar terrain of different altitudes. Subjects were told the red square symbolized a prior radar "scan area," on which hazards and LAPs had been identified. Subjects could overfly the red hazard areas but not land in them. All trials began targeting the center of the scan area as the LAP. In some trials, the LAP was automatically redesignated to one of the alternate LAPs (marked with cyan circles on the HSD).

The comm light – a small, rectangular box containing two circular lights – was located in the bottom-right corner of the HSD. When either of these lights illuminated, subjects responded and the speed and reliability of their responses provided a measure of their mental workload. This procedure is described in detail in Section 4.2.4.

Five areas of interest (AOIs) were identified around the fuel indicator, attitude indicator, altitude indicator, RoD indicator, and comm light. Each task accomplished by the operator was assumed to require foveal vision: errors between the actual and flight director specified attitude and RoD were not identifiable in peripheral vision unless they were very large, the fuel and altitude call-out tasks required the subjects to read the respective instruments, and the comm light illuminated with blue and green colors that were very close to the turquoise background color of this instrument. The size of foveal vision is approximately 2° visual angle, and so each AOI encompassed 1° visual angle margin around the instrument, as based on the reference position of subjects' heads. All instruments were separated by at least 2° visual angle to ensure that the AOIs did not overlap.

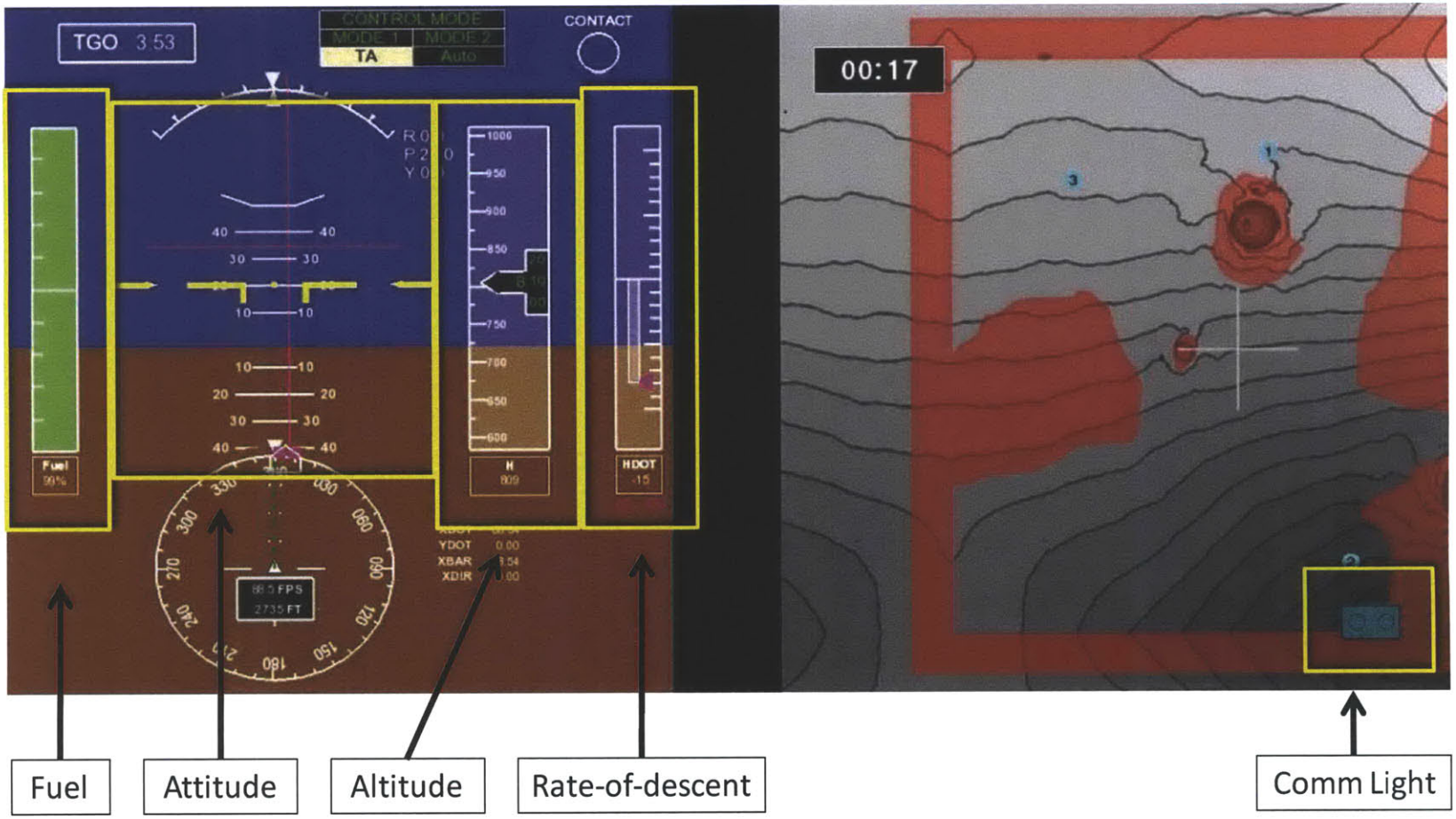


Figure 4. Lunar landing simulator displays (Stimpson, 2011). The black bar in the middle represents the gap between two separate computer screens, with the primary flight display on the left screen and the horizontal situation display on the right screen. The regions highlighted in yellow are the areas of interest for the five instruments involved in the experimental tasks.

In this thesis, the attitude and RoD indicators are referred to as the "flight instruments" because the subjects used them for the manual control flying tasks. The other instruments – the comm light, fuel indicator, and altitude indicator – are referred to as the "secondary instruments." This is because they were not required for the flying task, which subjects were told was their primary responsibility.

4.2.2 *Lunar Landing Simulator Controls, Dynamics, and Initial Conditions*

The experimental scenario was designed to replicate the landing phase of the Apollo mission design. Each trial began at an altitude of 820 ft (250 m) above the lunar surface, 2800 ft (945 m) downrange of the initial landing aim point (LAP) located in the center of the scan area. The vehicle was pitched up 18°, with a horizontal offset of ±200 ft (±61 m) cross-range from the initial LAP. The forward velocity was 94 ft/s (29 m/s), and the descent rate was -16 ft/s (-5 m/s). These initial conditions required control inputs immediately after the trial began in order to match the current vehicle attitude and RoD to the guidance trajectory.

The vehicle dynamics approximated the Apollo lunar module dynamics as detailed by Bilimoria (2009). Attitude inputs made to the right-hand rotational hand controller (RHC) were rate command/attitude hold, which was used in the Apollo lunar module (Duda et al., 2009). This meant that subjects could maintain a specified attitude without having to maintain a constant deflection of the RHC. When the RHC was returned to the neutral position, the attitude control computer would maintain the vehicle attitude until the next control input. The maximum angular rate, corresponding to full deflection of the RHC, was 8.57°/s in pitch and 9.68°/s in roll. There was a ≤1 s delay until this maximum angular rate was achieved, with the exact value depending on the joystick deflection (Appendix E). The RoD was commanded in ±1 ft/s (±0.3 m/s) increments by a throttle button located under the subject's left index finger. The actual RoD lagged behind the commanded RoD with a time constant[2] of ≈1 s. The vehicle could be operated in three distinct control modes: 1) Two-axis (pitch and roll) and rate-of-descent manual control (TA+RoD), 2) Two-axis manual control with automatic rate-of-descent control (TA), and 3) Automatic control of attitude and rate-of-descent (Auto). The difficulty of performing the required experimental tasks in a particular control mode was proportional to the number of manually-controlled loops.

4.2.3 *Recording Visual Attention*

Subjects' visual fixations were recorded using an eye tracker (Tobii Model x50; Stockholm, Sweden) mounted beneath the two simulator displays. The positioning of the tracker camera beneath the simulator displays allowed for gaze angles up to 40° visual angle above, left, and right of the camera (Tobii Technology, 2006). This encompassed all instruments on the display. Subjects were not required to keep their head fixed in space. Subjects' chairs were repositioned so their eyes were 2 ft (61 cm) from the eye

[2] The time for the actual RoD to reach 63.2% of the commanded value.

tracker and centered within the camera's range of vision. In this position, subjects were still able to comfortably reach the simulator controls.

Gaze position relative to the screen was almost always sampled at a rate of 50 Hz (Tobii Technology, 2006). If the eye tracker lost track of a subject's eyes due to rapid head or eye movements, the sampling rate would temporarily drop to a minimum of 10 Hz. Consecutive gaze points that occurred within a 50-pixel diameter – which corresponds to 2° visual angle, the size of foveal vision – for ≥100 ms were grouped together as a single fixation. The selection of this fixation duration was based on prior literature (Manor and Gordon, 2003; Blignaut, 2009). The x50 eye tracker specified accuracy of 1° across large head movements and over long periods of time (Tobii Technology, 2006).

At the beginning of each trial, subjects performed the x50 eye tracker's native 9-point calibration routine. This calibration left substantial, yet consistent, errors and so a secondary calibration was developed and applied to all fixations. Subjects fixated on each of the 9 calibration points five times, and the average horizontal and vertical errors about each point were determined. These nine "corrections" were applied *post-hoc* to each subsequent fixation, weighted by the inverse square distance between the fixation and the particular calibration point. After both calibrations, subjects performed a validation procedure. They looked at instruments on the display repeatedly, and the data was inspected to make sure the fixations fell within the area of interest (AOI) they were actually looking at. All subjects satisfied this validation step and proceeded to the training and collection of data. Another validation procedure was performed after the data-collection trials to ensure that any drift in the eye tracker measurements was minimal. All subjects satisfied this post-experiment validation requirement.

4.2.4 *Subjects and Experimental Protocol*

Twelve volunteer subjects (7 males, 5 females; average age 23.8±1 (mean ± standard error, throughout unless noted) participated in the study. All subjects completed the full experiment and were compensated for their time. Three subjects wore contact lenses for vision correction, but this did not impede visual fixation data collection. Seven subjects had no experience with any flight simulator, four had some experience, and one held a private pilot's license with instrument ground school training. The experimental protocol was approved by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. Further details about the subject demographics can be seen in Appendix A.

Each experimental trial began in one of the three control modes with the automated guidance targeted on the center of the scan area as the landing aim point (LAP). After 20 seconds a tone cued subjects to press a button, manually accepting a transition to one of the other two control modes. Subjects were told what the beginning and ending control mode of each trial would be before the trial began. Furthermore, the modes were displayed during the trial on a mode annunciator panel at the top of the primary flight display. There were six possible mode transitions (MTs), as seen in Table 1. In half of the trials, there was a landing point redesignation (LPR) in which the automated guidance changed the LAP.

The approach trajectory to this new LAP was also updated, which caused a step-change in the flight director guidance error. The agent flying – the subject or the autopilot – was required to nullify this error. For trials beginning in autopilot, or those beginning in manual control where subjects closely followed the reference trajectory, the average magnitude of this step-change in guidance was 6.2±0.04° in pitch and 7.0±0.03° in roll. Further details about the experimental test matrices can be seen in Appendix A.

Table 1. Subjects experienced six different control mode transitions during the simulated lunar landing experiment. It was expected that Mode Transitions 1-3 involved changing to a harder task, whereas the flying task became easier across Mode Transitions 4-6.

Direction of Change in LoA	Mode Transition	Starting Mode		Ending Mode	Change in # of Manual Control Loops
Low→High	1	TA	→	Auto	-2
	2	TA+RoD	→	Auto	-3
	3	TA+RoD	→	TA	-1
High→Low	4	Auto	→	TA	2
	5	Auto	→	TA+RoD	3
	6	TA	→	TA+RoD	1

In some trials, a system failure was introduced after ≥50 s of elapsed trial time. The trial ended when subjects correctly detected and diagnosed the failure. If the trial did not present a failure or if the subject did not detect the failure present, the trial ended after 80 s. This chapter focuses on subject behavior across the mode transitions, and not during failure detection and diagnosis. As a result, the data analysis considers only the first 50 s of each trial. For the analysis of subject failure detection and diagnosis, please refer to Kaderka (2014).

Mental workload and situation awareness were assessed with the same procedures used by Hainley et al. (2013). A two-choice visual secondary task was used to measure subject mental workload throughout each trial. At regular intervals, one of the circles on the comm light illuminated blue (always the left circle) or green (always the right) at a random time within the first 2 s of a 6-s window. Subjects were instructed to respond by pressing the corresponding left or right button on the rotational hand controller. The light extinguished when the subject responded with the correct button or after the 6-s window ended. The comm indicator illuminated 11 times in each trial, 8 of which were within the first 50 s of the trial.

Verbal call-outs of the vehicle's altitude, fuel remaining, and scan area landmarks at pre-determined intervals were used as a tertiary task to measure situation awareness. Subjects were told to report altitude every 100 ft (30.5 m) down to 500 ft (152 m), and at 50-ft (15-m) intervals thereafter. Fuel remaining was to be reported in 5% increments. Lastly, three scan-area landmarks were to be reported: 1) when the scan area came into view on the HSD, 2) when the selected LAP came into view on the HSD, and 3) when the

vehicle crossed over the edge of the scan area. Subjects were trained on the call-outs before the data collection trials began, and a placard was placed beside the displays to remind them of the call-outs during the trials. If subjects kept on the reference trajectory, there were 16 call-outs (8 altitude, 5 fuel, and 3 landmark) spread throughout the trial, with 10 (6 altitude, 4 fuel) in the first 50 s. The average time between consecutive call-outs of the same type in the first 50 s was 7.1±0.03 s for the altitude call-outs and 12.7±0.02 s for the fuel call-outs. Call-outs were scored as correct if they were verbalized within 1-2 s of their actual occurrence, as judged by the experimenter.

After the initial calibration trials described in Section 4.2.3, subjects were given standardized training to familiarize them with the simulator displays, controls, and tasks. They were explicitly instructed to give the flying task (control of the vehicle attitude and RoD) the highest priority, and to give the comm light response task a higher priority than the situation awareness call-outs. They were not told to prioritize one type of call-out over another. Subjects flew between 25-33 training trials until they felt comfortable and their performance on the experimental tasks had reached a satisfactory level (as judged by the experimenter). This was followed by two blocks of 18 data collection trials separated by a short break.

4.2.5 *Experimental Design and Data Analysis*

This experiment used a repeated-measures within-subjects design. There were two inter-trial independent variables: the 6 mode transitions (MTs) depicted in Table 1 and 2 levels of landing point redesignation (LPR), yes and no. Each of the 12 factor-level combinations of these two variables was repeated 3 times for a total of 36 trials. Trial order was pseudo-randomized so that the same ending control mode was not repeated in consecutive trials and the ages of the mode transitions within the experiment were approximately equal. The digital elevation map (3 different maps seen in Appendix A) and failure time window (3 levels: 50-55, 55-60, and 60-65 s after the MT) were randomized into the design using a Graeco-Latin square.

Each trial was divided into 6-second windows outward from the mode transition (Figure 5). The windows were then grouped into three phases: pre-, during-, and post-MT. As is stated later, the characteristics of the during-MT window set the 6-s window length across the trial. The beginning and ending time of each phase, and the number of 6-s windows in each phase (2 pre-MT, 1 during-MT, and 4 post-MT) can be seen in Figure 5. The pre-MT phase did not extend to the beginning of the trial to omit subjects' scan patterns during the time they spent nulling the initial flight director guidance errors. The during-MT phase lasted 6 seconds, because subjects spent this amount of time nulling the flight director guidance errors that occurred with an LPR. Furthermore, this phase contained the first comm light response, fuel call-out, and altitude call-out after the MT in nearly all trials (Comm Light: 97.2% of trials, Fuel: 98.8%, Altitude: 99.8%). Table 2 details the number of comm light illuminations and call-outs in each phase. The post-MT phase ended 30 s after the MT, which was the time of the earliest failure in any trial.

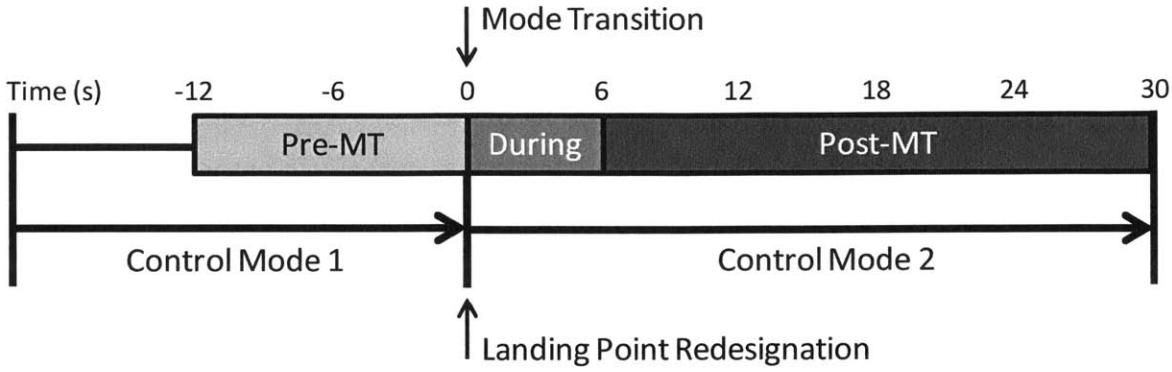


Figure 5. Each 50 s trial was partitioned into three phases based on the time of the mode transition.

Table 2. Number of comm light illuminations and call-outs in each phase.

Measurement	Phase			Total
	Pre-MT	During-MT	Post-MT	
Comm Light Illuminations	3	1	4	8
Altitude Call-outs	2	1	3	6
Fuel Call-outs	1	1	2	4

The analyses focused on the five instruments indicated in Figure 4: the two "flight instruments" (the attitude and RoD indicators) and three "secondary instruments" (the comm light, fuel indicator, and altitude indicator). Three metrics quantified subjects' visual attention on each instrument in each 6-s window. First, the percent of time spent on an instrument directly indicated how much attention subjects allocated to it. As a result, this metric, referred to as the "percent of attention," was the primary focus of the analyses. The two other metrics of attention – the number of fixations and the average dwell duration in each 6-s window – provided additional information about how subjects allocated their attention to a specific instrument.

Secondary task performance was quantified by two metrics: 1) the percent of correct comm light responses[3], and 2) for the correct responses, the time interval that elapsed between the illumination of the comm light and the response. Tertiary task performance was quantified by the percent of successful call-outs (call-outs verbalized within 1-2 s of their actual occurrence, as judged by the experimenter).

The visual fixation data was processed in MATLAB (MathWorks; Natick, MA) in preparation for analysis. These procedures involved applying the secondary calibration, identifying the area of interest (AOI) for each fixation, and grouping fixations into 6-s windows. If multiple consecutive fixations appeared on the same AOI, they were counted together as one longer fixation. This was done because it was assumed that subjects did not change the task they were performing if they looked at a different area

[3] An incorrect response occurred when the subject responded to the comm light while it was illuminated, but pushed the wrong color button. However, this occurred <1% of the time. Most of the time, subjects either responded correctly or missed the comm light completely.

of same AOI. Performance data on the comm light response task and call-outs was also processed in MATLAB. Thirty-three trials (7.6% of the data) were removed across all subjects due to insufficient eye tracker data, abnormal behaviors that were observed during the experiment (e.g. forgetting to control the RoD in TA+RoD), or flying performance that was an extreme outlier (beyond 3 times the interquartile range) from the subject's normal performance. Two subjects had no trials removed, and each of the other subjects had $\leq$6 trials removed. Statistical analyses on the remaining 399 trials were performed using SYSTAT (Systat Software; Chicago, IL). In all analyses significance was determined at a level of α=0.05. Post-hoc pairwise comparisons were Bonferroni-corrected.

Before statistical analyses were performed, the flying performance (mean square error between the actual and guidance-prescribed pitch, roll, and RoD) data was inspected for learning effects. Friedman tests were performed to see if subjects' steady-state mean square flying error (in the period 40-50 s after the trial commencement) decreased with later repetitions of the same experimental condition. No consistent trends were found across experimental conditions, which meant that subjects were sufficiently trained before the data-collection trials and their performance had plateaued. As a result, it was not necessary to include trial number in further analyses.

4.3 Hypotheses

4.3.1 *Attention across Mode Transitions*

As was stated in Section 4.1, subjects had a limited capacity of visual attention resources to allocate among all instruments in the display. Whenever workload increased to a level where there were insufficient resources for all tasks, subjects would need to sacrifice attention on lower-priority tasks in favor of higher-priority tasks. Controlling the attitude of the vehicle by nulling the flight director errors was a complex task that required frequent control inputs and attention from the subjects. As a result, it was expected that adding the primary flying task on top of the secondary and tertiary tasks would exceed their attention capacity. It was hypothesized that when transitioning from Auto to TA, subjects would decrease attention on the secondary instruments in order to increase attention on the flight instruments. This was expected to occur even without the presence of an LPR, an event that required additional attention on the attitude indicator immediately after the mode transition.

- ***Hypothesis 1: Subjects would increase attention on the attitude indicator and decrease attention on the fuel indicator, altitude indicator, and/or comm light across Auto→TA mode transitions without an LPR.***

Mode transitions between TA and TA+RoD in both directions were the focus of Hypothesis 2. Since flying the vehicle was the primary task, subjects were expected to give the attitude indicator enough attention for satisfactory control before and after these mode transitions. Any attention re-allocated to the RoD indicator across TA→TA+RoD mode transitions should have been taken from the secondary

instruments. Similarly, any attention removed from the RoD after TA+RoD→TA mode transitions should have been re-allocated to the secondary instruments. As with Hypothesis 1, these behaviors were expected to occur even when there was no LPR.

- ***Hypothesis 2.a: Subjects would not increase or decrease attention on the attitude indicator across TA→TA+RoD or TA+RoD→TA mode transitions without an LPR.***
- ***Hypothesis 2.b: Subjects would decrease attention on the fuel indicator, altitude indicator, and/or comm light across TA→TA+RoD mode transitions without an LPR.***
- ***Hypothesis 2.c: Subjects would increase attention on the fuel indicator, altitude indicator, and/or comm light across TA+RoD→TA mode transitions without an LPR.***

Hypotheses 1 and 2 are summarized in Table 3.

Table 3. Hypothesized changes in the percent of attention on instruments across particular mode transitions without an LPR.

	Change in attention across the MT		
Mode Transition	**Attitude Indicator**	**RoD Indicator**	**Secondary Instruments (Comm, Altitude, Fuel)**
Auto→TA	↑	=	↓
TA→TA+RoD	=	↑	↓
TA+RoD→TA	=	↓	↑

4.3.2 *Transient Effects after Mode Transitions*

The previous hypotheses addressed expected changes in subjects' visual scan patterns before and after a particular mode transition. The following hypotheses investigated how quickly subjects assumed these scan patterns after a mode transition, and if these scan patterns were constant throughout the post-MT phase.

It was expected that the presence of an LPR would cause transient effects in subjects' post-MT scan patterns. An LPR caused a step-change in the flight director guidance error, which subjects nulled when transitioning to a manual mode (TA or TA+RoD). This required subjects to temporarily allocate more attention to the attitude indicator. Since it was expected that subjects would already be working at their attention capacity in these manual control modes (per Hypothesis 1), they would not be able to increase attention on the attitude indicator without cost to their performance elsewhere. They would need to re-allocate attention away from the secondary instruments. Once the flight director guidance error was nulled, subjects would be able to decrease attention on the attitude indicator and increase attention on the attitude indicator, returning to their steady-state post-MT scan pattern. Without the presence of an LPR, subjects would be able to begin their steady-state scan pattern immediately after the mode transition.

- ***Hypothesis 3.a: When transitioning to a manual mode with an LPR, attention on the attitude indicator would temporarily increase immediately after the mode transition, and then decrease again in later windows.***
- ***Hypothesis 3.b: When transitioning to a manual mode with an LPR, attention on the fuel indicator, altitude indicator, and/or comm light would temporarily decrease immediately after the mode transition, and then increase again in later windows.***
- ***Hypothesis 3.c: When transitioning to a manual mode without an LPR, attention on the attitude indicator and the fuel indicator, altitude indicator, and/or comm would be constant across all windows after the mode transition.***

4.4 Results

4.4.1 *Mode Transitions from Auto→TA*

The first hypothesis predicted how subjects would re-allocate attention across Auto→TA mode transitions without an LPR. To test these claims, four mixed-model ANOVAs were constructed, one each for the attitude indicator, fuel indicator, altitude indicator, and comm light. The RoD indicator was not analyzed because it was not needed for either control mode. Each of these ANOVAs used the percent of attention on an instrument as the dependent variable, the subject as a random factor, and the phase (pre-, during, or post-MT) as a fixed factor. To satisfy assumptions about the normality and homoscedasticity of the ANOVA residuals, the percents of attention were transformed by the square root before being analyzed. Furthermore, windows with no attention on the particular instrument were excluded. When there were significant main effects (α=0.05), pairwise comparisons were conducted to see which phases differed. These pairwise comparisons were Bonferroni corrected within each ANOVA.

Figure 6 shows the data used in this Auto→TA analysis. The figure is a way of graphically characterizing the subjects' "visual attention budget" for each instrument by experimental phase. Similar figures, for each of the other mode transitions with and without an LPR, can be found in Appendix B. Included are the five instruments shown in Figure 4, as well as a group containing all other instruments in the simulator. Also included is the amount of attention "overhead" – the percent of time that subjects were not looking at any instrument. During this overhead time, subjects may have been moving their eyes from one instrument to another, looking at a space on the display outside of an instrument, or looking away from the screen. A mixed-model ANOVA (with the same factors and transformations as described in the preceding paragraph) finds no significant difference in the amount of overhead between phases ($F(2,178)=0.95$, $p=0.39$). This suggests that subjects did not change the total amount of attention on the instruments; instead, they re-partitioned it among the instruments.

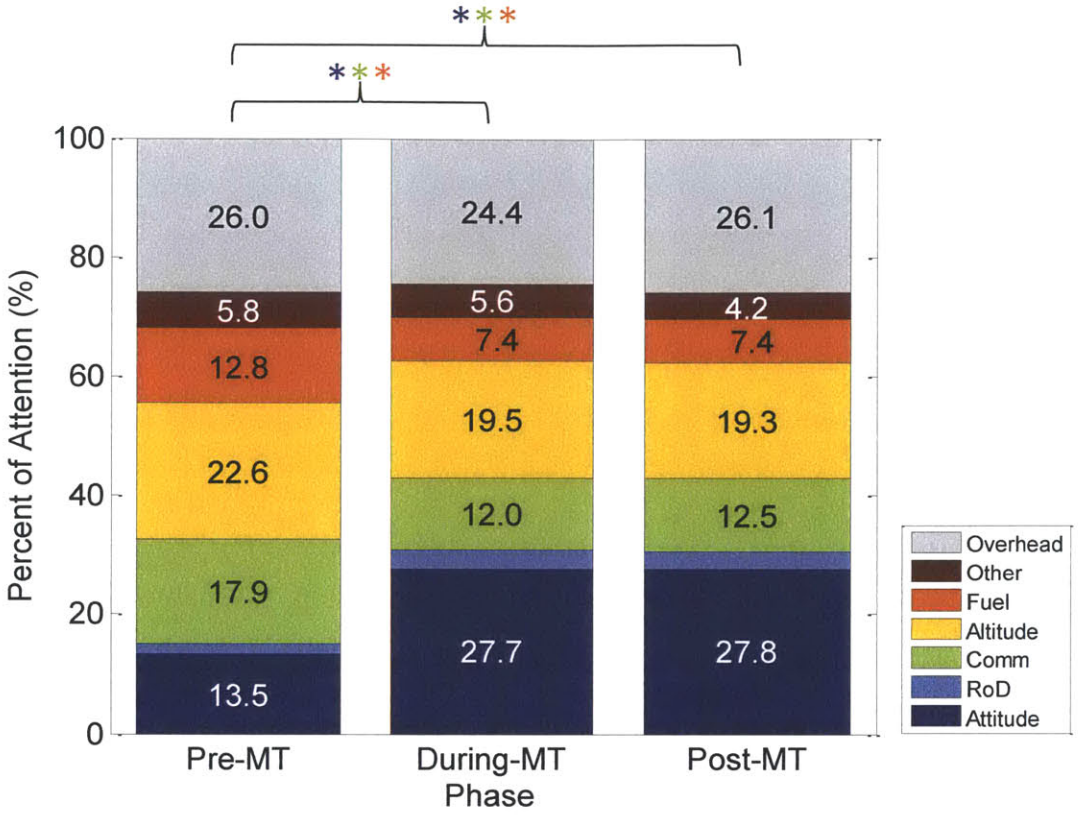


Figure 6. Percent of attention on each instrument by phase during Auto→TA mode transitions without a landing point redesignation. The instruments are listed in the legend in the same order as they appear in each stacked bar. Significant differences (α=0.05) were found between the pre-MT phase and each of the other two phases for the attitude indicator, comm light, and fuel indicator (marked with *).

Figure 7 shows the change in the percent of attention from the pre-MT phase to the during- and post-MT phases. There was a significant main effect of phase on the percent of attention on all instruments except the altitude indicator. (Attitude: $F(2,238)=44.4$, $p<0.0005$; Comm light: $F(2,214)=18.6$, $p<0.0005$; Altitude: $F(2,230)=2.82$, $p=0.062$; Fuel: $F(2,185)=13.5$, $p<0.0005$). Pairwise comparisons (indicated by the * symbols on Figure 7) showed that the percent of attention on the attitude indicator significantly increased across the Auto→TA mode transition. At the same time, the percents of attention on the fuel indicator and comm light significantly decreased across the mode transition. These findings supported Hypothesis 1.

It is surprising that subjects significantly decreased attention on the fuel and comm indicators but not the altitude indicator. Subjects were explicitly told during training that responding to the comm light was their secondary task, and that altitude and fuel call-outs were their tertiary task (with no priority specified between these two). So, why did subjects sacrifice attention on the secondary task before the tertiary task? And, why did subjects choose to sacrifice attention on the fuel indicator but not the altitude indicator?

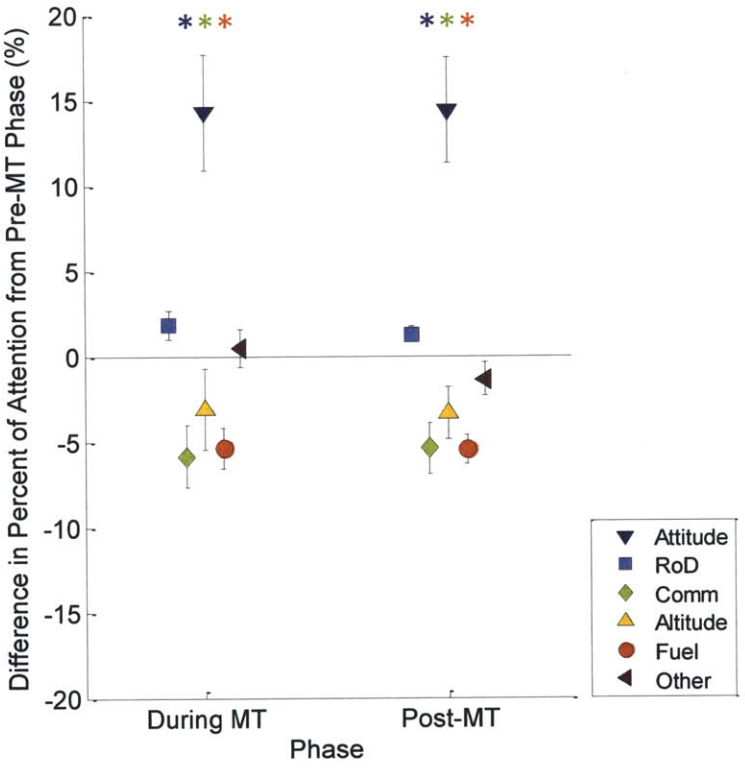


Figure 7. Change in percent of attention from the pre-MT phase to the phase noted on the x-axis, Auto→TA mode transitions without a landing point redesignation. Error bars indicate standard error. Significant differences (α=0.05) were found between the pre-MT phase and each of the other two phases for the attitude indicator, comm light, and fuel indicator (marked with *).

To answer these questions it is necessary to first see if the changes in attention on the comm light and fuel indicator corresponded to significant changes in the secondary and tertiary task performance. If attention changed across the Auto→TA mode transitions without significant changes in performance, it suggests that these instruments were oversampled during the Auto control mode.

As Figure 8 shows, the fuel call-out performance did significantly degrade over the mode transition (Friedman test for main effect of phase: p=0.003). In fact, performance on the fuel call-out during the mode transition was the worst across subjects. Subjects improved post-MT, but still consistently performed worse than during the pre-MT Auto mode. It is important to note that the numbers of call-outs that occurred in each phase were similar to one another but not identical (Table 2). Even so, averaging call-outs by phase does not grossly change the results.

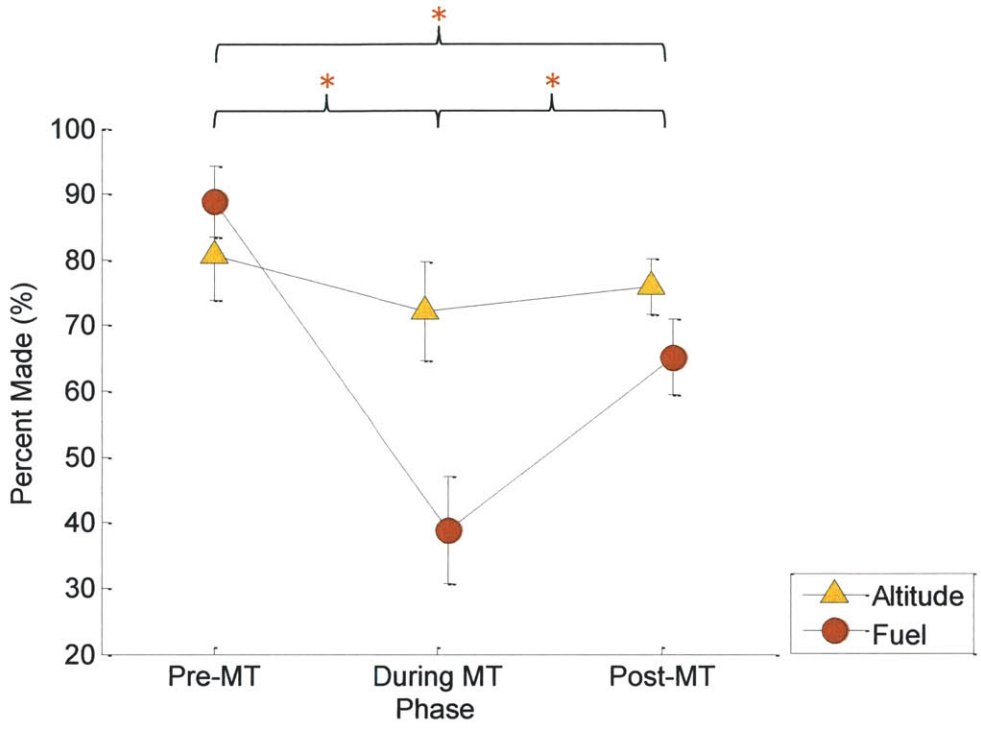


Figure 8. Performance on the fuel call-out decreased during the Auto→TA mode transition and then recovered, but not to the pre-MT performance. Error bars indicate standard error. Significant differences (α=0.05) between phases were found for the fuel indicator, and are marked with *.

Subjects may have chosen to shed the fuel call-out instead of the altitude call-out for two reasons: their timing and their location on the display. On the nominal trajectory fuel call-outs occurred approximately every 12.7±0.02 s, while the altitude callouts occurred approximately twice as often – every 7.1±0.03 s. It is likely that subjects gave priority to the call-outs that need to be made more frequently. This is consistent with the notion of expectancy, one of the top-down processes that influences attention allocation as described in Section 4.1.

The fuel and altitude indicators were also placed differently on the display. The altitude indicator was located in the middle of the combined display, near the instruments required for the flying task (8° visual angle separation between the altitude and attitude indicators, 4° visual angle separation between the altitude and RoD indicators). The fuel indicator was located on the left side of the display, set apart from the other instruments (9° visual angle separation between the fuel and attitude indicators, 18° visual angle separation between the fuel and RoD indicators). Therefore, it required relatively more effort for the subjects to move their eyes to the fuel than to the altitude indicator. This is in line with models of attention that have suggested the attention allocated to an instrument is inversely proportional to the relative distance one's eyes must be moved to fixate on it (Section 4.1).

Unlike the fuel call-outs, the comm light performance did not significantly change over Auto→TA mode transitions without an LPR (Friedman test, p=0.12) although the average response time increased slightly over the phases (Figure 9).

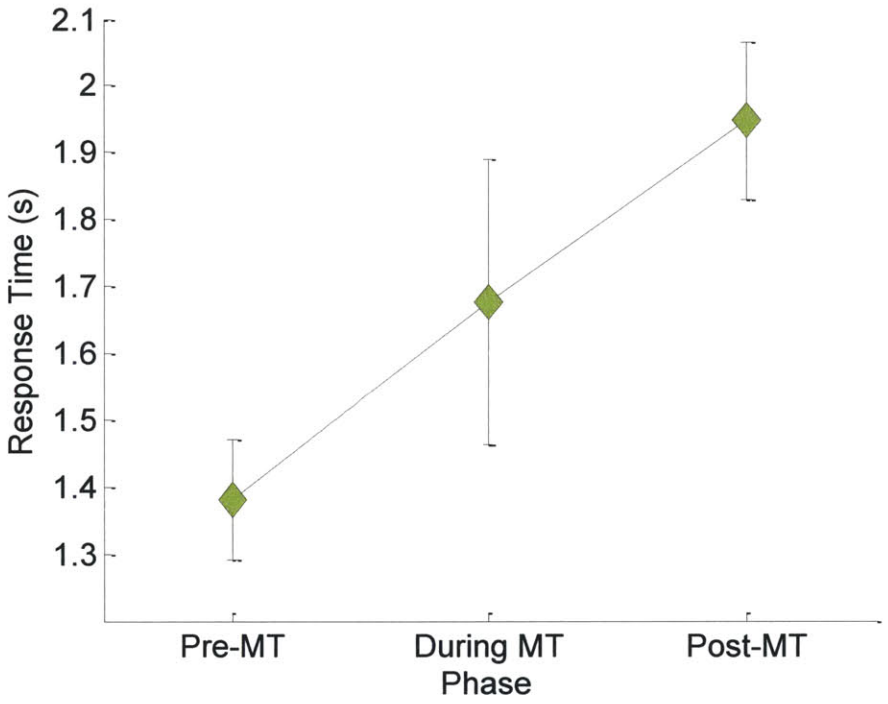


Figure 9. Response time to correctly identified comm lights responses increased slightly, but not significantly, across the Auto→TA mode transition. Error bars indicate standard error.

Subjects were not responsible for the flying task in Auto; therefore, during this phase the comm light response task became the primary task. This may explain why subjects allocated more attention to the comm light than was necessary for satisfactory performance. After transitioning to TA, they were able to sacrifice the "extra" attention without a significant effect on response time. The location of the comm light may have also helped to encourage this re-allocation. Like the fuel indicator, the comm light was located off to the far side of the display (27° visual angle separation between comm light and altitude indicator). As such, it required a comparatively large effort to fixate on it.

This analysis shows that subjects increased and decreased their attention on certain instruments in systematic, largely predictable ways, but it does not indicate the manner in which these changes occurred. Subjects may have increased or decreased the number of fixations on an instrument, the dwell duration, or both. Figure 10 shows the number of fixations/window (on the y-axis) and the dwell duration (indicated by the size of the data points) for the Auto→TA mode transition without an LPR. Similar figures, for all MTs with and without an LPR, can be found in Appendix B.

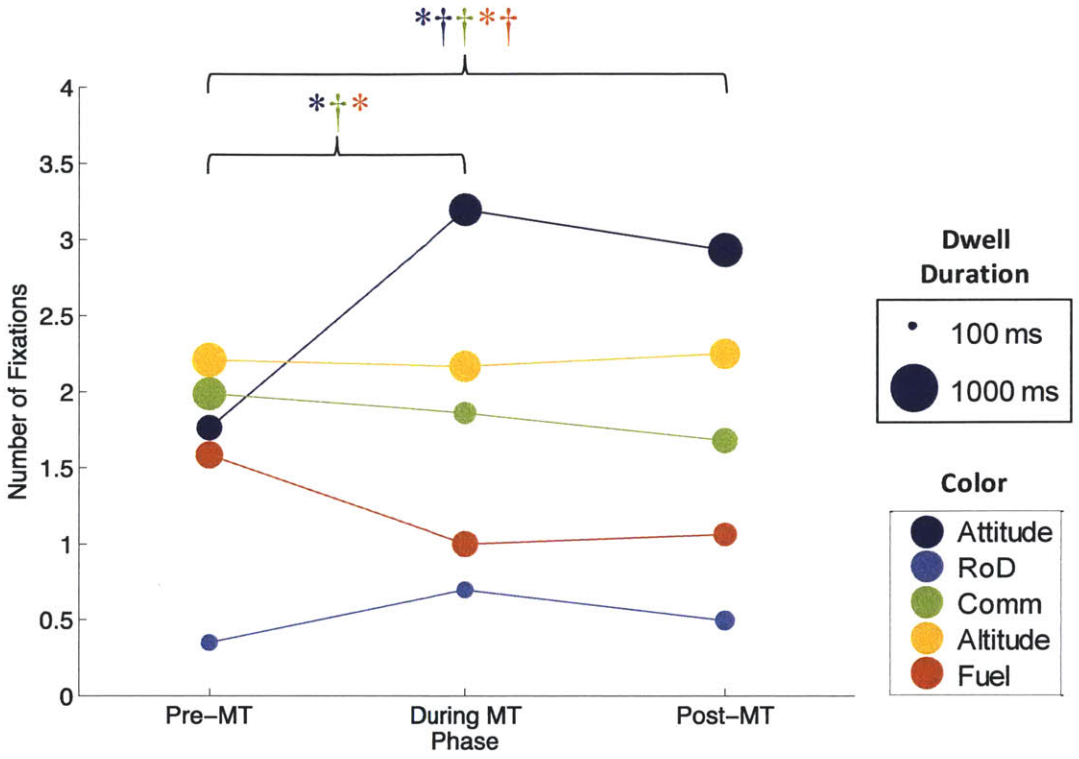


Figure 10. The number of fixations/window (y-axis) and the dwell duration (size of data point) across Auto→TA mode transitions. Error bars indicate standard error. Significant differences (α=0.05) in the number of fixations between phases are marked with *. Significant differences in the dwell duration between phases are marked with †.

Mixed-model ANOVAs were performed to test whether the number of fixations/window and dwell durations significantly changed across this mode transition for any instruments. Like the ANOVAs for the percent of attention, these ANOVAs used phase as a fixed factor and subject as a random factor. In order to make the residuals satisfy the Fisher criteria, the numbers of fixations were averaged together across trials for each subject and phase before analysis, and the dwell durations were transformed by the natural log. Significant main effects of phase were found for the number of fixations on the attitude and fuel indicators (Attitude: $F(2,94)=27.8$, $p<0.0005$; Comm light: $(2,94)=1.17$, $p=0.31$; Altitude: $(2,94)=0.10$; $p=0.90$; Fuel$(2,22)=12.2$; $p<0.0005$) and for the dwell duration on the attitude indicator, comm light, and fuel indicator (Attitude: $F(2,226)=7.77$, $p=0.0005$; Comm light: $F(2,210)=11.1$, $p<0.0005$; Altitude: $F(2,227)=1.89$, $p=0.15$; Fuel: $F(2,180)=8.18$, $p<0.005$).

While there were statistically significant differences in the number of fixations/window and dwell duration on the fuel indicator, they were not substantial (<1 fixation and <100 ms dwell duration). This highlights the advantage of analyzing the attention budget: insignificant changes in the number of fixations and dwell duration can combine to produce a significant change in the attention budget (Figure 7) and task performance (Figure 8). Unlike the fuel indicator, the statistically significant differences for

the attitude indicator and comm light were substantial. The fact that the dwell duration on the comm light decreased without a significant change in the number of fixations suggests that subjects dwelled on the comm light longer than was required for satisfactory performance. When subjects transitioned to TA, they could reduce this dwell duration without a significant change in performance. Finally, the significant increase in the number of fixations and the average dwell duration on the attitude indicator makes sense: the manual control task obviously requires more frequent attention (to make sure that the guidance errors never grow too large) and longer dwells (to make the necessary control inputs).

4.4.2 *Mode Transitions between TA and TA+RoD*

Hypothesis 2 predicted how subjects would re-allocate their attention across mode transitions between TA and TA+RoD without an LPR. As with the analyses for Auto→TA mode transitions, separate mixed-model ANOVAs were conducted for each mode transition and instrument. The percent of attention was the dependent variable (transformed by the square root, windows with no attention on an instrument removed), and subject and phase were random and fixed factors, respectively. Because the subject controlled the RoD before or after these two mode transitions, it was included in the analyses. The percent of attention spent on attention "overhead" was also analyzed to see if subjects changed the amount of overhead or if they just re-allocated attention among the instruments.

Figure 11 shows the results of the analyses. For mode transitions from TA→TA+RoD there was a significant main effect of phase for the flight instruments (Attitude: $F(2,210)=27.6$, $p<0.0005$; RoD: $F(2,145)=11.9$, $p<0.0005$), but not the secondary instruments (Comm light: $F(2,187)=0.24$, $p=0.79$; Altitude: $F(2,192)=0.58$, $p=0.56$; Fuel: $F(2,139)=0.33$, $p=0.72$) or overhead ($F(2,157)=0.37$, $p=0.69$). Bonferroni-corrected pairwise comparisons for the attitude and RoD indicators found that attention on the RoD indicator significantly increased across the mode transition (pre-MT compared to during- and post-MT). Contrary to Hypothesis 2.a., subjects unexpectedly re-allocated attention *away* from the attitude indicator across the mode transition. They chose to reduce attention on the primary instrument rather than on the secondary and tertiary instruments.

In the analysis of TA+RoD→TA mode transitions there was a significant main effect of phase on the percent of attention for all instruments except the comm light and the attention overhead (Attitude: $F(2,192)=7.44$, $p=0.0008$; RoD: $F(2,181)=6.29$, $p=0.002$; Comm light: $F(2,152)=2.39$, $p=0.095$; Altitude: $F(2,110)=32.6$, $p<0.0005$; Fuel: $F(2,209)=8.79$, $p<0.0005$; Overhead: $F(2,150)=3.58$, $p=0.03$). While the change in overhead was statistically significant, it was insubstantial, only -3.6%. As expected, attention on the RoD indicator decreased across the mode transition while attention on the altitude and fuel indicators increased. Figure 12 shows that the increased attention on the altitude and fuel call-outs lead to improved call-out performance across the mode transition. However, the effect was significant (per Friedman tests as in Section 4.4.1) only for the altitude call-outs (Altitude: $p=0.0006$; Fuel: $p=0.13$). Once again, contrary to Hypothesis 2.b., subjects' attention on the attitude indicator significantly decreased across the TA+RoD→TA mode transition.

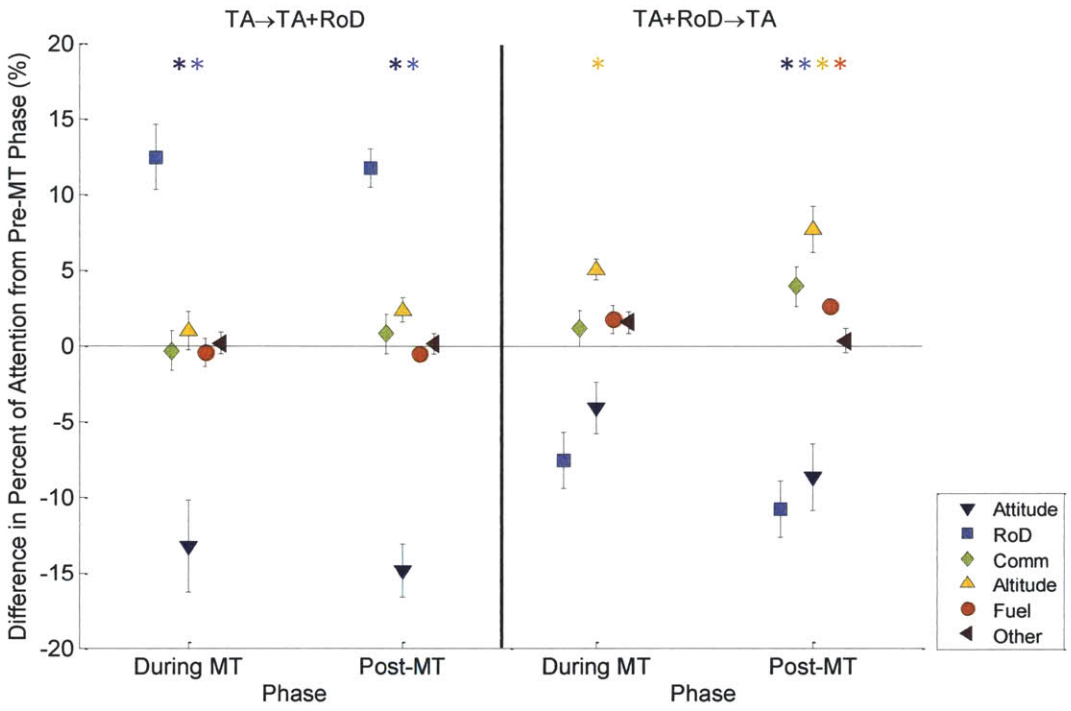


Figure 11. Change in percent of attention from the pre-MT phase to the phase noted on the x-axis. Mode transitions (indicated on the figures) did not feature a landing point redesignation. Error bars indicate standard error. Significant differences (α=0.05) are marked with *.

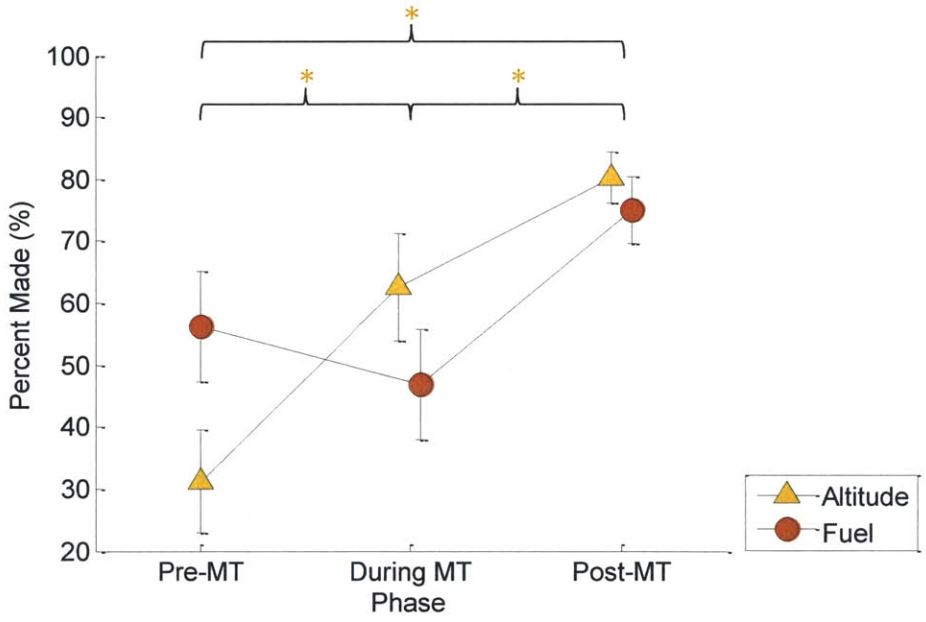


Figure 12. Performance on the altitude call-out increased during the TA+RoD→TA mode transition. Error bars indicate standard error. Significant differences (α=0.05) between phases were found for the altitude indicator, and are marked with *.

These analyses support Hypothesis 2.c., but reject 2.a. and 2.b. Subjects decreased attention on the attitude indicator across both TA→TA+RoD and TA+RoD→TA mode transitions. In retrospect, this behavior may have occurred because the rate-of-change of the attitude and RoD guidance reference trajectories was not constant over the trial. At the beginning of each trial, the reference trajectory directed the subjects to pitch the vehicle from 45° to ~7° in 20 s. This corresponded to a high reference trajectory rate-of-change, as is shown in Figure 13. With a high reference trajectory rate-of-change, the error between the guidance-prescribed and actual attitude grew quickly when a subject was not making control inputs. In order to keep this error low, the subject needed to make frequent control inputs, which required frequent fixations on the attitude indicator. The experimental data shows that there were 0.75 s between fixations on the attitude indicator in the first 10 s of a trial. After the mode transition at 20 s, the reference trajectory was flatter. The guidance error did not grow as fast, and subjects did not have to make control inputs and fixations on the attitude indicator as frequently. This is reflected in the experimental data: at the end of the trial (40-50 s after the trail beginning) there were 1.5 s between fixations on the attitude indicator. This reduction in the attention required on the attitude indicator meant that subjects were arguably freer to re-allocate attention to other instruments. Although subjects performed the flying task before and after the mode transition, in retrospect it was realized the nature of the flying task – and therefore the attention allocated to the attitude indicator – was arguably somewhat different.

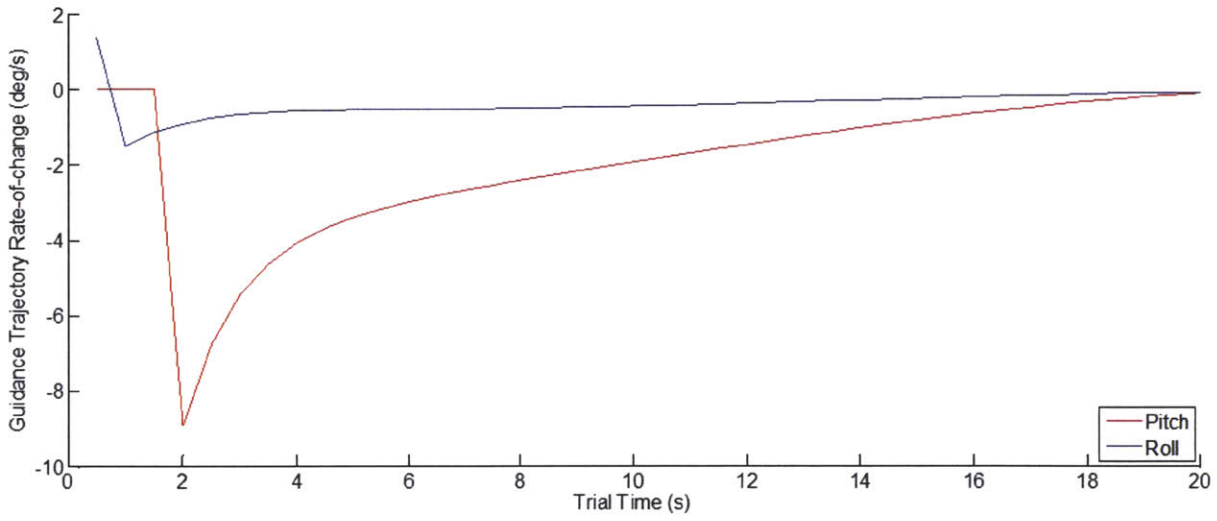


Figure 13. The pitch and roll reference trajectory rates-of-change throughout an example trial in Auto, with no mode transition or landing point redesignation.

4.4.3 *Transient Effects after Mode Transitions*

The third set of hypotheses asserted that that attention on the attitude indicator, fuel indicator, altitude indicator, and comm light would have a constant, steady-state value after transitions to manual control modes that lacked an LPR. Conversely, transitions to manual control modes accompanied by an

LPR would have required a temporary increase in attention on the attitude indicator and corresponding decrease in attention on the secondary instruments.

Mixed hierarchical regression was used to test these hypotheses. Two separate mixed hierarchical regressions were performed for each instrument: one for the trials with an LPR and one for those without an LPR. The percent of attention was transformed by the square root before analyzing the data and windows with no attention on an instrument were removed. The mode transition was included as a 4-level (mode transitions ending in TA or TA+RoD) categorical independent variable, the window number as a continuous independent variable, and subject as a random factor. The cross-effect of mode transition and window was also included in the models. The primary factor of interest was the main effect of window, which described the rate (%/window) at which the attention on an instrument increased or decreased after the mode transition.

In trials with an LPR, there was a significant effect of window for all four instruments (Attitude: $z = -11.4$, $p<0.0005$; Comm light: $z=3.52$, $p<0.0005$; Altitude: $z=3.98$, $p<0.0005$; Fuel: $z=2.61$, $p=0.009$). The change in the percent of attention on the attitude indicator was -3.51±0.31%/window (Figure 14). However, the effect of window was not as substantial for the secondary instruments. The change in the percent of attention was +1.20±0.30%/window for the altitude indicator, +0.83±0.24%/window for the comm light, and +0.60±0.23%/window for the fuel indicator (Appendix B).

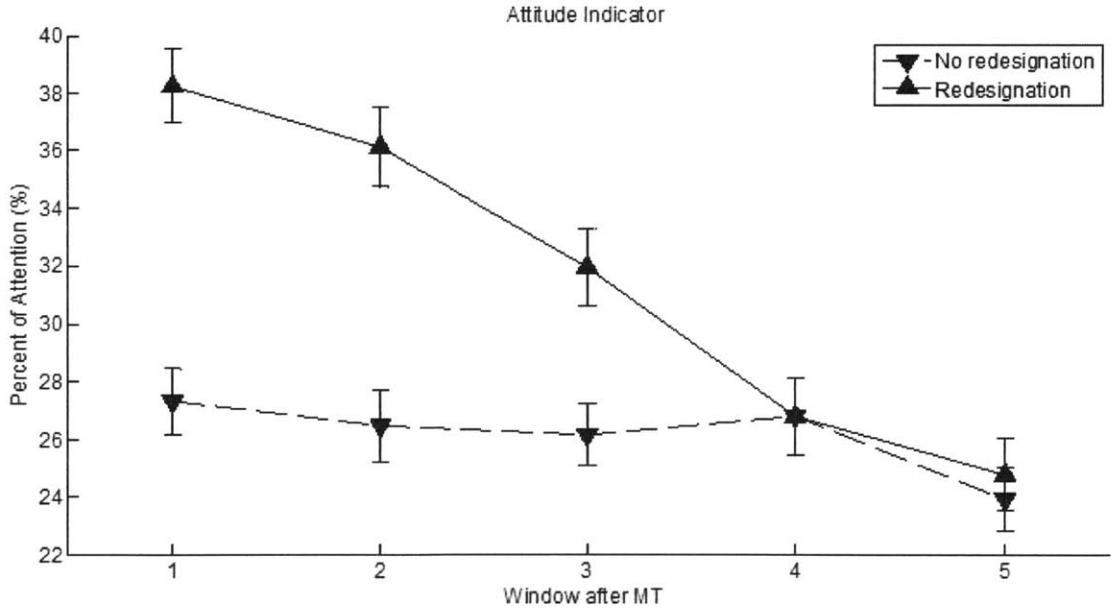


Figure 14. The percent of attention on the attitude indicator after the mode transition was greatly different depending on whether or not there was a landing point redesignation. Data was averaged over all four mode transitions that end in TA or TA+RoD (MTs 3-6). Error bars indicate standard error.

When there was no LPR, the effect of window was significant only on the attitude indicator (Attitude: z=-2.31, p=0.021; Comm light: z=3.52, p=0.75; Altitude: z=3.98, p=0.43; Fuel: not analyzed because the fit to the model was poor). Nevertheless, the effect was not substantial (-0.72±0.31%/window).

These results thus confirm Hypotheses 3.a. and 3.b. The LPR caused an increase in attention on the attitude indicator that was not present without an LPR. By the end of the post-MT phase, however, the two percents of attention on the attitude indicator – with and without LPR – were nearly identical (Figure 14). This suggests that subjects returned to a steady-state scan pattern after the LPR. This return to a steady-state scan pattern did not occur after the first 6-s window, in which subjects nulled the step change in flight guidance error caused by the LPR. Instead, the return to steady-state happened gradually over three 6-s windows. It is believed that this occurred because an LPR caused a longer-term change to the reference trajectory in addition to the step-change in guidance error. This effect can be seen in Figure 15. When there was no LPR the reference trajectory rate-of-change was small and required minimal control inputs. However, when there was an LPR, the reference trajectory had a greater rate-of-change even after the initial step-change. This required more control inputs, therefore a higher percentage of the subjects' attentions. Also note that the increase in attention on the attitude indicator with an LPR was not mirrored by an equal decrease in attention on the secondary instruments. This suggests that subjects had other attention resources they could draw on – perhaps the RoD indicator or the "other" instrument category.

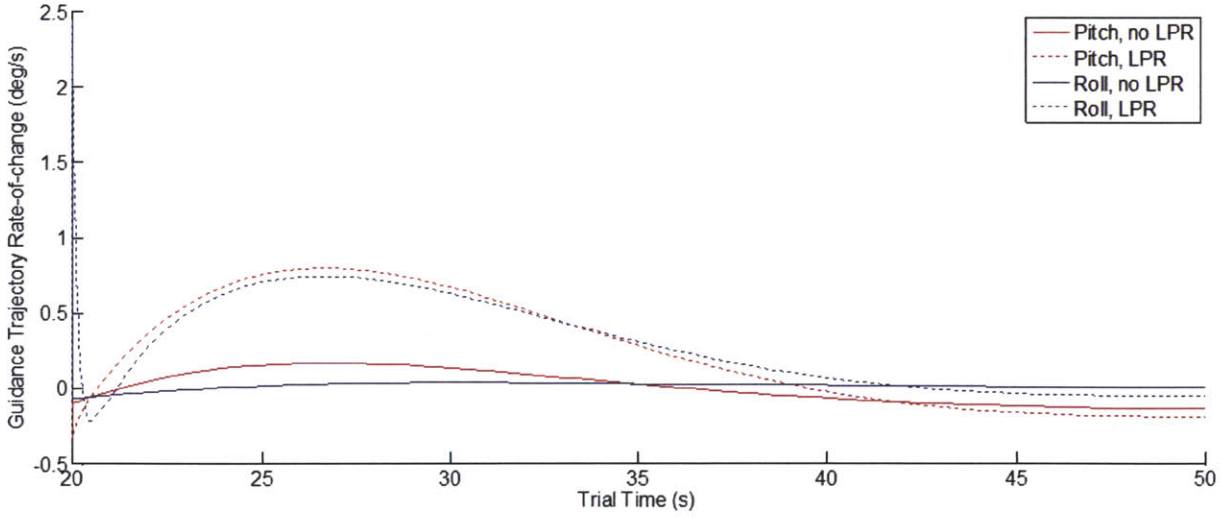


Figure 15. The shape of the reference trajectory was different depending on whether or not there was a landing point redesignation, which occurred at 20 s into the trial (the x-axis origin).

4.5 Conclusions

To the author's knowledge, this experiment was the first to measure operator visual attention, flying performance, mental workload, and situation awareness all together during control mode transitions in which specific control loops varied between passive monitoring and active multiloop control. The results help describe how subjects re-allocated their limited attention resources when their mental workload changed upon entering a new control mode. The experiment also demonstrated the effects that this re-allocation of visual attention had on subjects' situation awareness of vehicle states. Many experimental effects were large enough to suggest that the attention demands and the relative priority of supervisory and manual flying tasks should be considered in system design, particularly when mode transitions might occur. The behavioral trends observed in the present experiment can help designers predict which instruments will receive more or less attention after a mode transition.

This experiment found that adding a two-axis attitude manual control task (nulling pitch and roll flight director guidance errors) to the existing supervisory tasks (here, responding to a comm light visual secondary task and a tertiary task of making verbal situation awareness call-outs) reached or exceeded subjects' visual attention capacity. In order to maintain performance in the flying task, subjects chose to re-allocate attention away from the fuel indicator and the comm light. As a result, there was a significant decrease in subjects' situation awareness of the fuel indicator, as indicated by the fuel call-out performance. In a real lunar lander, such decreased situation awareness could have real operational consequences. The operator may not realize that he is nearly out of fuel, for example, or he may miss noticing a fuel leak. Knowing which instruments acquire a lower priority during and after a mode transition would allow designers to expect that the operator will be attending to them less frequently. If automation detects a system failure whose signature includes the lower priority instrument, the system could draw attention to that instrument using visual or auditory display cues to compensate for the operator's lowered visual attention and reduced awareness on that instrument.

System designers must understand that the priority of each of the operator's tasks is not only explicitly stated during flight training – as it was here in the pre-experiment training – but that it can also be influenced by the nature of the tasks and the design of the displays. In this experiment, the percent of attention on the altitude indicator (a tertiary task) was found to be higher and decrease less than the percent of attention on the comm light (the secondary task) and the fuel indicator (a tertiary task of equal priority). This behavior was likely influenced by the task priority, the frequency of the task (comm light illuminations or call-outs), and the relative effort required for subjects to move their eyes to the instrument. These results confirm the findings of previous visual attention research (Wickens et al., 2003).

Like the discrete comm light response and call-out tasks, attention on the continuous manual control task also appeared to be influenced by subjects' expectations of event frequency. However, the influential parameter was the guidance trajectory rates-of-change, which dictated the required frequency of control inputs. This was demonstrated in the decrease in attention on the attitude indicator across TA→TA+RoD

and TA+RoD→TA mode transitions (Hypotheses 2). The two-axis flying task remained subjects' first priority across these mode transitions, but the guidance trajectory rate-of-change decreased. Consequently, the percent of attention on the attitude indicator also decreased. The same behavior occurred when there was a landing point redesignation (LPR). Attention on the attitude indicator increased with the step-change in the flight guidance error caused by the LPR, and then gradually decreased to a steady-state (no LPR) value as the reference trajectory rate-of-change decreased. The mode transitions between TA and TA+RoD and the transient behavior after an LPR both demonstrate how faster rates-of-change in the guidance trajectory require more control inputs and greater attention.

4.5.1 *Limitations and Future Work*

Though the flying task used in this study has some technical face validity, only three control modes were employed to decrease experimental time, cost, and complexity. In reality, operational aerospace systems have many more control modes. Furthermore, the control modes in this experiment only varied in the number of manual control loops the subjects had to close. The supervisory tasks – monitoring the fuel level, altitude, and comm light – were never allocated to the automation. Future research can confirm and extend the results of this research by investigating subject performance across different control mode transitions in similar, simplified experimental settings.

Another fruitful area of future research would be to re-arrange the display elements to further test the role of the distance between flight instruments in determining how subjects allocated their attention. For example, researchers could exchange the fuel and altitude indicators, or move the comm light to a more central location. As an instrument moves closer to the flight instruments in the center of the combined display, it is expected that it will receive more attention. However, some research has suggested that the bandwidth of the information (how frequently the instrument needs to be sampled) is more important than the relative effort of moving one's eyes (Steelman et al., 2011). So, if the fuel and altitude instruments switch locations, it is likely that there will be a slight increase in subjects' attention on the fuel indicator and a slight decrease in subjects' attention on the attitude indicator. However, it is unlikely that the relative priority between the two will change.

One final limitation is that the simulator provided no motion cues, and flying it presented no real risk to the subjects comparable to that of a true lunar landing. Subjects' visual attention priorities may be quite different in a real lunar lander cockpit. For example, monitoring the altitude is critical when one is rapidly descending toward the real lunar surface, whereas our subjects did not need to sample the attitude indicator to fly the vehicle. They could simply focus on nulling flight director cues. Another limitation was that with one exception, the experimental subjects were nonpilots with little formal experience in manual control and instrument scanning. Recruiting highly experienced pilots might have altered the results (Bellenkes et al., 1997; Hayashi, 2003; Huemer et al., 2005). Future work could test expert pilots in the same and different experimental conditions to compare how they re-allocate attention across mode transitions.

This page intentionally left blank.

The model set forth here combines
Predictions of how pilots fly
With an important extension
To prior work with attention
That describes movements made with the eyes.

5.0 A Closed-Loop Model of Operator Attention and Control across Mode Transitions

This chapter proposes an integrated human-vehicle model to better understand the complex ways that visual attention, mental workload, and situation awareness change before, during, and after control mode transitions. The model replicates the lunar landing scenario described in Chapter 4, including identical vehicle dynamics and guidance algorithms. The human performance model is quantitative, and is based around the four-stage "perceive-think-decide-do" scheme for human information processing (Wickens & Hollands, 2000). It builds upon a number of established human performance models: the Crossover (McRuer et al., 1965) and Optimal Control (Kleinman et al., 1971) Models for manual control and SEEV models for visual attention (Wickens et al., 2003). One strength of this model is that it specifies attention for both manual control and supervisory tasks, unlike prior attention models that focus on one or the other. Furthermore, this model has the ability to capture changes in attention on specific control loops as they vary in priority and between passive monitoring and active multiloop control. Another strength of this model is that it is a closed-loop model, which allows it to simulate the effect that changes in visual attention have on operator situation awareness and manual flying performance, and the effect of manual control demands (i.e. mental workload) on attention. Very few models simulate all of these factors in concert, and those that do are incredibly detailed and require the specification of many parameters. The model is not intended to give absolute predictions of operator behavior; rather, it aims to illuminate the relative differences between simulated conditions. This allows researchers to identify unexpected emergent behaviors that warrant further targeted human subject experiments.

5.1 Background

5.1.1 *Models of Operator Manual Control*

5.1.1.1 The Crossover Model

The Crossover Model (McRuer et al., 1965) describes how a human operator adjusts his control strategy to provide a stable closed-loop system response with reasonably-low error (McRuer et al., 1965; Young, 1973; Jagacinski and Flach, 2002). The operator can provide control lead to the system to overcome his reaction time delay and neuromuscular lag, which are collectively represented as a time delay. Beyond the lead required to overcome these inherent limitations, the operator can also provide lead and lag to stabilize the closed-loop system response. The lead and lag are chosen strategically based on the vehicle dynamics. For a stable closed-loop response, the operator must provide lag for a zero-order plant (position control) and lead for a second-order plant (acceleration control). Once the operator has applied lead and lag to keep the system stable, he can apply a control gain to minimize the tracking error. These control strategies make the joint human-system open-loop transfer function appear as a gain, a time delay, and an integrator in the region of the crossover frequency (the frequency where the open-loop

amplitude ratio is 0 dB). The control of multiple integrated control loops is simulated by arbitrarily lengthening the time delays, lowering the loop gains, and increasing the remnant noise to represent the effects of divided attention (McRuer & Jex, 1967; McRuer & Krendel, 1974).

The Crossover Model is a simple, easy-to-understand representation of operator manual control. However, it has limitations: it is only valid in the region around the crossover frequency, and it is best suited for compensatory tasks in which the error between the actual and guidance-prescribed state is displayed.

5.1.1.2 The Optimal Control Model

The Optimal Control Model (OCM) predicts the control inputs an operator will make during a simple tracking task (Kleinman et al., 1971). The OCM combines the vehicle dynamics and a human performance model (HPM) into one linearized system. In the OCM, it is assumed that the operator does not directly perceive the actual system states needed for control. Instead, he perceives a subset of states perturbed by noise and delayed because of his inherent reaction time. The noise, also called the remnant, captures the non-linear portion of vehicle and human behavior that is not modeled by the linearized system (Levison et al., 1969).

The OCM makes estimates of the actual system states based on the concept of a Kalman filter (Kalman & Bucy, 1961). The Kalman filter is an algorithm that estimates the system state from noisy, delayed measurements using information about: 1) the dynamics of the system, 2) the operator's control inputs, 3) the relationship between the system dynamics and the control inputs, and 4) the properties of the external disturbances and measurement noise. Together, the components used by the Kalman filter represent the operator's "mental model" of the system and its behavior.

The operator's closed-loop control inputs to the system are based on the estimate of the actual state produced by the Kalman filter. These control inputs can be shaped to optimize a mathematical cost function, trading control input effort against state errors. The theoretical duality of the state estimation and control problems is exploited, giving the Optimal Control Model its name. While the full closed-loop OCM is only applicable to manual control tasks, the optimal state estimation portion of the model is more generally applicable. It can be used to model tasks where a passive observer needs to monitor certain state variables but does not have an immediate knowledge of the control inputs being made. One example of these passive monitoring tasks is human spatial orientation, and a number of Kalman filter models have been used to simulate this behavior (Oman, 1982; Merfeld et al., 1993).

5.1.2 *Models of Operator Attention*

5.1.2.1 The Optimal Scanning Model

The attention model based on the OCM – called the Optimal Scanning Model – predicts the amount of attention (measured in time) that is required on a display to produce the optimal control inputs (Baron & Kleinman, 1969; Baron et al., 1970). These models capture the shift of attention between different displays by varying the level of assumed perceptual noise. The noise on a given display is lower when the HPM is fixating on it and higher when the HPM is fixating elsewhere. This simulates foveal and peripheral vision, respectively. Visual acuity is higher in foveal vision than in peripheral vision, and so the perceptual noise is lower. The Optimal Scanning Model has been used as a design tool, predicting how changes to the flight director or displays will influence the operator's scan pattern (Kleinman, 1976; Hess, 1981). Sheridan (1970) created a similar mathematic model that determined the optimal frequencies with which an operator needed to sample an unknown process and make control inputs in order to maximize a given value function.

Because the Optimal Scanning Model is based on the full, closed-loop OCM, it can only predict the amount of attention required for manual tracking tasks. This is the primary limitation of the model. Manual tracking tasks are often the primary task in aerospace systems (e.g. controlling attitude to match a flight director), but they are rarely the only task that an operator must complete. Aerospace systems usually require the operator to perform additional supervisory tasks, in which he monitors the automation or the state of a display without directly controlling it. The amount of attention allocated to each supervisory task influences the operator's situation awareness of the related vehicle state (Chapter 1). The Optimal Scanning Model can be used to predict the amount of spare attention left over after the manual tracking task is completed, which is the attention that can be used to complete all supervisory tasks. The amount of this spare attention is assumed to be inversely proportional to the operator's mental workload. While the Optimal Scanning Model can be used to predict mental workload, it cannot predict situation awareness. It does not capture how much attention each supervisory task requires for satisfactory situation awareness, and therefore cannot predict if these requirements are being met.

5.1.2.2 SEEV Attention Models

The family of SEEV models predict attention for any set of tasks, not just manual control (Wickens et al., 2003). In SEEV, the percent of visual attention spent on a given instrument is driven by bottom-up and top-down processes. The bottom-up properties describe the characteristics of the display – the relative salience of the instrument (e.g. color and luminance) and the relative effort[4] required for an operator to

[4] It is important to note that moving one's eyes does not require much physical effort. The SEEV literature uses the term "effort" to denote the relative amount of time that it would take to move one's eyes from the current instrument to any other instrument. The instruments that are farther away take longer to move to, and therefore have a higher "effort." The model in this thesis uses the same effort primitive used in SEEV attention models, and so the term "effort" is used in place of "distance" or "movement time."

move his eyes to the instrument, in comparison to all other instruments. More influential, however, are the top-down processes: the expected frequency of events involving the instrument (called the "expectancy"), and the value of the instrument's information to the current tasks. The current tasks are also prioritized, and the instrument's value to each task is multiplied by that task's priority. These products are then summed, giving the instrument's overall value in the current task structure. These four parameters – salience, effort, expectancy, and value – are known as attention primitives. Each primitive can be weighted differently, describing its relative influence on the allocation of attention

The challenge in applying SEEV attention models to real-world scenarios is quantifying the primitives and their weightings in the context of SEEV's qualitative framework. In the original formulation of SEEV, the researcher or other subject-matter experts had to subjectively specify numbers for the SEEV primitives. These were often derived from qualitative rankings: for example, Gore et al. (2009) quantified an instrument's expectancy as 0, 0.333, 0.666, or 1 based on whether the event rate was "none," "low," "moderate," or "high" relative to the other instruments. More recently, quantitative methods have been developed to objectively determine the salience, effort, and expectancy for a given scenario or phase of flight. Computational models can determine the salience of static and dynamic displays based on the luminance, color, and orientation of instruments (Itti & Koch, 2000; Walther & Koch, 2006). The expectancy of an instrument is often taken as proportional to the rate of events. This value can be specified by the experimenter (Steelman-Allen, 2011) or learned by the model through training simulations (Wortelen et al., 2013a-c). Finally, the effort to move from one instrument to any other can be computed as the angular distance between the two on the display (Gore et al., 2009). In contrast to these primitives, the priority of tasks and the value of each instrument to each task remain subjective parameters.

Prior research has investigated the relative weightings of each attention primitive by building models that include and exclude certain primitives and comparing the model predictions to experimental data. The results found that value and expectancy dictated attention allocation more than effort (Wickens et al., 2005; Steelman et al., 2011) and salience (Steelman et al., 2011), and that expectancy dictated attention allocation more than instruments' value (Wickens et al., 2003). For these reasons, many implementations of the SEEV attention model leave out the salience and/or effort primitive (Wickens et al., 2003; Wickens et al., 2005; Horrey et al., 2006; Wortelen et al., 2013a-c).

One limitation of most SEEV models is that, like the Optimal Scanning Model, they predict the total percent of time fixated on each instrument and not the time between fixations (the sampling period). As discussed in Chapter 1, the sampling period of an instrument is an important parameter of the operator's situation awareness because it describes how frequently the operator can correct his estimates of the vehicle states. The SEEV expectancy primitive relates to the operator's desired sampling period of an instrument, but there is no way of determining if this desired rate is met or not.

Another limitation of SEEV models is that they cannot predict the effect of the attention allocation on task performance or operator situation awareness on their own. They must be integrated with models of human control and vehicle dynamics. One prominent integrated model is the Man-machine Integration Design and Analysis System (MIDAS), which simulates operator performance, workload, and situation awareness in aviation (Tyler et al., 1998; Gore et al., 2009; Hooey et al., 2010). MIDAS provides a continuous estimate of situation awareness, which increases whenever the SEEV attention model fixates on an area of interest with information relevant to a current task and decreases in the time between these relevant fixations. This situation awareness model is difficult to validate because there is no easy way to continuously measure situation awareness in an experiment (see Section 4.1 for a review of experimental situation awareness measurement techniques). Furthermore, in MIDAS the operator behavior is specified by a large number of unbounded parameters that can be adjusted to fit individual data sets. This can be useful when the model is used in a descriptive manner (e.g. explaining experimental results), but it limits MIDAS's use in novel situations ions where a new system or vehicle is still relatively undefined.

Other integrated models that employ SEEV or SEEV-inspired attention models include the Cognitive Architecture for Safety Critical Task Simulation (CASCaS) (Wortelen et al., 2013a-c), the N-SEEV attention model (Wickens et al., 2009), and the Attention-Situation Awareness (A-SA) model (Wickens et al., 2005). In CASCaS, attention is allocated to each task present based on its priority, which is specified by the researcher, and the expected rates of events involving that task, which the model learns by training on the scenario. CASCaS predicts the effects that a particular attention allocation has on operator task performance, but does not address how attention is influenced by workload or how attention influences situation awareness. The N-SEEV model predicts the time that it would take an operator to notice a particular event, which can be assumed inversely proportional to the operator's situation awareness (Wickens et al., 2009). Unlike MIDAS, N-SEEV predictions can be validated against similar experimental metrics; however, it does not address how attention influences workload or performance on other tasks. Finally, A-SA predicts continuous situation awareness in a manner similar to MIDAS, and has the same limitation regarding experimental validation.

5.2 Model Structure

The model described in this chapter is a closed-loop model of operator visual attention, flying performance, mental workload, and situation awareness. It describes how an operator makes estimates of the system state, adjusts these estimates by attending to and perceiving information on the displays, and uses these estimates to control the vehicle. The model contains an Attention block that builds upon the strengths of the Optimal Contol Model and SEEV and addresses the weaknesses of each. Similar to SEEV, the attention allocated to an instrument is a function of value and effort primitives. This model also describes a new primitive – the uncertainty in the estimate of an instrument's state – which draws on ideas from the Optimal Control Model. The uncertainty is similar to the SEEV expectancy primitive, but it directly relates to the OCM-inspired state estimator and mental model. Furthermore, the uncertainty primitive responds in a closed-loop manner with the chosen attention allocation and the demands of the

flying task. By integrating this new attention model with models of operator control and vehicle dynamics, predictions of manual control performance, mental workload, and situation awareness can be made. Workload and situation awareness are both predicted through two approaches: a cognitive first-principles approach and a performance-based approach that is validated against experimental data.

The model has two major components: a vehicle model and a human performance model (Figure 16). The current vehicle state and the guidance-prescribed states (the "reference trajectory") are communicated to the human performance model (HPM) through a simulated display. The HPM in turn influences the vehicle states through a simulated hand controller.

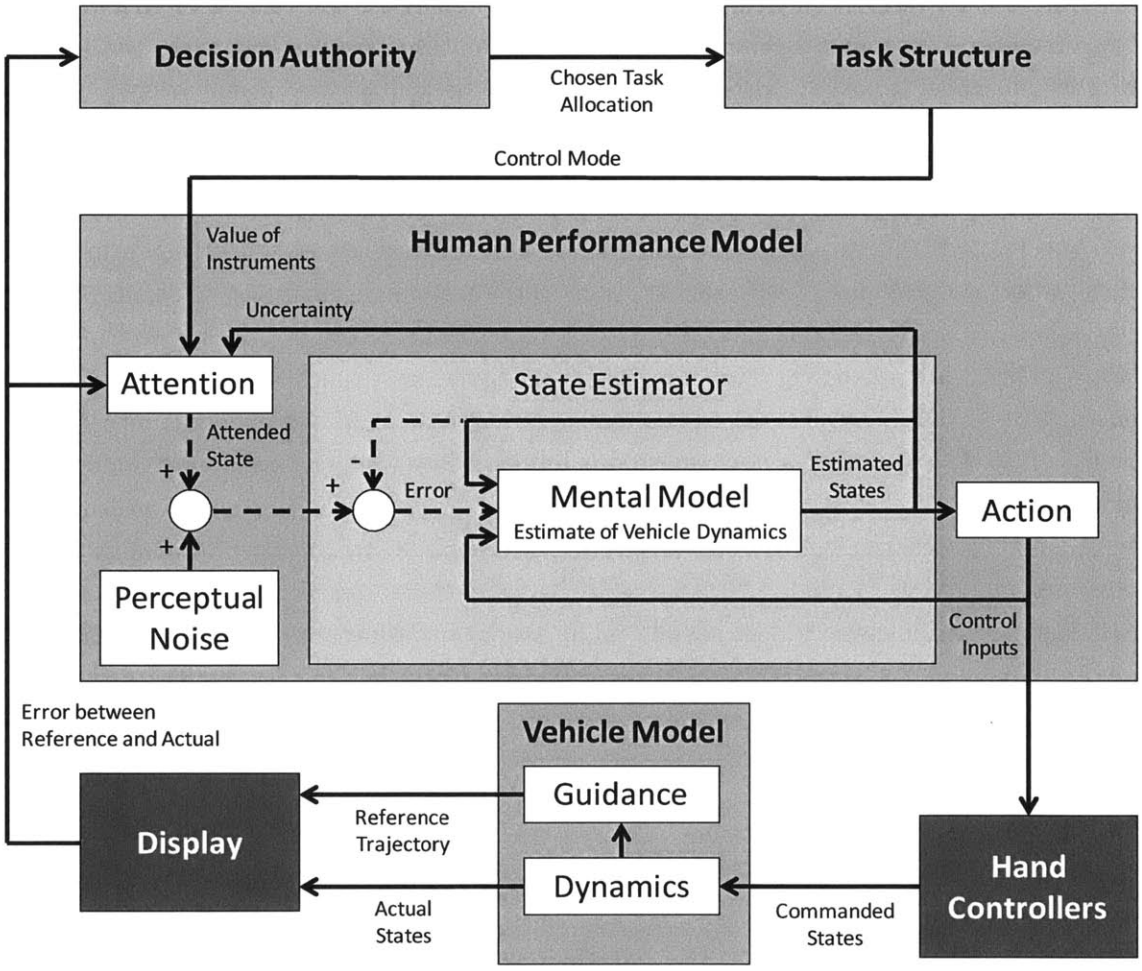


Figure 16. Structure of the integrated vehicle-human model.

The structure of the HPM in this thesis assumes that there are four human stages of information processing: 1) information acquisition, 2) information analysis, 3) decision and action selection, and 4) action implementation (Wickens & Hollands, 2000). The Attention and Perceptual Noise blocks represent the information acquisition stage. The Attention block determines the instrument to be attended during each time step. The Perceptual Noise block simulates an operator's inability to perfectly determine the

state of the vehicle. This perceptual noise is the result of physical human limitations (e.g. limited visual acuity), the design of the vehicle displays (e.g. the inherent quantization of a digital display), and non-linearities of the system dynamics (e.g. the control remnant). The State Estimator block represents the information analysis stage. The HPM continually estimates the system states based on an internal model of the system dynamics (the Mental Model block). The uncertainty in these estimates grows with time, until the instrument displaying a particular state is attended to and the estimate of that state is "corrected." Finally, the Action block represents the action selection and implementation stages. It decides upon an action (making manual control inputs, responding to an illuminated comm light, or making fuel and altitude call-outs) based on the output of the State Estimator.

The HPM has four prioritized tasks to complete, using information from five task-related instruments: 1) an attitude indicator (showing the actual and flight director-prescribed pitch and roll), 2) a rate-of-descent (RoD) indicator (showing the actual and flight-director prescribed state), 3) a cockpit light that illuminates to simulate a data communication ("comm light"), 4) a fuel indicator, and 5) an altitude indicator. The four tasks to be completed are to:

First priority:

- Minimize the error between the vehicle attitude and the guidance-prescribed attitude.
- Minimize the error between the vehicle RoD and the guidance-prescribed RoD.

Second priority:

- Respond to comm light illuminations.

Third priority

- Make call-outs of the vehicle fuel and altitude at pre-determined intervals.

When the control mode involves manual control of both the attitude and RoD, these two highest-priority tasks essentially become one task. While each task involves the same HPM blocks, the behavior of the Perceptual Noise, State Estimator, and Action blocks is different. The following sections discuss these blocks in the context of their task, after explanations of the Vehicle Model and Attention block.

5.2.1 *Vehicle Model*

The vehicle dynamics and guidance algorithms in the model are identical to those used in the experiment in Chapter 4. The vehicle dynamics and controls replicate those of the Apollo lunar module (Duda et al., 2009). There were three distinct control modes: 1) Two-axis (pitch and roll) and rate-of-descent manual control (TA+RoD), 2) Two-axis manual control with automatic rate-of-descent control (TA), and 3) Automatic control of attitude and rate-of-descent (Auto). In the first two modes, the HPM's task is to minimize the error between the vehicle state and the state required to converge with the optimal reference trajectory. This reference trajectory is continually updated based on the current vehicle state. It plots a smooth trajectory that brings the vehicle to a hover 150 ft. above the selected landing aim point (Bilimoria, 2009). In the Auto control mode, the vehicle automatically follows the reference trajectory.

The vehicle model can enact transitions between two control modes at any time. Model simulations discussed in this chapter all used the same initial conditions presented in Section 4.2.2, and each simulation runs lasted 50 s.

5.2.2 *Allocation of Visual Attention*

The Attention block in the HPM is based on the SEEV model of attention (Section 5.1.2.2), with a number of modifications. At each time step, the Attention block fixates on one of six instruments: 1) the attitude indicator, 2) the rate-of-descent indicator, 3) the comm light, 4) the altitude indicator, 5) the fuel indicator, or 6) a generic "other" indicator. The presence of the other indicator captures the amount of attention "overhead": the time that an operator spends looking at regions of the display outside the five task-related instruments, looking off of the display, and/or dwelling on a point less than the 100 ms threshold for a fixation (Section 4.2.3). The other indicator does not include pursuit movements between instruments, which are counted separately (Section 5.2.2.4).[5]

The model assumes that all tasks require foveal vision, and only one instrument can be attended to at a time. This is different than the Optimal Sampling Model, which models peripherally-viewed instruments with increased noise. Only modeling foveal vision is arguably a defensible assumption given the design of the displays and tasks. In the experiment, errors between the actual and flight director-specified attitude and RoD were not identifiable in peripheral vision unless they were very large, the fuel and altitude call-out tasks required the subjects to read the respective instruments, and the blue and green colors of the illuminated comm light were very close to the turquoise background color of this instrument. (See Section 4.2.4 for further details about the experimental tasks).

Unlike many SEEV models, which only predict the percent of attention allocated to each instrument, this model uses the SEEV primitives to drive attention around to different instruments in a real-time simulation. This is similar to more-recent implementations of the SEEV model (Steelman et al., 2011; Wortelen et al., 2013a-c). After the HPM completes a fixation on one instrument, it selects the next instrument based on three attention primitives: 1) the *value* of each instrument to the current task structure, 2) the level of *uncertainty* in the current estimate of each instrument's state, and 3) the *effort* required to move eyes to each instrument.

The value, uncertainty, and effort of an instrument i are combined to produce a weight for that instrument, $w(i)$, as described by the following formula:

$$w(i) = V \cdot \frac{v(i,m)}{v_{max}} + U \cdot \frac{u(\Delta t(i), T(i))}{u_{max}} - E \cdot \frac{e_j(i)}{e_{j\,max}} \tag{5.1}$$

[5] Considering the overhead and pursuit movements as two separate categories is different than in Chapter 4, in which the pursuit movements were contained within the overhead.

v is the value of instrument i, which is a function of the control mode, m. u is the uncertainty of instrument i, which is a function of the time since the last fixation on i, $\Delta t(i)$, and the desired sampling period of i, $T(i)$. e_j is the effort to move eyes from instrument j to instrument i. These primitives were calculated for each instrument from knowledge of the experimental scenario design and experimental data. The dataset used for parameter specification contained 60% of the 432 total experimental trials. This left 33% of the remaining data for use in validating the attention model (Section 5.3), excluding 7% of the data that was discarded due to insufficient eye tracker data or flying performance that was an extreme outlier from the subject's normal performance (the same as in Section 4.2.5). This 60/33% split in the data was necessary for each subject and experimental condition to be equally represented in the parameter-specification and validation data sets. Each subject performed 3 repetitions of each experimental condition, and one of these repetitions was randomly assigned to the validation data set. The others were then assigned to the parameter-specification data set.

Equation (5.1) shows that the value, uncertainty, and effort for instrument i are all normalized by the maximum value, uncertainty, or effort for any instrument at that time step. This ensures that each primitive is measured on the same 0-1 scale. With this normalization, the relative contribution of each primitive is specified by the weighting of each primitive noted by the capital letters V, U, and E. These weightings, which are consistent across instruments, can be adjusted by the researcher to best match experimental data. Section 5.3 describes how the V, U, and E weightings were chosen for this implementation of the model.

The probability of fixating on instrument i is a function of an instrument's weight, $w(i)$ per the following formula:

$$P(i) = (1 - O) \cdot \frac{w(i)}{\sum w} \tag{5.2}$$

In this equation, O is the probability of fixating on a generic "other" instrument. In the model, this was set to the percent of attention given to "overhead" during the experiment in Chapter 4 (approximately 30%). The model randomly selects the next instrument to fixate on based on these probabilities. The model also contains an inhibition-of-return parameter (Steelman et al., 2011), which does not allow for consecutive fixations on any instrument except the "other" instrument. The model also captures the time required to move the eyes between two instruments (Section 5.2.2.4), which is a function of the display design, and the dwell duration on each instrument (Section 5.2.2.5), which is derived from experimental data.

The next three sections discuss each of the attention primitives in more detail.

5.2.2.1 Value

The value of an instrument computed by the Attention block is similar to the SEEV value primitive. In SEEV, an instrument's value is the product of its relevance to each task and the priority of that task. In the experimental scenario used with this model, each instrument is only used for one task. Therefore, the value of an instrument is directly proportional to the priority of the task for which it is used. These priorities were determined from the instructions given to subjects during pre-experimental training (Section 4.2.4). As Table 4 shows, the priority of tasks changes with the control mode. This also changes the value of instruments in the Attention block.

Table 4. The value of instruments in the Attention block is proportional to the priority of the task for which they must be referenced.

Task	Task Priority	Instrument	Instrument Value under Control Mode		
			Auto	TA	TA+RoD
Control attitude	1	Attitude indicator	0	3	3
Control RoD	1	RoD indicator	0	3	3
Respond to comm light illuminations	2	Comm light	3	2	2
Make call-outs	3	Altitude indicator	2	1	1
		Fuel indicator	2	1	1

5.2.2.2 Effort

Effort in the Attention block is also similar to the SEEV effort primitive. The effort required to move eyes from one instrument to another instrument is proportional to the visual angle between the two instruments. This is a function of the display design. The visual angle between instruments (Table 5) was measured with respect to the locations indicated on Figure 17.

Table 5. Visual angle between instruments (°).

	Attitude	Altitude	RoD	Comm Light
Fuel	8.7	15.3	17.6	39.2
Attitude		7.6	11.2	33.9
Altitude			4.4	26.8
RoD				22.7

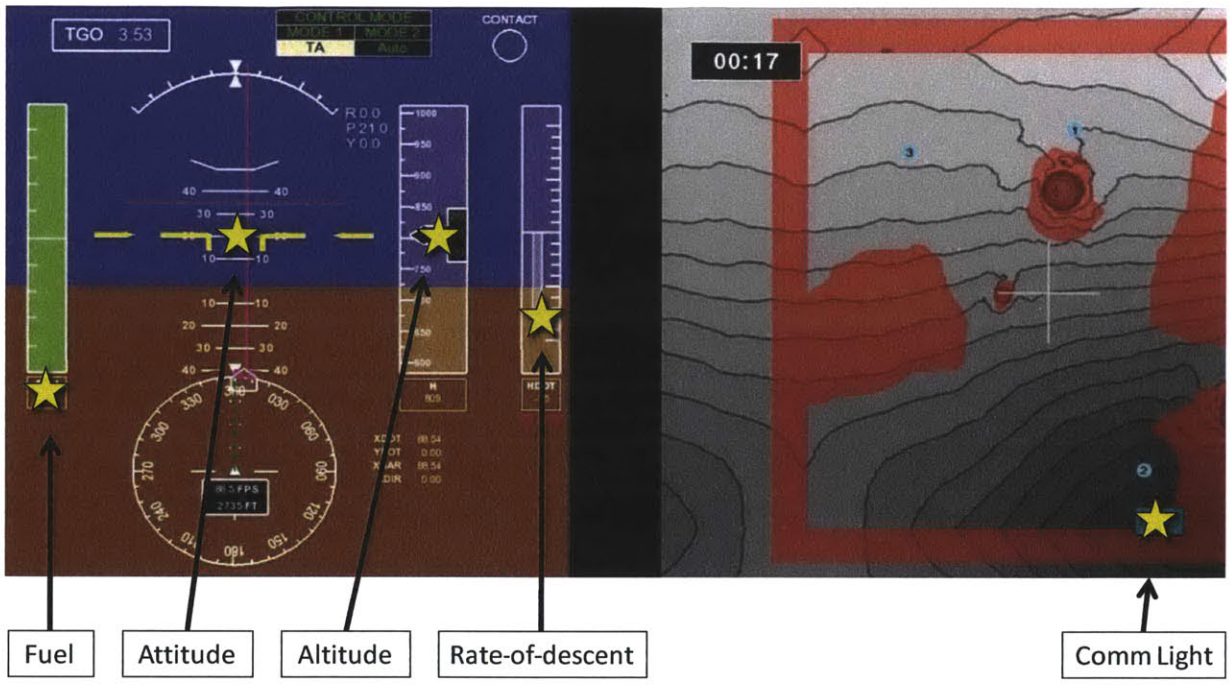


Figure 17. Lunar landing simulator displays with yellow stars marking the reference locations used in the calculation of the visual angle between each pair of instruments.

5.2.2.3 Uncertainty

Uncertainty is a new primitive developed in this thesis. It models the uncertainty in the estimate of an instrument's state, which grows between fixations. When an instrument is being attended, the uncertainty is 0. When an instrument is not being attended, its uncertainty grows according to Equation (5.3).

$$u(\Delta t(i), T(i)) = \frac{\Delta t(i)}{T(i)} \tag{5.3}$$

where $\Delta t(i)$ is the time since the last fixation on instrument i and $T(i)$ is the desired sampling period (seconds between fixations) of instrument i. Figure 18 shows how the uncertainty of each instrument grows and then "resets" to 0 when it is attended to. If an instrument is attended to when the uncertainty is ≤ 1, the desired sampling period has been met. Whenever an instrument's uncertainty is >1, more time has elapsed between fixations than is desired. The desired sampling period has not been met.

The desired sampling period of each instrument is specified by the researcher, and is a function of the amplitude and frequency of the displayed signals. As the amplitude and the frequency increases, the desired sampling period decreases (i.e. the sampling frequency increases). The following subsections discuss how the uncertainty was determined for each instrument in this thesis's experimental scenario.

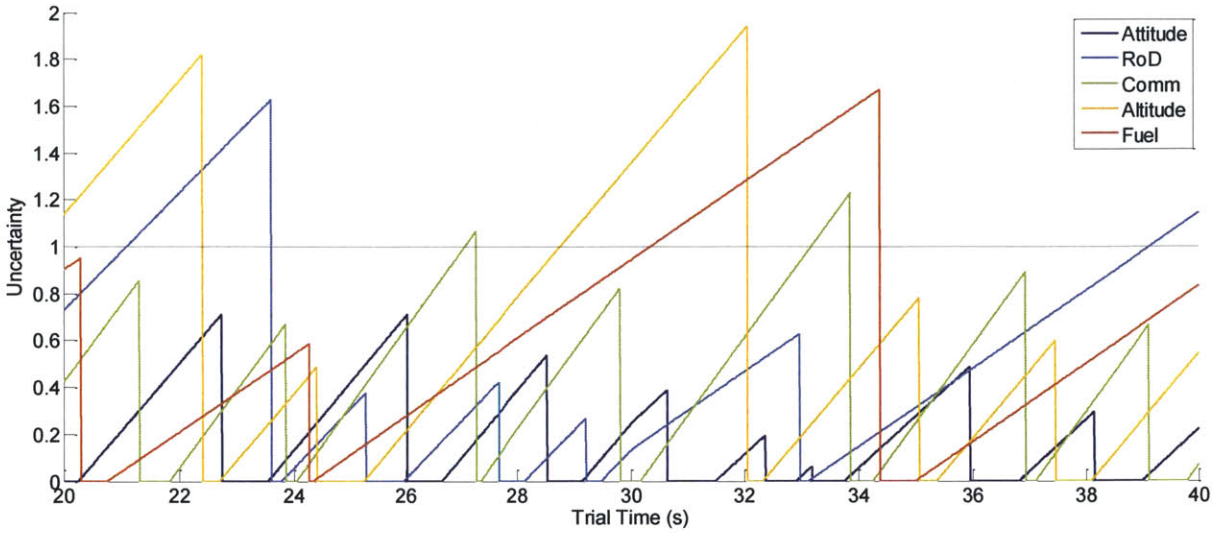


Figure 18. The uncertainty of an instrument is 0 when it is attended, and grows when attention is fixated elsewhere. This data is from the 20-40 s after the commencement of an example simulation run in TA+RoD with no landing point redesignation.

Secondary Instruments – The desired sampling period of the secondary instruments – the comm light, altitude indicator, and fuel indicator – is a function of the frequency of the respective task – the comm light illumination, altitude call-out, or fuel call-out. This is similar to how prior research has quantitatively determined the expectancy of instruments (Steelman-Allen, 2011; Wortelen et al., 2013a-c). The time between events was known from the experimental scenario (Table 6). In the experiment, these three event periods did not change much during one trial, and so the sampling periods are held constant in a simulation run. It was assumed that operators sample the secondary instruments at the Nyquist sampling rate (i.e. the sampling period of the instrument was half the period of the related task event).

Table 6. The desired sampling period on the secondary instruments is approximately half the rate of the associated event.

Instrument	**Event Period from the Experiment** (seconds between task)	**Desired Sampling Period in the Model** (seconds between fixations)
Comm Light	6.0±0.01	3
Altitude Indicator	7.1±0.03	3.5
Fuel Indicator	12.7±0.02	6.5

Flight Instruments – The event period that drives attention on the attitude and rate-of-descent indicators is the time that it takes the error between the actual and guidance-prescribed state to grow above a certain threshold. This threshold is the error that operators allow when they consider their closed-loop performance to be satisfactory. It is a function of the display resolution, the size of the displayed

signals, and the operator's control behavior. Given the lunar landing displays and controllers used in the experiment, these thresholds were set in the model at 1.5° for the attitude error and 1 fps for the RoD error. Experimental data confirms that subjects' steady-state flying errors were of this order of magnitude.

The time that it takes the attitude and RoD error to grow beyond threshold is a function of multiple factors: 1) the shape of the guidance-prescribed reference trajectory, 2) the perceptual noise in the attitude and RoD displays, and 3) the control remnant, which is the non-linear portion of vehicle and human behavior not captured by the linearized system. It is assumed that operators know the general shape of the reference trajectory and how it changes with time. This is an appropriate assumption for this experiment because subjects were given extensive training in which the reference trajectory did not change. If an operator does not know the reference trajectory or vehicle dynamics before flying the landing, or if the reference trajectory changes unpredictably, he will need to sample the attitude indicator frequently to gain an understanding of its behavior. Then, once the behavior is understood, he can re-adjust his sampling period. In principle, the sampling interval could be set by computing the autocorrelation function of the attitude error. As the attitude error becomes more random, the sampling period will have to decrease.

The operator can estimate the shape of the reference trajectory, the perceptual noise, and the control remnant, which allows him to set his sampling period. Figure 19 shows the guidance-prescribed pitch and roll for example simulation runs in Auto (with no mode transition), with and without a landing point redesignation (LPR). In the experiment, the reference trajectory rate-of-change was faster at the beginning of the trial than at the end (Figure 19, top). So, the sampling period of the attitude must be shorter. The presence of a landing point redesignation (LPR) 20 s into the trial also had an effect on the shape of the attitude reference trajectory (Figure 19, bottom). When there is an LPR, the reference trajectory rate-of-change was faster, and the desired sampling period of the attitude must decrease. The desired sampling period of the RoD is unaffected by the LPR.

Because of the design of the lunar lander displays and controls, subjects could sample the vehicle pitch and roll in the same fixation and make simultaneous control inputs. This coupling decreased the amount of attention needed to control the attitude. Because the pitch rate-of-change was higher than the roll rate-of-change, sampling the pitch with the desired period inherently satisfied the desired sampling period on the roll. If the pitch and roll were displayed separately, subjects would have to spend extra time sampling the roll indicator. This would leave less time to attend the RoD and secondary instruments, increasing mental workload and decreasing situation awareness.

The perceptual noise in the attitude and RoD displays can be estimated from knowledge of the display design, but the control remnant is difficult to model. For this reason, the desired sampling periods on the attitude and RoD were calculated from a dataset containing 60% of the experimental data (Section 5.2.2). Each trial was split into five 10-s windows, and the average attitude and RoD rates-of-change (i.e. the derivatives of the signals) were determined in each window. From this, the average time for the error to grow 1.5° (attitude) or 1 fps (RoD) was calculated.

Table 7 shows the desired sampling period ($T(i)$, measured in seconds between fixations) for the attitude and RoD indicators.

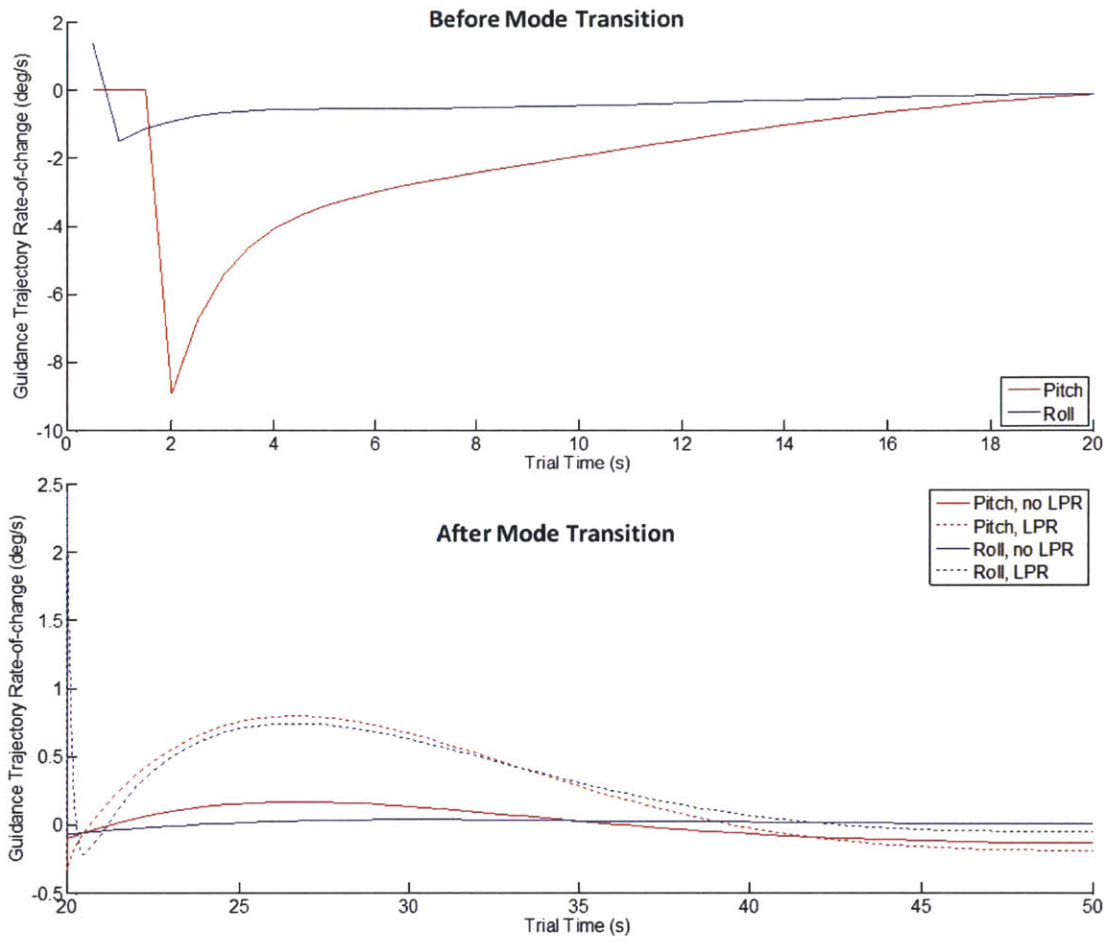


Figure 19. The shape of the reference trajectory changes with trial time, and with the presence of a landing point redesignation at the mode transition (20 s into the trial). Note that the x-axis and y-axis scales of the top and bottom graphs are different.

Table 7. The desired sampling period on the flight instruments changes with the trial time, based on the shape of the reference trajectory.

Trial Time (s)	Desired Sampling Period (seconds between fixations)		
	Attitude		RoD
	with LPR	without LPR	
0 – 10	0.5		4
10 – 20	1.5		4
20 – 30	1.5	3.5	4
30 – 40	3	4.5	6
40 – 50	4.5	4.5	7

5.2.2.4 Eye Movement Time

The model calculates the time required to move the eyes (μ, measured in ms) between two instruments i and j by the following equation (Carpenter, 1988):

$$\mu = 2.2\gamma(i,j) + 21 \tag{5.4}$$

In Equation (5.4), $\gamma(i,j)$ is the visual angle between the two instruments, as detailed in Table 5. Because the "other" instrument represents a group of locations on the display, there is no physical meaning of the visual angle between it and one of the task-related instruments. So, this visual angle was set to 18.7°, which is the average of the visual angles between all instrument pairs.

5.2.2.5 Dwell Duration on Instruments

When attention is placed on an instrument, it remains there for a specified dwell duration (Table 8). This dwell duration is randomly selected from a normal distribution, whose mean and variance are calculated from a dataset comprising 60% of the experimental data (Section 5.2.2). Randomly-selected dwell duration <100 ms are rounded up to 100 ms, the minimum fixation duration (Section 4.2.3). If the model fixates on the attitude indicator in a manual control mode (TA or TA+RoD), and the error between the actual and guidance-prescribed pitch and/or roll is above a prescribed threshold (1.5°, as described in Section 5.2.2.3) at the end of the dwell duration, attention remains on the attitude indicator until the errors fall below the threshold.

Table 8. Dwell durations (mean ± standard deviation) in ms computed from experimental data.

	Control Mode		
Instrument	**TA+RoD**	**TA**	**Auto**
Fuel	400±160	420±210	440±200
Attitude	530±320	600±450	350±190
Altitude	510±250	550±320	610±240
RoD	580±330	340±170	290±170
Comm	410±190	420±210	500±260
Other	220±200	230±260	260±230

5.2.2.6 Attention-based Measures of Mental Workload and Situation Awareness

As has been stated throughout this thesis, attention, mental workload, and situation awareness are closely related. Changes in the task structure can increase mental workload, which can force subjects to decrease attention on certain instruments. This, in turn, leads to a reduction in situation awareness. Mental workload is commonly assumed to be inversely proportional to the amount of spare attention – that is, the amount of attention that is not allocated to the primary task (the attitude and RoD indicators). Spare attention can be computed directly from the attention budget produced as the output of a simulation run.

As is stated in Section 5.2.2.3, the desired sampling period of a system state is chosen to give satisfactory performance on the associated task. It is assumed that this is also the sampling rate required for satisfactory situation awareness. If the state is sampled more frequently than desired, there is no additional benefit to situation awareness. However, if the state is sampled less frequently than desired, situation awareness is assumed to be lower than satisfactory.

The model estimates the situation awareness of a particular system state over a period of time by calculating the "average peak uncertainty" (APU). The "peak uncertainty" of an instrument is the uncertainty at the moment it is attended to (Figure 18). Averaging this peak value over a period of time (i.e. pre-, during-, and post-mode transition) tells whether the desired sampling period is being met in that period. If the APU is ≤1, then the sampling period is at (APU=1) or below (APU<1) the desired period and situation awareness is assumed to be satisfactory. If the APU is >1, the actual sampling period is longer than desired and situation awareness is assumed to be degraded.

5.2.2.7 Summary of Attention Block Parameters

Table 9 summarizes the parameters used in by the Attention block and how they were determined.

Table 9. Parameters used in the attention block.

Parameter			Determined by
Value	$v(I,m)$	For each instrument, in each control mode	Scenario design
Distances	$\gamma(I,j)$	Between all pairs of instruments	Scenario design
Desired sampling period	$T(i)$	On fuel, altitude, and comm	Scenario design / First principles (Nyquist sampling rate)
		On attitude and RoD, as a function of trial time	Experimental data
Threshold of operator control		For attitude and RoD	Experimental data
Weighting of each primitive	V, U, E	Of each attention primitive	Model fitting
Eye movement time	μ	Between instruments	Scenario design / Existing literature (Carpenter, 1988)
Dwell durations		For each instrument, in each control mode	Experimental data

5.2.3 *Attitude Perception, Estimation, and Control*

5.2.3.1 Attitude Perception and State Estimation

The Perceptual Noise and State Estimator blocks for the attitude indicator are structurally similar to the OCM, as described in Section 5.1.1.2. When attention is fixated on the attitude indicator, the State Estimator cannot directly perceive the exact vehicle pitch and roll. Instead, it must estimate the actual pitch and roll from a noisy measurement. To perform this estimation, the State Estimator employs a Kalman filter which represents the operator's mental model of the system: he understands the dynamics of the vehicle he is flying, how his control inputs influence these dynamics, and the characteristics of his perceptual noise. Operators use all this information to estimate the state of the vehicle. As Figure 16 indicates, these estimates are compared to the perceived attitude and the difference is weighted and used to adjust the state estimates. Once adjusted, operators make control inputs to the system based on these estimates.

The Kalman filter used in the Mental Model block is based on a linearized model of the actual lunar lander system dynamics (Appendix E). The operator inputs a joystick deflection (δ), which controls the vehicle pitch and roll rates through first-order lags. The rates are then integrated to give the pitch (θ) and roll (ϕ). Equations (5.5) and (5.6) show these transfer functions:

$$\theta = \frac{8.57}{0.12s^2 + s}\delta_\theta \tag{5.5}$$

$$\phi = \frac{9.68}{0.12s^2 + s}\delta_\phi \tag{5.6}$$

In the model, most of the operator remnant is assumed to arise from perceptual noise (Levison et al., 1969). The lunar lander in this model is assumed to be operating in a non-atmospheric environment free of turbulence, and operators would not expect to be subjected to a process noise due to environmental disturbances. System nonlinearities, such as the joystick deflection limits, could also contribute to the remnant. However, as Section 5.2.3.2 describes, most subjects chose a control strategy that did not encounter this nonlinearity. The perceptual noise represents the fact that subjects cannot perceive the actual pitch and roll of the vehicle, but instead simply estimate it to the nearest 5°. This value was chosen because the pitch ladder in the lunar landing simulator was marked at 10° increments, and it would have been easy for subjects to estimate whether they were at, or between, a marked decimation.

The Kalman filter uses the operator's knowledge of the system dynamics and the control input to propagate the vehicle state estimate from the previous time step forwards to the current time step. It then computes the error between this *a priori* estimate and the observation. The *a priori* estimate is updated by this error, multiplied by a Kalman gain matrix. The Kalman gain matrix is chosen to minimize the *a*

posteriori error between the estimated and the actual state. Essentially, this gain matrix describes how much the observation can be trusted. If the perceptual noise is known to be high, the *a priori* estimate will be weighted more heavily than the observation. If the perceptual noise is known to be low, then the observation will be weighted more heavily.

During a simulation, the Kalman gain is recomputed with a Ricatti solver in real-time over the duration of a fixation. This is different than prior optimal state estimation models, which use steady-state Kalman gains (Kleinman et al., 1971; Merfeld et al., 1993). These models assume that the operator continually perceives the system state and the uncertainty in the state estimate remains constant. In contrast, the model described in this thesis assumes the operator only perceives the attitude during visual fixations. As a result, the uncertainty in the operator's estimate of the attitude grows between fixations and is greatest when the operator first looks at the attitude indicator. Consequently, the Kalman gain is at its highest at this time (i.e. the internal estimate is not trusted much). The error between the estimated and actual state quickly decreases as the operator perceives the noisy state of the attitude indicator during a fixation. This is the "correction" in the state estimate. As the error decreases, the Kalman gain also decreases and the internal estimate is trusted more. This process occurs faster than the duration of a fixation.

For more details about the Kalman filter used in the model, please refer to Appendix E.

5.2.3.2 Attitude Control

The Action block models the control inputs made by the operator. This block receives the estimated vehicle state from the State Estimator block, compares it to the guidance-prescribed error, and makes a control input to reduce this error. All of these tasks – perception, estimation, and control – occur simultaneously. The Action block in this model was not based on the OCM, and therefore does not perform "optimal" control inputs. Rather, the Action block is based on the Crossover Model of operator control (McRuer et al., 1965), as reviewed in Section 5.1.1.1.

An operator's control input is a function of three parts: an operator gain, a first-order lag that represents the time required for neuromuscular activation, and a reaction time delay. The latter two are fundamental limitations of the human operator that affect the nature of his control strategy (Sheridan & Ferrell, 1974). In the Action block, the neuromuscular activation and reaction time are simplified and modeled as a combined time delay of 0.30 s. The lag included in this model is an approximation to the lead/lag that might be employed if higher-order terms of the integrated human-vehicle model were used.

The operator gain in the Action block can be varied to simulate different control strategies. Ideally, an operator should aim to null the attitude errors with the minimum amount of attention. In order to do this, the operator should increase his control gain to increase the speed of the system response to errors. However, increasing the gain too much leads to overshoot. If the vehicle overshoots the guidance-prescribed attitude by more than 1.5°, the HPM will have to spend additional attention correcting this

overshoot. So, selecting the operator gain is a tradeoff between maximizing response speed and minimizing overshoot. Furthermore, the selected gain must maintain system stability.

A linearized model of the pitch dynamics was used to identify the range of operator gains that give this desirable performance. Further details of this linearized model, and an explanation of how response speed and overshoot affect attention, are shown Appendix E. The pitch dynamics were the same as described in Equation (5.5). This transfer function was multiplied by the operator control gain, G, a first-order lag for the neuromuscular activation, and a first-order Pade approximation of the reaction time delay:

$$\theta = \left[\frac{8.57}{0.12s^2 + s} \cdot G \cdot \frac{1}{0.15s + 1} \cdot \frac{-s + 13.33}{s + 13.33} \right] \delta_\theta \tag{5.7}$$

Analyzing the system with the SISO (single-input/single-output) Design GUI in MATLAB (MathWorks; Natick, MA) showed that a gain of G=0.217 produced a 1.5° overshoot for a 7° step change (the approximate step-change in guidance that was caused by an LPR in experimental trials). This is the maximum gain that the human should employ. The analysis also indicated that a gain of G=0.092 gave the fastest system response with no overshoot. The open-loop Bode plots of these two gains show that the linearized system model has sufficient gain and phase margin for closed-loop stability. However, it is important to note that the linearized model is only an approximation, and the stability of the linearized model does not necessarily imply the stability of the complete, non-linearized model. Appendix E shows the open-loop root-locus and Bode plots of the linearized system with these operator gains.

One limitation of this linearized system is that it ignores the joystick deflection constraints present in the actual simulator. This can cause situations where certain large operator gains prescribe joystick deflections that are greater than physically allowable. Adding the joystick deflection constraints limits the maximum pitch rate of the vehicle, which decreases the amount of overshoot with large operator gains. An analysis finds that a gain of G=0.25 with the joystick limits will produce a 1.5° overshoot for the 7° step change. This is a 15% increase from the maximum gain of G=0.217 without the joystick limits.

With the joystick inputs limited to ±1, two different strategies of operator control emerge. One of them is "bang-bang" control where the operator controls the vehicle by deflecting the joystick to the extreme limits. This produces a rapid system response, but has a greater potential for overshoot. Only one experimental subject, #7, exhibited this behavior (Figure 20, left). Bang-bang control can be simulated in the Action block by using high values of the operator gain. The other control strategy is continuous control, where the operator makes smaller, continuous inputs to the joystick. This provides smoother control and less overshoot, but has a slower response time. The majority of subjects in the experiment exhibited continuous control, making smaller inputs (Figure 20, right). Continuous control is simulated in the model by smaller values of the operator gain.

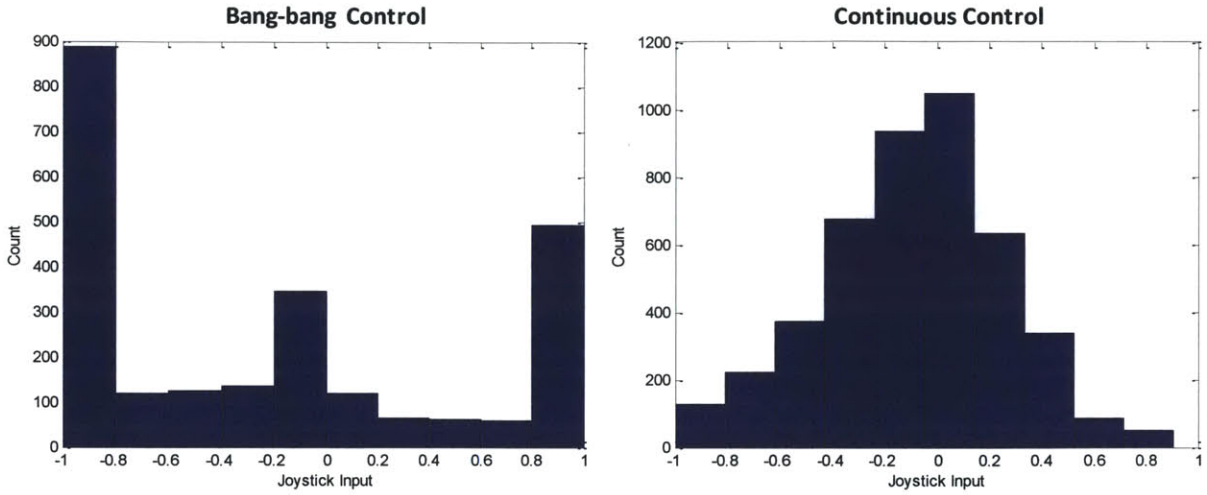


Figure 20. Subject 7 (left) exhibited bang-bang control – most joystick inputs were at the ±1 limit. Other subjects (#12, shown at right) exhibited smooth control and made small joystick inputs. Both histograms only consider non-zero joystick inputs.

5.2.3.3 Summary of Attitude Perception, Estimation, and Control Parameters

Table 10 summarizes the attitude parameters used by the Perceptual Noise, State Estimator, and Action blocks. It mentions whether the parameters were determined by engineering first principles of optimal state estimation and manual control theory, knowledge of the scenario design, or existing literature.

Table 10. Model parameters related to the attitude control task.

Block	Parameter		Determined by
Perceptual Noise and State Estimator	Estimated dynamics	Of vehicle pitch and roll	Scenario design / First principles
	Standard deviation	Of process noise	First principles
		Of measurement noise	Scenario design
Action (Manual Control)	Operator control gain		Scenario design / First principles
	Neuromuscular time delay		Existing literature (Sheridan & Ferrell, 1974)
	Reaction time delay		Existing literature (Sheridan & Ferrell, 1974)
	Threshold of operator control		Scenario design

5.2.4 *Rate-of-descent Perception, Estimation, and Control*

Whereas attitude control occurs in parallel (perception, estimation, and control all occur simultaneously) control of the rate-of-descent occurs in series. Control of the RoD happens after the perception and estimation have concluded for a fixation. This occurs because there is a substantial lag in the RoD dynamics, with a time constant on the order of 1 s. Experimental subjects were alerted to this lag, and were encouraged to not spend time watching the RoD after a control input. Instead, they were encouraged to fixate on another instrument and return their attention to the RoD later.

During the fixation on the RoD, the State Estimator block estimates how many control inputs are necessary. The RoD control system replicates the design of the Apollo lunar lander, where RoD must be commanded in ±1 ft/s (±0.3 m/s) increments (Duda et al., 2009). The RoD display in the simulator was demarcated in 2 ft/s (0.6 m/s) increments, and so it was assumed that operators could perceive the RoD to the nearest 1 ft/s (0.3 m/s) (i.e. they could determine if the RoD was at or between one of the demarcations).

The perceived RoD signal is highly trusted by the State Estimator block for two reasons. As was just stated, the perceptual noise is the same order of magnitude as the required control inputs. There is no need to estimate the actual state with more precision. Secondly, the display noise encountered in the experiment is added to the RoD signal (Appendix E). However, it was limited to ±0.25 ft/s (±0.08 m/s), four times smaller than the perceptual noise. Because the perceived RoD signal can be highly trusted, a Kalman filter is not necessary like it is for attitude Perception block.

Table 11 summarizes the RoD parameters used by the Perceptual Noise, State Estimator, and Control/Action blocks.

Table 11. Model parameters related to the RoD control task.

Block	Parameter		Determined by
Perceptual Noise and State Estimator	Noise	In the RoD indicator	Scenario design
	Quantization	Of RoD perception	Scenario design
Action (Manual Control)	Quantization	Of RoD control	Scenario design

5.2.5 *Fuel and Altitude Perception and Simulated Call-out Performance*

5.2.5.1 Fuel and Altitude Perception

When attending the fuel and altitude, the operator can only perceive a quantized version of the actual state. These quantizations are functions of the simulator displays and the system behavior. While the actual fuel level (% fuel remaining) in the lunar landing experiment was a decimal, the digital fuel display only presented it rounded down to the nearest integer. Therefore, subjects in the experiment could only perceive fuel in 1% increments. The digital altitude display also presented the actual altitude in feet to the nearest integer. However, the altitude decreased too quickly through much of the trial for subjects to perceive it at this resolution. In the model it is assumed that the ones digit of the altitude readout is unreadable, and that the operator relies on the tens digit to estimate the altitude. With the maximum rate-of-descent of -16 ft/s (-4.9 m/s), each value in the ones digit is visible for 63 ms. This is less than the minimum duration of a fixation and is too brief to perceive. However, each value in the tens digit remains visible for 625 ms, which is the same order of magnitude as the average experimental dwell duration on the altitude indicator (which is between 510 and 610 ms, depending on the control mode). For these reasons, it is assumed that the simulated operator is able to perceive the altitude in 10-ft (3-m) increments.

The altitude quantization is assumed to be the same throughout the trial, but it would be possible to change this and have the quantization as a function of the rate-of-descent. This assumes that as the rate-of-descent decreases, the operator can estimate the altitude with more precision and the quantization step size of the altitude estimation decreases as well. For example, at a rate-of-descent of -4 ft/s (-1.2 m/s) (which is achieved near the end of a trial) the ones digit is visible for 250 ms, which is more than twice the minimum length of a fixation. This may still be too fast for the operator to perceive the altitude in 1 ft-increments (0.3 m), but his perception should be better than 10-ft (3-m) increments. However, because the call-outs occur in 100- or 50-ft (30.5- or 15.2-m) increments, more precise perception is not required for the tertiary task.

5.2.5.2 Simulated Call-out Performance

The HPM uses its knowledge of the fuel and altitude state to make call-outs at predetermined intervals. Verbal call-outs of vehicle state have been used as a real-time measure of situation awareness with high face validity in a number of previous experiments (Chapter 4; Hainley et al., 2013; Kaderka, 2014). In the model, altitude call-outs occur every 100 ft (30.5 m) down to 500 ft (152.4 m), and at 50-ft (15.2-m) intervals thereafter, and fuel call-outs occur at 5% increments. If the Attention block fixates on the altitude or fuel indicator within ±1 quantization unit (10 ft (3 m) for the altitude, 1% for the fuel) of a call-out, this call-out is successfully "made." If a call-out passes without being fixated on, it is "missed." The percent of call-outs made gives a measure of situation awareness on these two indicators that can be compared to similar experimental data. The higher the response percent, the greater the operator's situation awareness.

5.2.5.3 Summary of Fuel and Altitude Parameters

Table 12 summarizes the fuel and altitude parameters used by the Perceptual Noise and Action blocks:

Table 12. Model parameters related to the call-out task.

Block	Parameter		Determined by
Perceptual Noise and State Estimator	Quantization	Of fuel and altitude perception	Scenario design
Action (Call-out Response)	Timing	Of fuel and altitude call-outs	Scenario design
	Acceptable range	For "making" a call-out	Scenario design

5.2.6 *Comm Light Response Task*

The model simulates operator response to comm light illuminations to provide a measure of spare visual attention and therefore a proxy for mental workload. As with the simulated call-out performance, the model-predicted mental workload can be compared to similar experimental measures. As in the experiment, the comm light "illuminates" at a random time within the first 2 s of a 6-s window. These illuminations occur throughout the trial. This timing of the comm light illuminations is determined by the scenario design, and is the only model parameter related to the comm light response task.

If the Attention block fixates on the comm light while it is illuminated, it is assumed to be noticed and the response considered as successfully "made." The model records this success, along with the time since the beginning of the illumination. If no response is made to the comm light, it "extinguishes" after the 6-s window ends and is recorded as "missed." Unlike the experiment, there is no regard to the color of the comm light in the model. The time that experimental subjects spent determining the color is already captured by the dwell duration on the comm light. Furthermore, subjects responded incorrectly to less than 1% of the comm lights in the experiment.

5.3 Validation of the Attention Block against Experimental Data

Ideally, the attention budgets produced by the HPM would be validated against the results of an independent experiment. Such data was not available, and so the simulated attention budgets were validated against a data set containing 33% of the experimental trials (Section 5.2.2). Simulations were run with six different sets of value, uncertainty, and effort (VUE) weightings (Table 13).

Table 13. VUE weightings used in the Attention block validation.

Value	Uncertainty	Effort
1	1	1
1	0	0
0	1	0
1	1	0
2	1	1
1	2	1

These six weightings were selected from a larger possible set in order to explore the sensitivity of the model to the attention primitives. Holding all primitives equal (VUE = 111) gives a baseline performance. Simplified models were run with only the value (VUE = 100), only the uncertainty (VUE = 010), and without the effort (VUE = 110). A model with only effort was not considered because prior research found that value and expectancy (which is similar to the uncertainty) were more influential than effort in determining attention allocation (Wickens et al., 2005; Steelman et al., 2011). For this reason, the value and uncertainty weightings were always ≥ the effort weighting. The final two models explored the importance of the value and uncertainty parameters. Each was weighted twice as much as the other primitives (VUE = 211 and VUE = 121, respectively).

For each set of VUE weightings, 5 simulation runs were performed for each of the 12 mode transitions (the 6 mode transitions (MT) specified in Table 1 of Chapter 4, each with and without a landing point redesignation). The attention budgets for each run were averaged by mode transition, phase (pre-MT, during-MT, or post-MT), and instrument. Then, these averaged attention budgets were each compared to experimental data under the same conditions. This gave 36 matching pairs (12 MTs * 3 phases) for each of the 6 instruments. The mean square error (MSE) between the model and experimental attention budgets is shown in Figure 21 for each set of VUE weightings.

As Figure 21 shows, the different VUE weightings gave a similar error for most instruments. In many cases the standard error bars overlapped, suggesting that there are few statistically significant differences between weighting sets. This implies that the other Attention block parameters detailed in Table 9 have a strong influence on the attention budget produced by the model. Because these parameters were determined independent of the model fit, they were not varied in a sensitivity analysis.

The most prominent trend of the different VUE weightings in Figure 21 relates to the MSE in the predicted attention on the comm light and the fuel indicator. These instruments are located on the extreme left and right edges of the display (Figure 4) and it takes relatively more effort to move the eyes to these instruments. The attention budget MSE on these instruments was greatest when effort was weighted equally to value and uncertainty (VUE=111). This suggests that subjects' attention was not impeded by high-effort movements to the comm light and the fuel indicator if the value and uncertainty of this information was high. The trend is congruent with prior experimental findings performed in different scenarios (Wickens et al., 2005; Steelman et al., 2011).

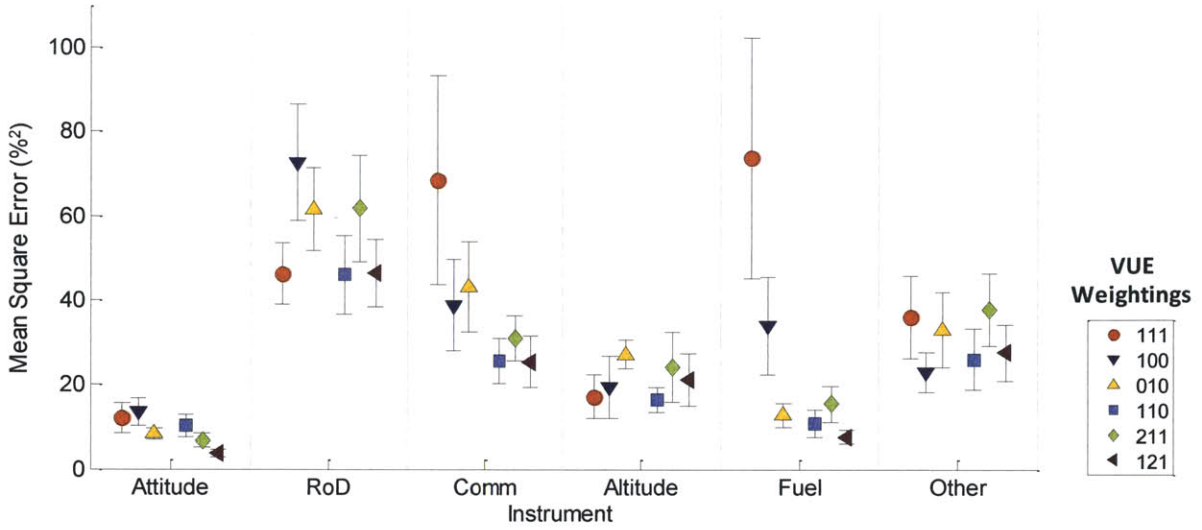


Figure 21. Mean square error between the experimental and simulated attention budgets for five different sets of VUE weightings. Data points shown have been averaged over all mode transitions and phases. Error bars indicate standard error. The units of MSE are the percent of attention allocated to an instrument, squared (not a percent difference).

For the remainder of the analyses in this chapter, a weighting of VUE=121 was chosen because it gave the smallest or nearly-smallest MSE across all instruments. Furthermore, the uncertainty primitive is the new parameter developed for this model, and it was expected that it play a significant role in the determination of attention, task performance, and situation awareness.

Figure 22 shows a scatter plot of all 216 matched pairs (12 mode transitions * 3 phases * 6 instruments) for the runs with VUE=121. The diagonal black line indicates a perfect match between the experimental and simulated data. As Figure 22 shows, the data is generally clustered close to the black line, representing a good match between the simulated and experimental data. There are some systematic biases: the attention on the RoD indicator is overestimated (most of the data points appear above the black line), and the attention on the altitude and "other" instruments are underestimated. However, the average difference between the experimental and simulated data (across levels of mode transition and phase) is small for all instruments (Attitude: 1.4%±1.1; RoD: 3.6%±0.5; Comm light: 0.4%±0.5; Altitude: -3.3%±0.6; Fuel: 0.5%±0.3; Other: -2.5%±0.8).

The poorest fit between the experimental and simulated data occurs in a manual control mode before the mode transition. This causes the large overestimation of attitude seen at the top right of Figure 22. This overestimation occurred regardless of the VUE weighting, as Figure 23 shows.

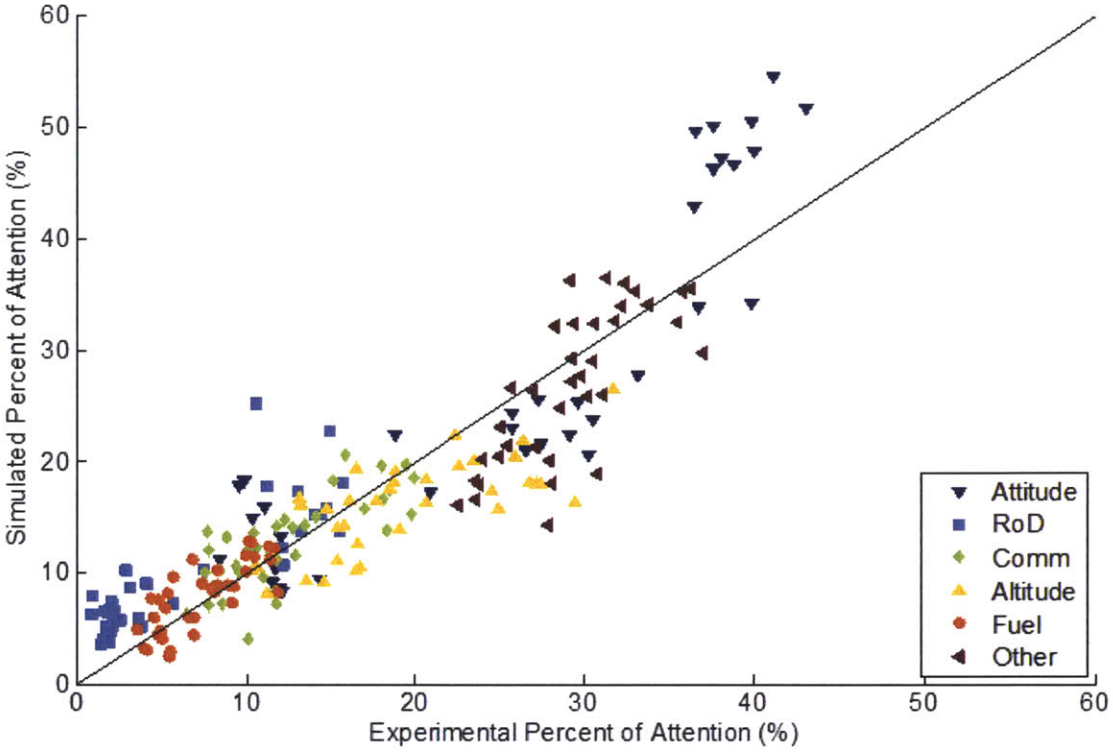


Figure 22. Comparison between simulated and experimental attention budgets when the uncertainty is weighted twice as much as the value and effort (VUE = 121).

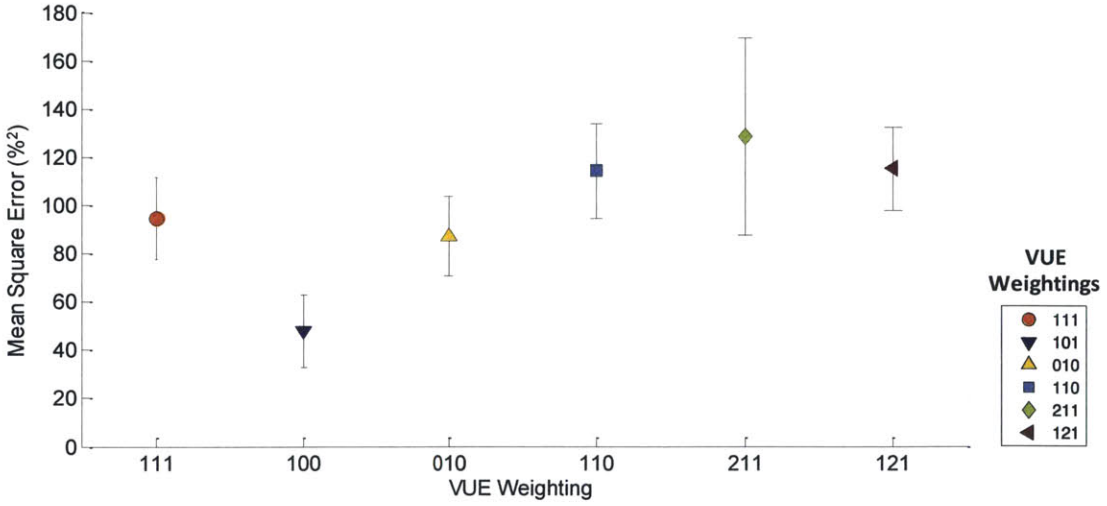


Figure 23. The percent of attention on the attitude indicator in a manual mode before the mode transition is overestimated for all VUE weightings. Error bars indicate standard error. The units of MSE are the percent of attention allocated to an instrument, squared (not a percent difference).

During this phase the reference trajectory rate-of-change is high, and the HPM must allocate a large amount of attention to the attitude indicator. One possible reason for the model's overestimation in these conditions may be a difference between the simulated control gain and the actual gain employed by the experimental subjects. Subjects may have used a higher gain when nulling these errors, leading to a faster convergence of the actual and guidance-prescribed pitch. This would have required less attention on the attitude indicator. The model also assumes that subjects kept attention on the attitude indicator while making control inputs. If the error was large and subjects felt comfortable with controlling the simulated lunar lander, they may have glanced away from the attitude indicator while continuing to make control inputs. This also would have lessened the amount of attention on the attitude indicator.

Appendix F shows that simulated attention budgets with VUE=121 show similar changes in attention allocation across all experimental control mode transitions.

5.4 Situation Awareness and Workload

As discussed earlier, the model makes predictions of operator mental workload and situation awareness in two ways. Based on the formal definitions of mental workload and satisfactory task performance (cognitive first principles), workload is measured by the spare attention not allocated to the flying task and situation awareness is assumed to be proportional to the uncertainty in the estimate of a system state (Section 5.2.2.6). The performance-based measurements replicate the same metrics used in the experiment: responses to a comm light illumination (Section 5.2.6) are used to predict mental workload and verbal call-outs of vehicle state (Section 5.2.5.2) are used to predict situation awareness.

5.4.1 *Relationship between Spare Attention and Performance-based Measurements of Mental Workload and Situation Awareness – Experimental and Model-Predicted Data*

Spare attention and mental workload/situation awareness task performance were all measured during the experiment, and so the model predictions can be compared to this data. Figure 24 and Figure 25 relate the performance-based measures of workload and situation awareness (y-axes) to the spare attention (x-axes) for both the experimental and model-predicted data. The experimental results were taken from the validation data set containing 33% of the experimental data, as described in Section 5.2.2. In both figures, note that spare attention is plotted so it is greatest on the left side of the x-axes.

Comparing experiment results and model predictions, performance on the comm light response task and call-outs is generally similar: As the spare attention decreases (from left to right on the x-axis), more fuel and altitude call-outs are missed and the comm light response time increases. The number of responses to comm lights shows a different trend, as it remains relatively constant until the spare attention reaches approximately 50-60%. It is only at the lowest levels of spare attention that the comm light response percent falls. The increase in mental workload has an effect on call out response latency time, but it is not enough to make the operator completely abandon this task at times.

While the general trends in Figure 24 and Figure 25 are the same, there are some differences. The model under-predicts task performance and spare attention at the beginning of a trial (the red "pre-MT" data points) in any control mode. This is believed to be a product of the over-prediction of attention on the attitude indicator during this phase (Section 5.3). As a result, there is less attention for the secondary and tertiary tasks. The model also over-predicts fuel call-out performance during mode transitions to TA and TA+RoD and comm light response time and altitude call-out performance during mode transitions to TA+RoD with an LPR. During these mode transitions subjects gained task(s) (the attitude and/or the RoD control task), sometimes at the same time as a change in the guidance-prescribed reference trajectory. It is possible that subjects temporarily changed their visual scan pattern in these high-workload situations and made unpredictable allocations of attention. Such behavior could not be predicted by the model.

5.4.2 *Relationship between Average Peak Uncertainty, Spare Attention and Performance-based Measurements of Mental Workload and Situation Awareness – Model-Predicted Data*

There is no way to measure the uncertainty in an experimental subject's estimates of system states, and so the model's prediction of situation awareness through the average peak uncertainty (APU) cannot be directly compared to similar experimental data. However, the model predictions can be inspected to see if they follow the expected behavior: an increase in the APU should be correlated with decreases in the amount of spare attention and situation awareness and an increase in mental workload.

Figure 26 and Figure 27 show how the APU for the comm light, altitude indicator, and fuel indicator correlate with the amount of spare attention and the respective task performance. On Figure 26, the y-axes have been reversed so that a higher APU (i.e. a longer sampling period than desired) is towards the bottom. These figures show the expected behavior: an increase in the APU is correlated with a decrease in the spare attention, and a decrease in task performance is correlated with an increase in the APU.

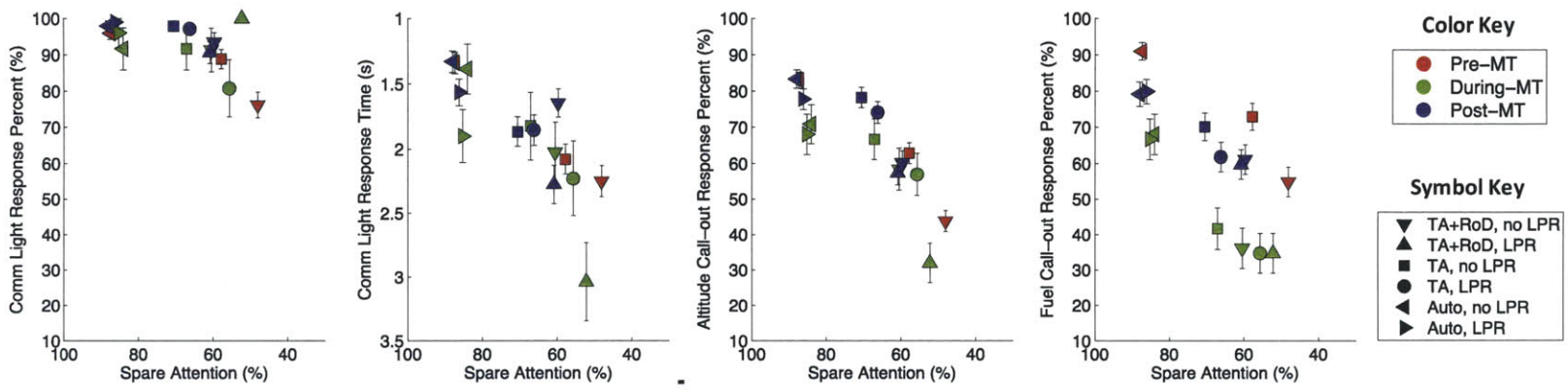


Figure 24. Experimental data showing the relationship between comm light response task/call-out performance and the amount of spare attention.

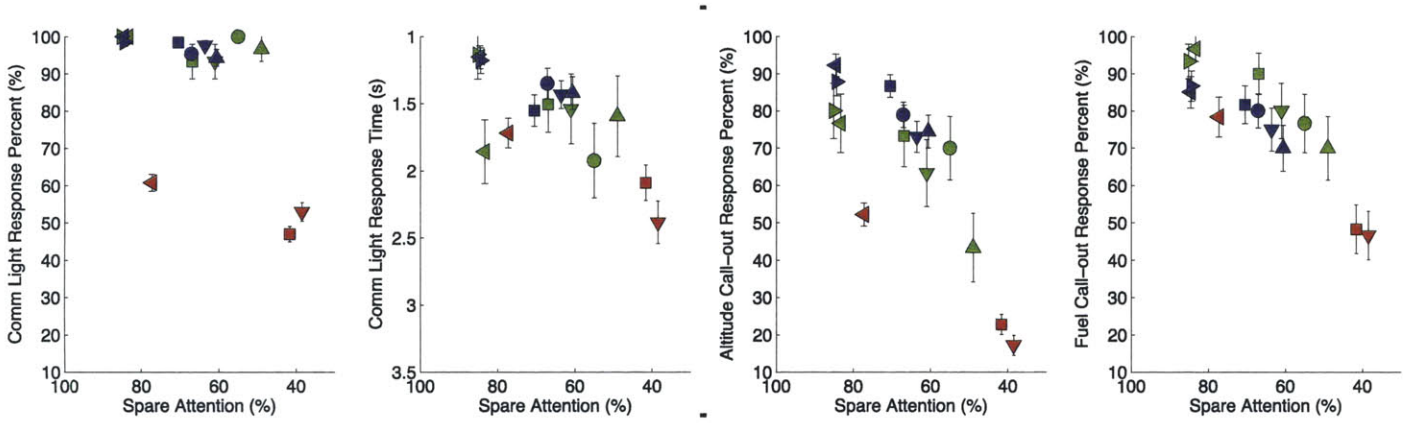


Figure 25. Model-predicted data showing the relationship between comm light response task/call-out performance and the amount of spare attention.

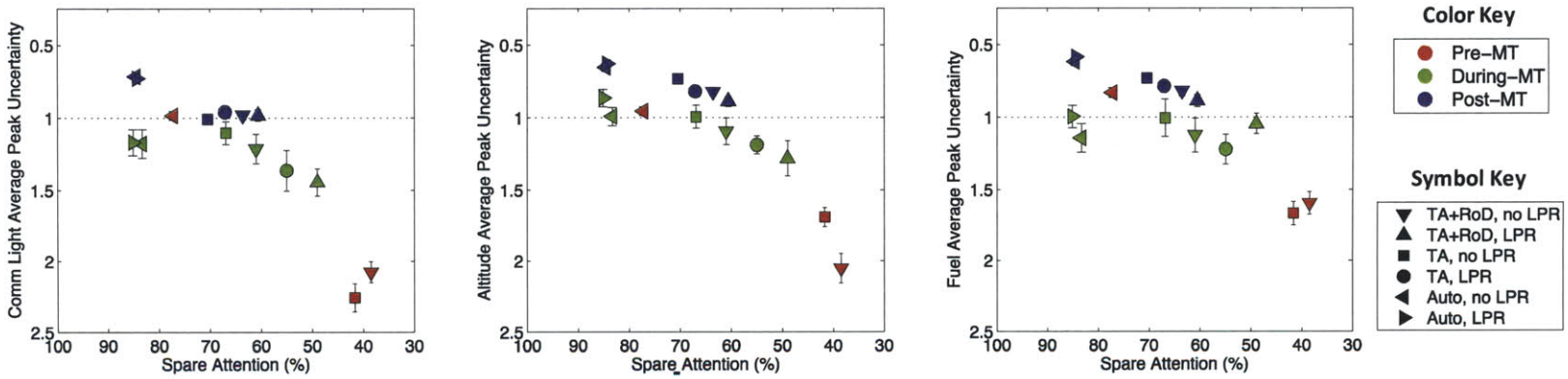


Figure 26. Model-predicted data showing the relationship between the average peak uncertainty on the comm light, fuel indicator, and altitude indicator and the amount of spare attention.

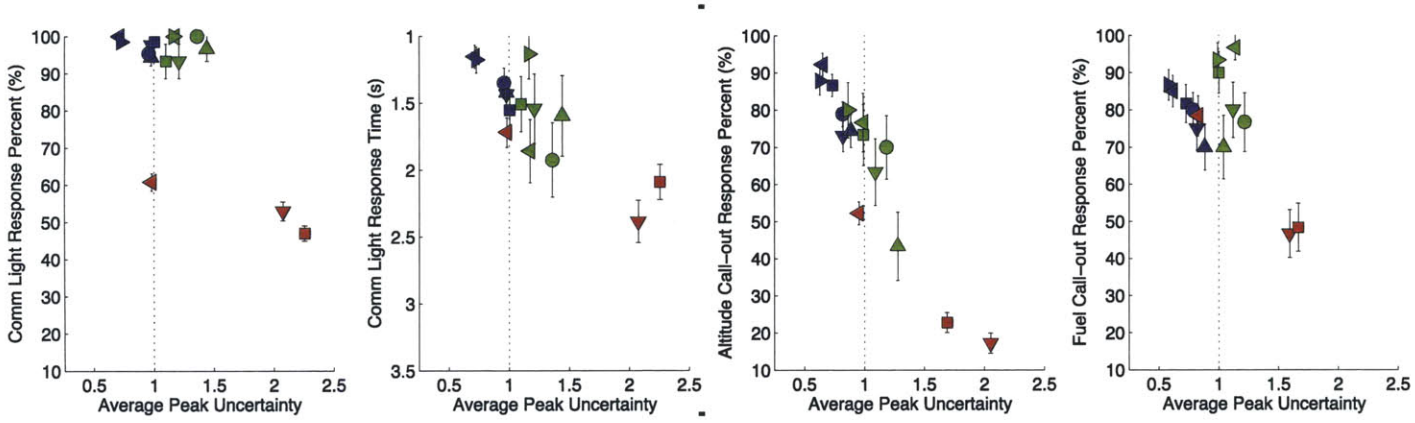


Figure 27. Model-predicted data showing the relationship between the comm light response task/call-out performance and the average peak uncertainty.

5.5 Flying Performance

The model described in this chapter does not only prescribe attention; it is a closed-loop model that is able to predict the effect of attention on an operator's flying performance. Figure 28 shows how the simulated performance on the flying task (measured by the MSE between the actual and guidance-prescribed vehicle state) compared to the experimental data from the 33% validation dataset. Each data point represents matched conditions (phase and control mode, with or without LPR) between the experimental data (on the x-axis) and the simulated data (on the y-axis).

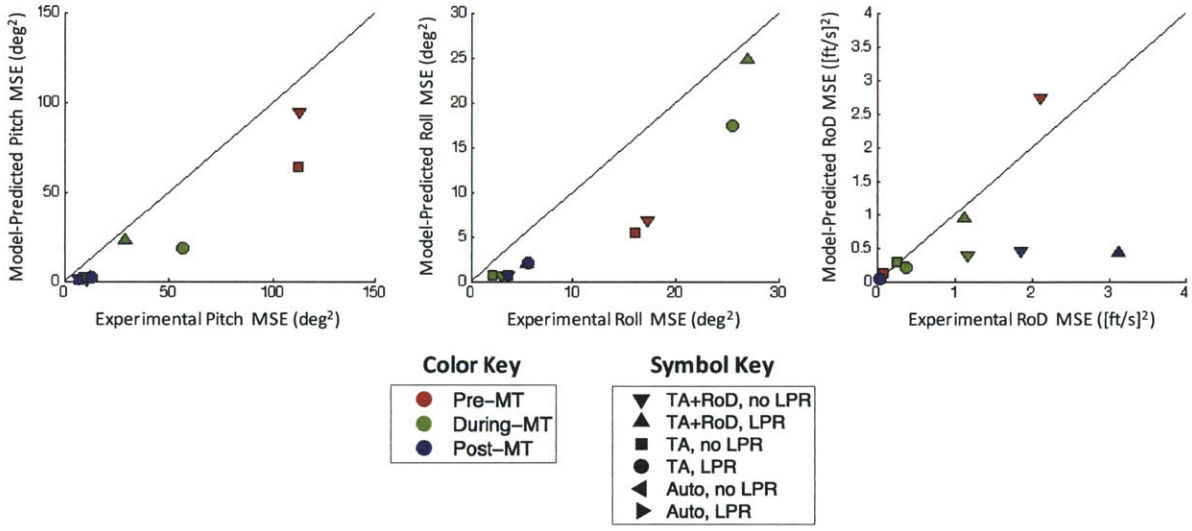


Figure 28. MSE between the actual and guidance-prescribed vehicle state, as compared to the autopilot MSE in the same phase.

As can be seen in Figure 28, the model performs the flying task with a smaller MSE than the experimental subjects did. Differences in the pre-mode transition phase may have occurred because this measure is sensitive to the time required to make the first control input at the beginning of the trial. The error between the actual and guidance-prescribed attitude is very large at the beginning of the trial because the guidance-prescribed pitch is 24° higher than the actual pitch (45° vs. 21°) and the guidance-prescribed roll is 6° higher than the actual roll (6° vs. 0°). If the operator delays in nulling these errors, the MSE quickly builds. In the simulated data set used in Figure 28, the first fixation on the attitude indicator was made less than half a second into the trial (0.44±0.04 s). The model responds to this error so quickly because the attitude indicator is the top priority and the probability of fixating on it (Equation (5.2)) is high at the beginning of a trial. While subjects knew that they had to null attitude errors right from the beginning of the trial, it is likely that they took longer than a half second to get acclimated to a new trial. Future work could investigate the time it took subjects to make their first control input and include this as a "trial initialization" period in the model. During this time, the HPM could be prevented from making control inputs.

The model's quick response to attitude errors is likely responsible for the model's under-prediction of MSE during mode transitions with an LPR present. The LPR introduces a step-change in the error between the actual and guidance-prescribed pitch and roll, which must be nulled quickly. As in the beginning of the trial, the probability of fixating on the attitude indicator is high. This leads to a fast response to the step-change in guidance, and a lower MSE.

Differences between the experimental and simulated data in other phases and control modes may also indicate that subjects' tolerance of errors was larger than the 1.5° pitch and roll and 1 ft/s RoD thresholds used in the model. Future work could devise experiments to fully characterize the size of the threshold at different parts in the experimental trial. These values could then be incorporated into future versions of the model.

5.6 Conclusion

This chapter presents a closed-loop vehicle-operator model designed to help better understand how the changes in operator attention and mental workload affect task performance and situation awareness. Specifically, the model simulates how these parameters change across control mode transitions where the priority of tasks changes and control loops shift from monitoring to manual control.

The human performance model employed in this model follows the stages of information processing as outlined by Wickens & Hollands (2000). The Attention block determines the instrument that is being fixated at each time step. Like the SEEV model of Wickens et al. (2003), the attention process is determined by the value of each instrument to the current task structure (which is dictated by the control mode) and the effort required to move one's eyes to each instrument. The effect of system state bandwidth is handled differently than in SEEV models. This model simulates growing uncertainty in the estimate of each system state between visual samples, which is assumed to depend on the rate-of-change of the displayed signal (i.e. both amplitude and bandwidth). This directly links attention and situation awareness, which is assumed to be inversely proportional to the uncertainty of a state estimate. The Attention block stochastically simulates visual fixations on various instruments and estimates the visual attention budget and fixation rate using Monte Carlo techniques. Parameters in the attention block were developed in part with a dataset containing 60% of the experimental data and then validated against a separate dataset with 33% of the experimental data. There was good agreement between the experimental and model-predicted attention budgets in models where the effort primitive was weighted lower than the value, uncertainty, or both.

The Perceptual Noise block simulates an operator's inability to perfectly estimate the actual state of the vehicle (i.e. the "situation" of the vehicle). The State Estimator block describes how the operator estimates the actual state of the vehicle from the noisy perceived value using a Mental Model of the system dynamics and system states rates-of-change. It mimics the decrease in situation awareness that occurs at high mental workload levels. Finally, the Action block makes pitch, roll, and rate-of-descent

control inputs to the vehicle, and responds to simulated fuel and altitude call-outs and comm light illuminations. Control gain is assumed to be chosen to minimize closed-loop response overshoot.

The key strength of this model is that it is a closed-loop model that concurrently simulates the operator and vehicle. Beyond quantitatively modeling visual attention, it predicts the effect of a given attention allocation on mental workload, situation awareness, and flying performance. In the model, mental workload and situation awareness are measured two ways: through cognitive first principles and task performance-based metrics that can be compared to experimental results. Mental workload is measured by the amount of spare attention not allocated to the flying task and the performance on a secondary comm light response task. Situation awareness is measured by the average peak uncertainty when an instrument is fixated on and the performance on simulated call-outs of vehicle fuel and altitude.

This chapter compared the comm light response performance and call-out performance to the amount of spare attention for both experimental and model-predicted data. There was good qualitative agreement between the two, with task performance decreasing in both as spare attention decreased. There were differences, primarily in how the model under-predicted performance and spare attention at the beginning of a trial. This chapter also showed that the average peak uncertainty of an instrument increased as the spare attention decreased, and the comm light and call-out task performance decreased as the average peak uncertainty increased. Finally, this chapter showed that the model underpredicted the MSE between the actual and guidance-prescribed attitude, perhaps because of this metric's sensitivity to large errors at the beginning of a trial and at the time of an LPR.

5.7 Limitations and Future Work

The generalizability of the particular model outlined in this thesis is low because many of the parameters are tied to the experimental scenario or experimental data. However, the modeling methodology of uniting the SEEV Model of attention (Wickens et al., 2003), the state estimation and prediction components of the Optimal Control Model (Kleinman et al., 1971), and the Crossover Model of multi-loop manual control (McRuer et al., 1965) is generalizable to other vehicles and scenarios that feature multi-loop manual control. Future researchers investigating these other domains can populate a similar model structure with information about the task, control modes, and displays to determine how attention and mental workload affect situation awareness and task performance. These investigations will help to highlight emergent behaviors resulting from the human-automation integration that require further study.

Another limitation of this modeling effort is that one experimental data set was used to both set parameters in the model and validate the model's Attention block. Future work could validate the model against separate experimental data. This experiment could be similar to the one described in Chapter 4, but with a different set of subjects and/or slightly different conditions. Or, the experiment could involve a completely separate vehicle and task set – a pilot flying approach and landing in a commercial aircraft, for

example. Validating against data taken in a completely separate domain would increase the generalizability of the model, but would require researchers to re-specify all the parameters in the model.

In the model, it was assumed that subjects kept attention on the attitude indicator while controlling pitch and roll. However, experienced operators who are more comfortable controlling the vehicle may be able to fixate on the attitude display, perceive the error, and then look away while making control inputs. The current model is easily adaptable to this modification because it contains blocks for both attention and manual control. If control inputs are continually made between discrete fixations, the sampling period acts as an additional time delay in the linearized system. Including the first-order Pade approximation of this time delay into Equation (5.7) gives:

$$\theta = \left[\frac{8.57}{0.12s^2+s}\cdot G\cdot\frac{1}{0.15s+1}\cdot\frac{-s+13.33}{s+13.33}\right]\cdot\left[\frac{-s+2T}{s+2T}\right]\delta_\theta \tag{5.8}$$

where T is the sampling period as in Equation (5.3). The presence of this new term reduces the maximum gain (G) that the operator can provide without too much overshoot. Conducting a similar analysis as in Section 5.2.3.2, the maximum gain for a 1.5° overshoot on a 7° step-change in the guidance-prescribed pitch is 0.09. This is less than half the maximum gain for the same overshoot when the operator is only controlling the attitude when looking at the indicator (G=0.217). Future work could use the desired sampling periods on the attitude indicator (Table 7) and knowledge of the reference trajectory rate-of-change throughout the trial (which gives the effective size of a guidance-prescribed pitch step-change) to determine the necessary operator gain throughout a trial.

A final limitation of the current model is that the uncertainty parameter, $u(\Delta t(i),T(i))$, is fixed prior to simulation runs. In future iterations the Mental Model could include memory of the error between the actual and estimated states, which could be used to adapt the uncertainty parameter in real-time during a simulation run. If the error in a particular state is large every time the operator looks at the related instrument, he knows that his Mental Model does not match reality. This implies one of two scenarios: either the Mental Model is inaccurate, or the Mental Model is accurate and the system is behaving abnormally. An inaccurate Mental Model may be the result of complex, non-linear system dynamics that are not easy for the operator to approximate. Or, the operator may be inexperienced and not have had enough training to fully understand the dynamics. Abnormal system behavior may the result of a system failure (see Kaderka, 2014 for a discussion of failure detection and diagnosis in the same lunar landing environment), or the operator may have missed a task being re-allocated from the automation to manual control. Whatever the reason for the increased error, the operator must sample the related instrument more frequently to reduce the uncertainty in the estimate. By allowing the uncertainty parameter to vary in response to the error, this behavior can be represented in the model.

This page intentionally left blank.

There has been much knowledge accrued.
Hypotheses, posed and pursued.
This, along with some text,
About steps to take next
Are put forth as the thesis concludes.

6.0 Conclusions

6.1 Thesis Executive Summary

Modern complex aerospace systems employ flight deck automation to increase the efficiency and safety of systems while reducing operator workload. However, too much automation can lead to overtrust, complacency, and a decrease in operator situation awareness. In an attempt to prevent these from occurring, the operator and the automation often share responsibility for performing tasks. The tasks allocated to each agent are rarely fixed; instead, they can be dynamically re-allocated throughout operations.

This thesis has been structured around a fundamental hypothesis concerning how a human operator responds to this dynamic task allocation, presented in Chapter 1 and replicated here:

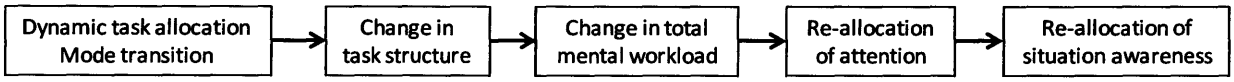


Figure 29. Chapters of this thesis have been structured around this proposed hypothesis describing operator response to dynamic task allocation.

It is important for system designers to understand these operator behaviors in order to design task re-allocations that occur "gracefully," meaning that they preserve performance and situation awareness while minimizing workload (Hainley et al., 2013). When tasks are re-allocated, the operator has a new set of tasks to perform. Performing each of these tasks at a desired level requires a minimum amount of cognitive and attention resources, which is called the mental workload (O'Donnell & Eggemeier, 1986). If the mental workload of all tasks together exceeds the operator's limited capacity, he must sacrifice attention on lower-priority tasks in favor of higher-priority ones. If attention falls below the minimum requirement, this decrease can lead to a noticeable decrease in task performance.

Operators continuously estimate the states of the entire system in which they operate, and this knowledge is called their situation awareness. These estimates are based on a mental model of the system dynamics and the system states' rates-of-change. These estimates all have an uncertainty, which grows between fixations as the estimate is projected farther into the future. The rate at which this uncertainty grows is a function of the system state's rate-of-change and the accuracy of the operator's mental model. Visual fixations on the display instruments give opportunities to "correct" these estimates with knowledge of the actual state. When attention is reduced on a lower-priority instrument in response to a task re-allocation, there are less frequent opportunities to correct the estimate of this state. As a result, the average uncertainty in the estimate grows and the situation awareness decreases.

This thesis begins by discussing the first part of the hypothesis detailed in Figure 29: the ways in which the task allocation can change. Dynamic task allocation is often adaptive, meaning that it occurs in

response to changes in the state of the system, environment, or operator. Dynamic task allocation can also be workload balancing, meaning that the task-reallocation occurs with the intent of keeping the operator in control as much as possible while remaining at a moderate level of mental workload. There are two specific forms of workload-balancing, adaptive dynamic task allocation: adaptive automation and adaptable automation. In the former, the automation has the decision authority to re-allocate tasks; in the latter the operator has this authority. There are also intermediate levels, in which both the automation and operator have varying degrees of authority. When adaptive automation is present in a system, it requires a triggering system to determine when to re-allocate tasks. Most commonly, the task re-allocation is triggered on the operator's mental workload, either self-assessed through subjective scales, assumed from their performance on the primary or secondary task, estimated from psychophysiological measurements, or determined by a human performance model that synthesizes multiple different measurements.

Much of the academic research literature is enthusiastic about adaptive automation and has focused on investigating its benefits and developing the enabling technologies and techniques. However, in practice, the pilot almost always remains the sole decision authority during real-world aviation operations. This thesis performed a hierarchical task analysis (HTA) of approach and landing in the Boeing 767 to investigate if and how dynamic task allocation was implemented. This HTA found that automation makes the landing more efficient and predictable, but it is not required for a safe landing. The pilots have the ability to turn the automation on and off, but the automation never decides to take over a task on its own. This is true even during unanticipated high-workload events, like a go-around. The procedure can be pre-programmed into the automation, but it is still the pilot's responsibility to push a button allowing the automation to perform the go-around.

Newer commercial aircraft contain adaptive systems, but these are not necessarily examples of adaptive automation. For example, flight envelope protection is an adaptive system that deters (in Boeing aircraft) or outright prevents (in Airbus aircraft) a pilot from exceeding the speed and attitude limits of the aircraft. Boeing flight envelope protection is not an example of adaptive automation because there is no re-allocation of tasks from the pilot to the automation. Instead, it is an example of adaptive control, in which the automation modifies the effect that the operator's control inputs has on the control surface deflections. Airbus flight envelope protection is not adaptive automation because the intent of this action is to prevent damage to the aircraft, and not to balance the operator's workload. Adaptive automation associate systems have also been developed and tested in military aircraft and rotorcraft, but they have not been operationally implemented to the best of this operator's knowledge.

This thesis hypothesized three reasons why aircraft manufacturers and operators have not fully embraced adaptive automation: 1) difficulties in fully testing and evaluating adaptive automation, 2) manufacturers' desire to make the pilot the ultimate authority, and 3) the reluctance of operators to trust adaptive automation. The remainder of this thesis addressed the first point and discussed the impact that dynamic task allocation has on human operators' visual attention, mental workload, situation awareness, and task performance through experimental investigations and closed-loop modeling.

In the experimental study, twelve subjects flew mode transitions between three unique control modes: 1) Two-axis (pitch and roll) and rate-of-descent manual control (TA+RoD), 2) Two-axis manual control with automatic rate-of-descent control (TA), and 3) Automatic control of attitude and rate-of-descent (Auto). A landing point redesignation (LPR) occurred during half of these mode transitions, causing a change to the guidance-prescribed reference trajectory. Subjects' visual fixations on the simulator displays were recorded using an eye tracking system. Five areas of interest were specified on the display: 1) the attitude indicator, 2) the rate-of-descent (RoD) indicator, 3) a comm light, 4) the altitude indicator, and 5) the fuel indicator. Measures of visual attention were compared to measures of mental workload, assessed by subjects' performance on the secondary task of responding to the comm light illuminations, and situation awareness, assessed by subjects' performance of tertiary verbal call-outs of the vehicle fuel and altitude at predetermined intervals.

Data analyses investigated how attention was re-allocated across three specific mode transitions by calculating the "attention budget," or the percent of attention that was allocated to each instrument in a given phase of the trial. The results confirmed the findings of prior work done on the same project that found mental workload and situation awareness to be dependent upon the number of manual control loops (Hainley et al., 2013). Furthermore, it extended this prior work by identifying other factors that also influenced the attention allocation.

Attention on the attitude indicator significantly increased across mode transitions from Auto→TA without an LPR, while attention on the fuel and comm light significantly decreased. This decrease in attention corresponded to a statistically significant decrease in fuel call-out performance, but not a significant change in comm light response time. It was suggested that attention decreased on the fuel indicator – but not the altitude indicator, an instrument of equal priority – because of the relative difference in the location of the instruments on the display and the timing of the call-outs. The fuel indicator was located on the side of the display apart from other instruments, while the altitude indicator was located in the middle of the display close to the flight instruments (attitude and RoD). Furthermore, the altitude call-outs occurred approximately twice as often as fuel call-outs.

Unexpectedly, attention on the attitude indicator significantly decreased across TA→TA+RoD and TA+RoD→TA mode transitions, even though the attitude control task was not re-allocated to or from the subjects. It was suggested that this occurred because the guidance trajectory rate-of-change decreased throughout each trial. The faster-changing portion of the guidance trajectory required more frequent control inputs and therefore more frequent visual fixations. The guidance trajectory rate-of-change also appeared to influence the allocation of attention after mode transitions. When there was an LPR at the mode transition, subjects' attention on the attitude indicator increased before linearly decreasing over the remainder of the trial. This same linear decrease was not seen without an LPR. It is believed that these behaviors were the result of the increased guidance trajectory rate-of-change caused by the LPR.

This thesis developed a closed-loop vehicle-human model to further investigate these hypotheses about the factors that influenced subjects' attention allocation across control mode transitions. It was also used to investigate how attention allocations correlate to mental workload, situation awareness, and flying performance. The model was structured around the "perceive-think-decide-do" scheme of human information processing. Like the state estimator in the Optimal Control Model, it simulated how operators constantly estimate system states with a mental model of system dynamics and state rates-of-change, correct these estimates with information perceived from the displays, and use these estimates to make control inputs (Figure 31). These control inputs were based on the Crossover Model of manual control. In the model, operator attention was influenced by three "primitives": 1) the value of each instrument to the current task structure, which is a function of task priority; 2) the effort required to move eyes to each instrument, which is a function of the display layout; and 3) the level of uncertainty in the current estimate of each instrument's state, which is a function of the desired and actual sampling periods of the instrument. The first two primitives were similar to those in the SEEV models of attention, but the third primitive was new to this thesis. The uncertainty primitive was developed to directly relate attention and situation awareness, which is inversely proportional to the uncertainty of a system state estimate.

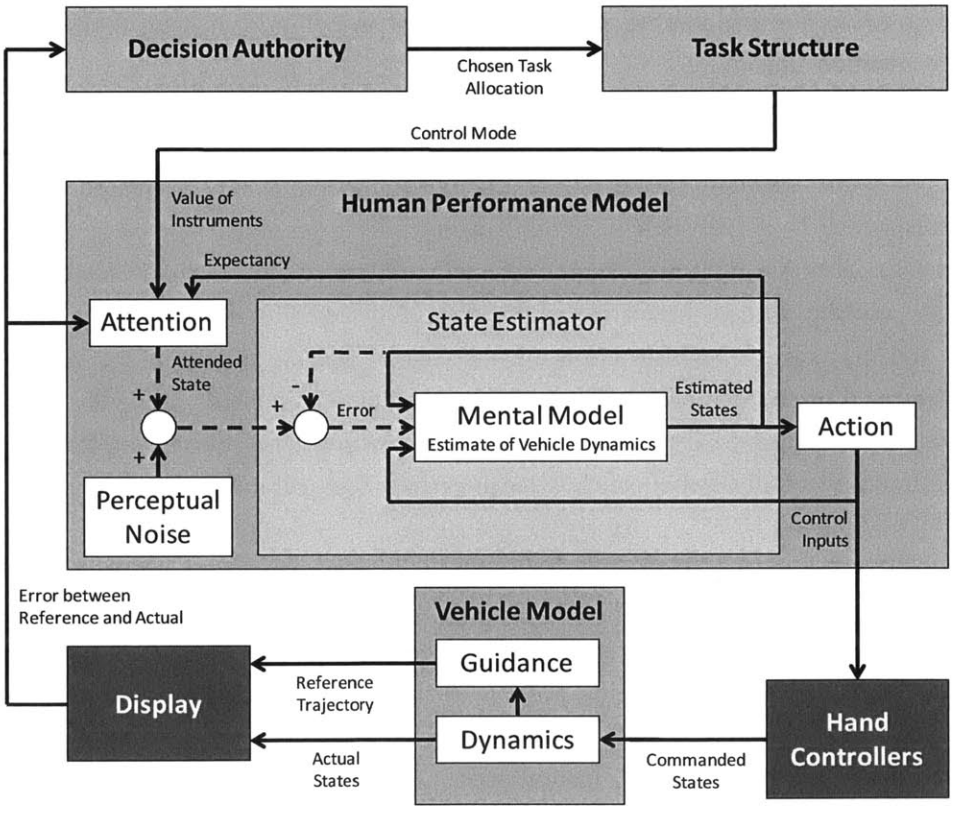


Figure 30. Structure of the integrated vehicle-human model. Figure repeated from Chapter 5.

These three primitives, and other parameters describing attention, were specified from existing literature, engineering first principles, knowledge of the experimental scenario design, or based on a data

set containing 60% of the experimental data. The model was used to simulate multiple runs of all mode transitions, and the attention budget was compared to a separate data set containing a different 33% of the experimental data. The weightings of the value, uncertainty, and effort were varied to investigate the sensitivity of the model outputs to these parameters. The results show a good match between the simulated and experimental data, and suggest that the value and uncertainty primitives influence attention more than effort, as prior research has found (Wickens et al., 2005; Steelman et al., 2011).

The model was also used to investigate the relationship between mental workload, situation awareness, and task performance. Other prior human performance models, like CASCaS and N-SEEV, do not have the ability to simulate all of these together with visual attention. Mental workload was measured by the spare attention not allocated to the flight instruments and by the response to simulated comm light illuminations. Situation awareness was estimated by the average peak uncertainty (APU) of an instrument when it was fixated and by simulated call-outs. Unlike the MIDAS or A-SA models of situation awareness, the prediction of situation awareness by simulated call-outs could be compared to similar experimental data.

The correlation between performance on the comm light response task and call-outs was compared to the spare attention for both the simulated and experimental data. While there were specific differences, the overall trends were the same for both data sets. As spare attention decreased, task performance also decreased. The model-predicted APU was then compared to the amount of spare attention and task performance. As expected, an increase in APU was correlated with decreases in spare attention and task performance. Finally, the model was used to predict performance on the primary flying task, and the results were compared to the experimental data. The model was found to underpredict the MSE between the actual and guidance-prescribed attitude, which may be due to the sensitivity of this metric to large errors at the beginning of a trial and at the time of an LPR.

6.2 Recommendations for the Adaptive Automation Field

The conclusions to each chapter discuss future work specific to that chapter's contents. This section provides additional suggestions for future work for the entire adaptive automation field. These recommendations discuss areas that, in the opinion of the author, should be investigated to promote the future implementation of adaptive automation in operational systems.

6.2.1 *Sensitivity and Specificity of Workload Measurement Techniques*

As described in Section 2.1.2, adaptive automation requires a system for triggering a task re-allocation. This triggering process can be viewed as a binary classification test, and it must have a high probability of correctly detecting a change in the flight events or an increase in the pilots' workload (i.e. it must have high sensitivity), while minimizing the number of incorrect re-allocations (i.e. it must have high specificity). Sensitivity and specificity thus both influence how well the state of the system and the operator can be determined. If the automation incorrectly classifies the system or operator state due to

noise in the event or workload detectors, it may re-allocate tasks at an unexpected or inappropriate time. This mismatch between pilot expectations and automation actions may lead, at best, to distrust in the automation, increased reluctance to use it, and/or extra time and effort spent making sure it is functioning properly (Parasuraman & Riley, 1997). At worst, it can cause a fatal accident. These issues will occur both if the automation determines that an operator's mental workload is above a given threshold when it is not (a "false positive" or type I error) or if the automation fails to recognize situations in which the operator workload is actually above the threshold (a "false negative" or type II error). It is critical that the automation's sensors and data-fusion algorithms minimize the number of type I and II errors to a level where pilots can be confident in the automation's predictability.

Adaptive automation clearly requires reliable detection of unexpected events, poor pilot performance, high mental workload, and/or the loss of situation awareness. But, how reliably can these states be detected with a given adaptive automation triggering system? One can frame the question using signal detection theory (Green & Swets, 1966) and the concept of a receiver operating characteristic (ROC) curve.

The ROC curve helps to clarify the tradeoff between type I and type II errors. It plots the true positive rate on the y-axis and the false positive rate on the x-axis. A simplified diagram of an ROC curve can be seen in Figure 31.

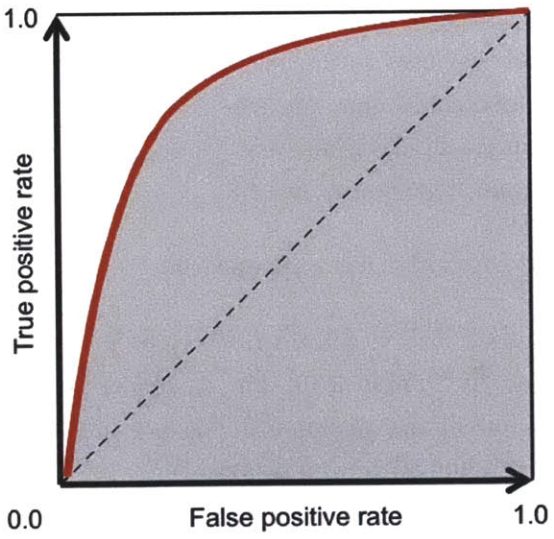


Figure 31. A representational receiver operator characteristic (ROC) curve is plotted in red. The grey shaded area represents the area under the curve (AUC).

The shape of an ROC curve is set by the separation between the signal and the noise in the underlying processes and the design of the instrument. The experimenter or system designer can set the detection threshold to operate the instrument at any point along the ROC curve, trading off sensitivity and

specificity. The chosen operating point should depend on the relative costs of false alarms and benefits of correct decisions in the operational context.

Because of this, the area under the curve (AUC) is an accepted summary metric that gives one number of the detection system's discriminatory power (i.e. the detection system's reliability). It does not depend on the chosen operating point. An AUC of 1 is perfect performance, and an AUC of 0.5 represents no discriminatory power (the detection system is randomly guessing). The medical literature suggests that AUC values of 0.97 or higher have "very high clinical value," due to the potentially high cost of an errant diagnosis (Fan et al., 2006). However, this domain is different from aviation and spaceflight, and the required AUC may be different as well.

These questions are closely related to research that has shown how decreasing automation reliability decreases operator trust, which in turn influences them to use less automation (Lee & Moray, 1992; Lee & Moray, 1994; Dzindolet et al., 2003; de Vries et al., 2003; Lee & See, 2004). Future research can build off of these studies, investigating the proper level of reliability (i.e. the minimum AUC) that gives the highest operator-automation system performance. This minimum AUC is likely to be different with different levels of operator experience, domains, environments, and types of automation. Future research should investigate these questions, testing how automation reliability affects the trust of novice and expert commercial aviation pilots or how the requirements for reliability are different depending on the stage of information processing completed by the automation.

Most researchers have contented themselves with characterizing the effect of automation using traditional tests of statistical significance. They manipulate the level of automation or the task difficulty and measure the effect on operator performance, workload, and situation awareness through a variety of techniques (see Section 2.1.2 and Section 4.1 for details on these measurement techniques). The measurement technique is considered sensitive to the changes in the experimental conditions if the data shows a statistically-significant difference between the performance, workload, and situation awareness measures across conditions. This is an incomplete evaluation of a measurement technique's capabilities because it does not fully characterize the specificity. Future research should characterize the ROC curve and the AUC to fully evaluate sensors and data-fusion algorithms to measure changes in operator performance, workload, or situation awareness.

Recommendation 1: Future research into adaptive automation triggers should investigate how operator trust of the trigger is affected by its reliability (quantified by the AUC) and the chosen operating point on the ROC curve under different levels of operator experience, domains, environments, and types of automation.

Recommendation 2: Future research into adaptive automation triggers should characterize the ROC curve and AUC of the measurements(s) and compare these to the recommendations made by the research addressing Recommendation 1.

6.2.2 *Ecological Validity of Simplified Experimental Conditions*

There is no question that research conducted with simplified experimental conditions is a necessary precursor to more detailed applied research and implementation. The noise and heterogeneity present in actual operational systems and scenarios are greatly reduced, which makes it easier to manipulate a variable and measure the effect. However, it is not clear whether the results from laboratory studies of adaptive automation have ecological validity, meaning that the measured effects would also be observed in real-world operational systems.

One way to determine the ecological validity of research conducted with simplified experimental conditions is to construct full systems and human-in-the-loop evaluations that employ adaptive automation in a complex environment that closely represents an operational system. A few projects have constructed such systems to serve as technology demonstrations and experimental platforms. However, as far as the author is aware, none of these demonstration systems have yet transitioned to real-world operations.

One example is the Pilot's Associate (PA) and Rotorcraft Pilot's Associate (RPA) described in Section 3.3. A more recent system employed adaptive automation for naval command and control tasks (Arciszewski et al., 2009; de Greef & Arciszewski, 2009; de Greef et al., 2010). Before operations, the user set "working agreements" that dictated when and how the automation should increase its authority. During operations, the automation used operator performance measurements and a model of operator cognition to estimate operator workload and trigger the adaptive automation. Lastly, the Playbook concept is a method of adaptable automation that makes it easy for the operator to delegate tasks to automated unmanned air vehicles (UAVs) (Miller, 2005; Miller & Parasuraman, 2007; Shaw et al., 2010; Miller et al., 2011). A "playbook" consists of a number of levels of automation that are available to the operator – manual control, selecting a pre-defined automated "script" for one UAV, or selecting a pre-defined automated "play" for multiple UAVs. While Playbook is a full, implementable system employing dynamic task allocation, it has only been tested in laboratory simulation environments.

Future research should take the results of research conducted with simplified experimental conditions and see how they hold in more complex, realistic environments, either physical or virtual. This will highlight areas of adaptive automation that are well understood, and provide guidelines for its implementation in future operational systems. Additionally, this research will identify complexities of real-world systems that change the way operators interact with adaptive automation. These complexities can then be investigated further in simplified experiments, closing the loop.

One important caveat to this recommendation is that the validation and verification of environments increases as the complexity increases. This occurs because a larger number of trials is required to encounter and statistically characterize all possible behaviors. As was stated in Section 3.4, there were difficulties in fully evaluating the PA and RPA associate systems because of the high-fidelity

environment (Smith and Geddes, 2003). If the simulation environments cannot be validated, then they cannot be used for the detailed design, verification, or certification of operational adaptive automation systems. However, they can still be used for investigative research. It is also important for future researchers to build upon well-validated research conducted with simplified experimental conditions and to compare the results of models and virtual environments to similar experimental data, when possible.

Recommendation 3: Future research should investigate implementations of dynamic task allocation in more complex and realistic scenarios, using the results of prior research conducted with simplified experimental conditions as a starting point.

6.2.3 *Using Pilot Flying-Pilot Monitoring Interaction as a Model for Adaptive Automation*

The interaction between the pilot flying (PF) and pilot monitoring (PM) in commercial aviation can serve as a model for how adaptive automation can be best implemented in operational systems. During flight, the PF is responsible for the aircraft's horizontal and vertical flight path and energy management regardless of the automation mode. The PF must be engaged in either manual control, directly hand-flying the aircraft, or supervisory control, monitoring the automation when it is flying the aircraft. The PM (also called the pilot not flying) has the responsibility for cross-checking the PF, performing tasks requested by the PF, and monitoring the state of the aircraft. The aircraft state to be monitored includes the flight director and flight management system (e.g. target airspeed and heading), aircraft systems (e.g. engine performance), and aircraft configuration (e.g. flap and slat settings). The PF and PM roles are distinct from the captain / first officer designation, and either member of the crew can perform the roles of the PF or PM. However, the captain remains the final authority in the cockpit.

The task structure between the PF and PM can change during flight, such as in response to a change in the phase of flight. When this occurs, the PF does not have to tell the PM what to do. Instead, both the PF and PM assume their new tasks as codified by the airline's standard operating procedures (SOP). The primary purpose of these procedures is to "identify and describe the standard tasks and duties of the flight crew for each flight phase" (Flight Safety Foundation, 1998c, p. 1). At their core, SOPs prescribe mutually agreed-upon strategies for a given phase of flight or off-nominal scenario that help the pilots to develop a shared mental model. At any point in the flight, the PF and PM know what tasks are their own responsibility and what tasks are the other pilot's responsibility. When each pilot follows the SOPs during dynamic task re-allocation, they both have the authority to shed old tasks and begin new tasks. The pilots are required to communicate verbally and by physically pointing to altered control settings, procedures that reinforce the shared mental model between them.

Pilot flying-pilot monitoring interaction can serve as a model for pilot-automation interactions and highlight areas where pilot-automation interaction is currently deficient. One such area is the shared mental model described in the previous paragraph. To complete the flight successfully, both pilots must understand the dynamics and control of the aircraft, the task structure and how to complete each task,

their own role in the task structure, and "the knowledge, skills, abilities, preferences, and other task-relevant attributes of their teammates" (Cannon-Bowers et al., 1993, p. 232). It is important to note that the shared mental model involves understanding of the other by both agents. In order for the pilots and the automation to share a mental model, each needs to understand how the other behaves.

Developing a shared mental model between the pilots and the automation is difficult because of communication barriers between the two. Communication is necessary for the pilot and automation to "access" each other's models, test their own knowledge of the system and environmental states, and update the other's knowledge (Sheridan & Verplank, 1978). Pilot flying-pilot monitoring communication can occur naturally, whereas pilot-automation communication must be mediated through an interface. Two pilots can engage in nonverbal "consequential communication," where information emerges as a consequence of the pilots' actions instead of being communicated intentionally (Segal, 1994). In other words, the pilots can see what each other is doing at all times and understand what tasks are being performed. This nonverbal communication is essential for developing the shared mental model between pilots. It allows each pilot to gain a concrete understanding of the state of the other pilot and the tasks that he is completing. Because the automation and the pilot cannot communicate in this manner, it can be harder for each to understand what the other is doing. They must make a conscious decision to communicate their actions and intentions to each other through text and words.

These limitations in pilot-automation communication and shared mental models are barriers to the implementation of adaptive automation in operational systems (Scerbo, 1996). If communication and the shared mental model are improved, the automation's behaviors will be more predictable and it is more likely that pilots will accept the automation having some authority to re-allocate tasks. Future research should use the interaction between two pilots as a model for developing automation capabilities and interfaces that enable the pilot and the automation to easily share mental models. Considering the automation as a team member has been proposed conceptually in the literature for some time (Sheridan & Verplank, 1978; Christoffersen & Woods, 2002; Klein et al., 2004; Goldman & Degani, 2012), and new cockpit interfaces have been proposed that attempt to clearly communicate information to the pilot based on "etiquette rules" (Miller & Funk, 2001) or crew resource management protocols (Geiselman et al., 2013). However, these principles have not yet fully implemented in operational automation and interface design.

Recommendation 4: Future research should use the interaction between two human pilots as a model for developing automation capabilities and interfaces that enable the pilot and the automation to easily share mental models.

6.3 Research Contributions

In summary, this thesis has made the following contributions to the field of dynamic task allocation research, including:

Chapters 1 and 2

1. Developed a 3-part taxonomy for classifying systems and procedures based on whether or not they: 1) feature the *dynamic re-allocation of tasks* between the operator and automation, 2) are *adaptive*, responding to the state of the operator, system, and environment, and 3) are *workload balancing*, meaning that their intent is to keep the operator in control as much as possible while remaining at a moderate level of mental workload.
2. Clarified the definition of *adaptive automation* as systems that feature adaptive, workload-balancing dynamic task allocation initiated by the automation (Chapter 2).

Chapter 3

3. Found that adaptive automation, per this definition, has not been implemented in commercial aviation approach and landing, adaptive systems in recent commercial aircraft, and operational military aircraft and rotorcraft. In commercial aviation approach and landing, the dynamic allocation of tasks is pilot-initiated *adaptable automation* (Section 3.1). The adaptive systems in recent commercial aircraft cannot be considered adaptive automation because they do not initiate a re-allocation of tasks, and/or they are not workload balancing (Section 3.2). While adaptive automation associate systems have been developed for military aircraft and rotorcraft, they have not been operationally implemented to the best of the author's knowledge (Section 3.3).
4. Hypothesized three reasons for this lack of adaptive automation in operational aviation systems: 1) difficulties in fully testing and evaluating adaptive automation, 2) manufacturers' desire to make the pilot the ultimate authority, and 3) the reluctance of operators to trust adaptive automation (Section 3.4).

Chapter 4

5. Conducted the first experimental investigation that measured subjects' visual fixations, flying performance, mental workload, and situation awareness all together during control mode transitions in which specific control loops varied between passive monitoring and active multiloop control.
6. Quantified how subjects increased and decreased attention on instruments across control mode reversions from autopilot to two-axis manual control and the effect that this had on performance-based measures of workload and situation awareness (Section 4.4.1).

a. A 14% increase in attention on the attitude indicator was accompanied by 5% decreases in attention on both the fuel indicator and the comm light. There was no significant change in performance on the comm light response task, but there was a significant decrease in performance on the tertiary fuel call-out task: subjects responding to 89% of the call-outs pre-mode transition, 39% during the mode transition, and 65% after the mode transition.

7. Suggested that attention on the attitude indicator across control mode transitions where the flying task was not re-allocated was influenced by the rate-of-change of the guidance trajectory, which dictated the required frequency of control inputs (Section 4.4.2).
    a. Attention on the attitude indicator significantly decreased 13% across mode transitions from two-axis rate-of-descent control to two-axis control. This occurred even though attention removed from the rate-of-descent indicator could have been allocated to the attitude indicator, which was required for the primary task. This likely occurred because the guidance trajectory rate-of-change was greater before the mode transition, which required more control inputs and a greater amount of attention.
8. Also suggested that attention on the attitude indicator after a mode transition with a landing point redesignation was influenced by the rate-of-change of the guidance trajectory (Section 4.4.3).
    a. Subjects increased attention on the attitude indicator in response to the step-change in flight guidance error caused by a landing point redesignation and then decreased their attention on this instrument at a rate of -3.5%/6-s window after the mode transition. Suggested that this occurred because the landing point redesignation also caused an increase in the guidance trajectory rate-of-change that lasted for approximately 18 s.

Chapter 5

9. Developed a closed-loop human-vehicle model that simulates attention allocation across control mode transitions and predicts operator mental workload, situation awareness, and flight performance.
10. Extended the SEEV model of attention in two ways: 1) capturing the influence of the uncertainty in system state estimates that grows between visual fixations on the related instrument, and 2) predicting the sampling period instead of just the total percent of attention (Section 5.2.2.3). The uncertainty parameter is specified by a desired sampling period, and the model compares this to the achieved sampling period to give a measure of situation awareness based on cognitive first principles (Section 5.2.2.6).
11. Investigated how weighting the different primitives that determine attention allocation (the *value* of an instrument to the current task structure, the *uncertainty* of the instrument's estimate at the current time step, and the *effort* to move the eyes to the instrument)

influenced the difference between the experimental and model-predicted attention budgets (the percent of attention allocated to each instrument) (Section 5.3). Found that when the uncertainty was weighted twice as much as the value and effort that the average difference on an instrument was ≤3.6%.

12. Found that when the effort primitive was weighted equally to the value and uncertainty primitives, the model under-predicted the attention on the instruments located on the peripheries of the lunar landing display (Section 5.3).
    a. The mean square error between the experimental and model-predicted attention on the fuel indicator and comm light were 4.5 and 2 time higher, respectively, than the other primitive weighting sets. This suggested that subjects were not impeded from making high-effort eye movements to an instrument when its value and uncertainty were high.
13. Predicted operator mental workload and situation awareness by cognitive first principles and performance-based measures that could be compared to experimental data. Based on cognitive first principles, workload was measured by the spare attention not allocated to the flying task and situation awareness was assumed to be proportional to the uncertainty in the estimate of a system state (Section 5.2.2.6). Performance-based metrics used were responses to a comm light illumination to predict mental workload (Section 5.2.6) and verbal call-outs of vehicle state to predict situation awareness (Section 5.2.5.2).
14. Took steps to validate this model's predictions of visual attention, mental workload (as measured by spare attention and secondary task performance), situation awareness (as measured by tertiary task performance), and flying performance against a portion of the experimental data not used to specify model parameters (Section 5.4).

Chapter 6

15. Recommended directions for future adaptive automation research to address research gaps and to help enable the implementation of the concept in operational systems (Section 6.2).

This page intentionally left blank.

7.0 References

1. *A Review of FAA's Efforts to Reduce Costs anad Ensure Safety and Efficiency Through Realignment and Consolidation*: Hearings before the Subcommittee on Aviation, Transportation and Infrastructure Committee, House of Representatives, 112th Congress. (2012). (Testimony of Paul M. Rinaldi). Retrieved from http://www.natca.org/ULWSiteResources/natcaweb/Resources/file/Legislative%20Center/Congressional%20Testimony/PaulHousdeMay312012.pdf
2. Abbott, K. (2001). Human Factors Engineering and Flight Deck Design. In C. R. Spitzer (Ed.), *The Avionics Handbook* (9-1–9-15). Boca Raton, FL: CRC Press.
3. Airbus SAS, & Flight Safety Foundation. (2006). *Flight Operations Briefing Notes – Standard Operating Procedures: Optimum Use of Automation*. Retrieved from http://www.skybrary.aero/bookshelf/books/190.pdf.
4. Annett, J. (2004). Hierarchical Task Analysis. In D. Diape and N.A. Stanton (Eds.), *The Handbook of Task Analysis for Human-Computer Interaction* (67-82). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
5. Arciszewski, H. F. R., de Greef, T. E., & van Delft, J. H. (2009). Adaptive Automation in a Naval Combat Management System. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 39(6), 1188-1199.
6. 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.
7. Bainbridge, L. (1983). Ironies of Automation. *Automatica* 19(6), 775-779.
8. Banks, S. B. & Lizza, C. S. (1991). Pilot's Associate: A Cooperative, Knowledge-based System Application. *IEEE Expert*, 6(3), 18-29.
9. Baron, S. & Kleinman, D. L. (1969). The Human as an Optimal Controller and Information Processor. *IEEE Transactions on Man-Machine Systems*, 10(1), 9-17.
10. Baron, S., Kleinman, D. L., & Levison, W. H. (1970). An Optimal Control Model of Human Response. Part II: Prediction of Human Performance in a Complex Task. *Automatica*, 6(3), 371-383.
11. Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997) Visual Scanning and Pilot Expertise: the Role of Attentional Flexibility and Mental Model Development. *Aviation, Space, and Environmental Medicine*, 68(7), 569-579.
12. Bennett, F. B. (1972). *Apollo Experience Report – Mission Planning for Lunar Module Descent and* Ascent. (NASA Technical Note D-6846). Washington, DC: National Aeronautics and Space Administration.
13. Bilimoria, K. D. (2009). Effects of Control Power and Guidance Cues on Lunar Lander Handling Qualities. *AIAA Journal of Spacecraft and Rockets*, 46(6), 1261-1271.

14. Billings, C. E. (1996). *Aviation Automation: The Search for a Human-Centered Approach.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
15. Billings, C. E. (1996). *Human-Centered Aviation Automation: Principles and Guidelines.* (Technical memorandum 110381). Moffett Field, CA: National Aeronautics and Space Administration.
16. Blignaut, P. (2009). Fixation Identification: The Optimum Threshold for a Dispersion Algorithm. *Attention, Perception, and Psychophysics*, 71(4), 881-895.
17. The Boeing Company. (2003). *767 Flight Crew Training Manual* (3rd rev.). Seattle: The Boeing Company.
18. The Boeing Company. (2003). *767-300 Operations Manual* (6th rev.). (Document Number D632T001-300G). Seattle: The Boeing Company.
19. The Boeing Company. (2011). *777-200/-200LR/-300/-300ER/F Flight Crew Operations Manual* (49th rev.). (Document Number D632W001-TBC). Seattle: The Boeing Company.
20. Cain, B. (2007). A *Review of the Mental Workload Literature.* (Technical Report RTO-TR-HFM-121-Part II). Toronto: Defense Research and Development Canada.
21. Calhoun, G.L., Ward, V.B.R., & Ruff H.A. (2011). Performance-based Adaptive Automation for Supervisory Control. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (2059-2063). Thousand Oaks, CA: SAGE Publications.
22. Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared Mental Models in Expert Team Decision Making. In N. J. Castellan, Jr. (Ed.), *Individual and Group Decision Making* (221-246). Hillsdale, NJ: Lawrence Erlbaum.
23. Carpenter, R. H. S. (1988). *Movements of the Eyes* (2nd ed.). London: Pion Limited.
24. Casali, J. G., & Wierwille, W. W. (1984). On the Measurement of Pilot Perceptual Workload: a Comparison of Assessment Techniques addressing Sensitivity and Intrusion Issues. *Ergonomics*, 27(10), 1033-1050.
25. Christoffersen, K. & Woods, D. D. (2002). How to Make Automated Systems Team Players. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research*, Vol. 2 (1-12). Bingley, UK: Emerald Group Publishing Ltd.
26. 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* (342-346). Thousand Oaks, CA: SAGE Publications.
27. Clamann, M.P. & Kaber, D.B. (2003). Authority in Adaptive Automation Applied to Various Stages of Human-Machine System Information Processing. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (543-547). Thousand Oaks, CA: SAGE Publications.
28. Cummings, M. L. (2010). Technology Impedances to Augmented Cognition. *Ergonomics in Design*, 18(2), 25-27.
29. de Crespigny, R. (2012). *QF32.* Sydney: Pan Macmillan.

30. de Greef, T. E. & Arciszewski, H. F. R. (2009). Triggering Adaptive Automation in Naval Command and Control. In S. Cong (Ed.), *Frontiers in Adaptive Control* (165-188). Vienna: InTech.
31. de Greef, T. E., Arciszewski, H. F. R., & Neerincx, M. A. (2010). Adaptive Automation Based on an Object-Oriented Task Model: Implementation and Evaluation in a Realistic C2 Environment. *Journal of Cognitive Engineering and Decision Making*, 4(2), 152-182.
32. de Greef, T. E., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation Foundations of Augmented Cognition. *Proceedings of the 5th International Conference on Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (219-228). Berlin: Springer.
33. de Visser, E. & 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.
34. de Vries, P., Midden, C., & Bouwhuis, D. (2003). The Effects of Errors on System Trust, Self-confidence, and the Allocation of Control in Route Planning. *International Journal of Human-Computer Studies*, 58(6), 719-735.
35. 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.
36. Duda, K. R., Johnson, M. C., & Fill, T.J. (2009). Design and Analysis of Lunar Lander Manual Control Modes. IEEE Aerospace Conference, Big Sky, MT.
37. Duda, K. R., Johnson, M. C., Fill, T. J., Major, L. M., & Zimpfer, D. J. (2010). Design and Analysis of an Attitude Command Velocity Hold/Hover Hold plus Incremental Position Command Blended Control Mode for Piloted Lunar Landing. AIAA GNC Conference, Toronto, Ontario.
38. Durso, F. T., Truitt, T. R., Hackworth, C. A., Albright, C. A., Bleckley, M. K., & Manning, C. A. (1995). *Reduced Flight Progress Strips in en route ATC Mixed Environments*. (Technical report DOT/FAA/AM-98/26). Washington, DC: Office of Aviation Medicine.
39. 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.
40. Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference* (789-795). New York, NY: IEEE.
41. Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), 32-64.
42. Endsley, M. R., & Kaber, D. B. (1999). Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. *Ergonomics*, 42(3), 462-492.

43. Ephrath, A. R. & Curry, R. E. (1977). Detection by Pilots of System Failures during Instrument Landings. *IEEE Transactions on Systems, Man and Cybernetics*, 7(12): 841-848.
44. Fan, J., Upadhye, S., & Worster, A. (2006). Understanding Receiver Operating Characteristic (ROC) Curves. *Canadian Journal of Emergency Medicine*, 8(1), 19-20.
45. Federal Aviation Administration. (2003). *Standard Operating Procedures for Flight Deck Crewmembers*. (Advisory Circular 120-71A). Washington, DC: U.S. Department of Transportation.
46. Federal Aviation Administration. (2013). *Manual Flight Operations*. (Safety Alert for Operators 13002). Washington, DC: U.S. Department of Transportation.
47. Federal Aviation Administration, Flight Deck Automation Working Group. (2013). *Operational Use of Flight Path Management Systems*. K. Abbott, D. McKenney, P. and Railsback (Eds.), Washington, DC: U.S. Department of Transportation.
48. Fidopiastis, C. M., Drexler, J., Barber, D., Cosenzo, K., Barnes, M., Chen, J. Y. C., & Nicholson, D. (2009). Impact of Automation and Task Load on Unmanned System Operator's Eye Movement Patterns. In D. D. Schmorrow, I. V. Estabrooke, and M. Grootjen (Eds.), *Foundations of Augmented Cognition, Neuroergonomics and Operational Neuroscience*. Berlin: Springer.
49. Fitts, P.M. (Ed.). (1951). *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. (Technical report ATI 133 954). Washington, D.C.: National Research Council.
50. Fletcher, J. W., Lusardi, J., Mansur, M. H., Moralez, E., III, Arterburn, D. R., Cherepinsky, I., Driscoll, J., Morse, C. S., & Kalinowski, K. F. (2008). UH-60M Upgrade Fly-By-Wire Flight Control Risk Reduction Using the RASCAL JUH-60A in-Flight Simulator. *Proceedings of the American Helicopter Society 64th Annual Forum* (1363-1389). Alexandria, VA: American Helicopter Society.
51. Flight Safety Foundation. (1998a). Being Prepared to Go Around. In *Approach and Landing Accident Reduction Toolkit* (Briefing note 6.1.). Alexandria, VA: Flight Safety Foundation.
52. Flight Safety Foundation. (1998b). Manual Go-Around. In *Approach and Landing Accident Reduction Toolkit* (Briefing note 6.2.). Alexandria, VA: Flight Safety Foundation.
53. Flight Safety Foundation. (1998c). Operating Philosophy. In *Approach and Landing Accident Reduction Toolkit* (Briefing note 1.1.). Alexandria, VA: Flight Safety Foundation.
54. Geiselman, E. E., Johnson, C. M., Buck, D. R., & Patrick, T. (2013). Flight Deck Automation: A Call for Context-Aware Logic to Improve Safety. *Ergonomics in Designs*, 21(4), 13-28.
55. Gillen, M. W. (2010, July). Diminishing Skills? *AeroSafety World*, 30-34. Retrieved from http://flightsafety.org/aerosafety-world-magazine/july-2010/diminishing-skills
56. Goldman, C. V., & Degani, A. (2012). A Team-Oriented Framework for Human-Automation Interaction: Implication for the Design of an Advanced Cruise Control System. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (2354-2358). Thousand Oaks, CA: SAGE Publications.

57. Goodrich, M. A., Boer, E. R., Crandall, J. W., Ricks, R. W., & Quigley, M. L. (2004). Behavioral Entropy in Human-Robot Interaction. *Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop* (93-100). Gaithersburg, MD: National Institute of Standards and Technology.
58. Gore, B. F., Hooey, B. L., Wickens, C. D., & Scott-Nash, S. (2009). A Computational Implementation of a Human Attention Guiding Mechanism in MIDAS v5. The 13th International Conference on Human-Computer Interaction, San Diego, CA.
59. Green, D. M., and Swets J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.
60. Hainley, C. J., Duda, K. R., Oman C. M., & Natapoff, A. (2013). Pilot Performance, Workload, and Situation Awareness During Lunar Landing Mode Transitions. *Journal of Spacecraft and Rockets*, 50(4): 793-801.
61. Hammer, J. M. (2009). Intelligent Interfaces. In J. A. Wise, V. D. Hopkin, and D. J. Garland (Eds.), *Handbook of Aviation Human Factors* (2nd ed.) (24-1 – 24-16). Boca Raton, FL: CRC Press.
62. Hancock, P. A. (2007). On the Process of Automation Transition in Multitask Human-Machine Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 37(4), 586-598.
63. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52, 139-183.
64. Hauss, Y., & Eyferth, K. (2003). Securing Future ATM-concepts' Safety by Measuring Situation Awareness in ATC. *Aerospace Science and Technology*, 7(6), 417-427.
65. Hayashi, M. (2003). Hidden Markov Models to Identify Pilot Instrument Scanning and Attention Patterns. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetic* (2889-2896). Piscataway, NJ: IEEE Press.
66. Hebb. D. O. (1955). Drives and the C.N.S. (Conceptual Nervous Systems). P*sychological Review*, 62, 243-254.
67. Hess, R. A. (1981). Aircraft Control-Display Analysis and Design Using the Optimal Control Model of the Human Pilot. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(7), 465-480.
68. Hooey, B. L., Gore, B. F., Wickens, C. D., Scott-Nash, S., Socash, C., Salud, E., & Foyle, D. C. (2010). Modeling Pilot Situation Awareness. *Proceedings of the Human Modelling in Assisted Transportation Conference* (207-213). Milan: Springer.
69. Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling Drivers' Visual Attention Allocation while Interacting with In-vehicle Technologies. *Journal of Experimental Psychology: Applied*, 12(2), 67-78.
70. Huemer, V. A., Hayashi, M., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005). Characterizing Scan Patterns in a Spacecraft Cockpit Simulator: Expert vs. Novice

Performance. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (83-86). Thousand Oaks, CA: SAGE Publications.

71. Inagaki, T. (2003). Adaptive Automation: Sharing and Trading of Control. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (147-169). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
72. Itti, L. & Koch, C. (2000). A Saliency-based Search Mechanism for Overt and Covert Shifts of Visual Attention. *Vision Research*, 40(10), 1489-1506.
73. Jagacinski, R. J. & Flach, J. M. (2002). *Control Theory for Humans: Quantitative Approaches to Modeling Performance.* Boca Raton, FL: CRC Press.
74. Kaber, D. B., & Endsley, M.R. (1997). Out-of-the-loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety. *Process Safety Progress*, 16(3), 126-131.
75. Kaber, D. B., & Endsley M. R. (2003). 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.
76. Kaber, D. B., & Riley J. M. (1999). Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement. *International Journal of Cognitive Ergonomics*, 3(3), 169-187.
77. 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.
78. Kaderka, J. D. (2014). *Experiments and a Model of Pilot System Failure Detection during Simulated Lunar Landing* (Doctoral dissertation). Retrieved from Dspace@MIT database.
79. Kalman, R. E., & Bucy, R. S. (1961). New Results in Linear Filtering and Prediction theory. *ASME Journal of Basic Engineering*, 83(1), 95-108.
80. Karwowski, W., Haas, M., & Salvendy, G. (2006). *A Review and Reappraisal of Adaptive Human-Computer Interfaces in Complex Control Systems*. (Technical report AFRL-HE-WP-TR-2006-0123). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.
81. Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems*, 19(6), 91-95.
82. Klein, M., & van Lambalgen, R. (2011). Design of an Optimal Automation System: Finding a Balance between a Human's Task Engagement and Exhaustion. In K. G. Mehrotra, C. K. Mohan, J. C. Oh, P. K. Varshney, & M. Ali (Eds.), *Modern Approaches in Applied Intelligence* (98-108). Berlin: Springer.
83. Kleinman, D. L. (1976). Solving the Optimal Attention Allocation Problem in Manual Control. *IEEE Transactions on Automatic Control*, 21(6), 813-822.
84. Kleinman, D. L., Baron, S., & Levison, W. H. (1971). A Control Theoretic Approach to Manned-Vehicle Systems Analysis. *IEEE Transactions on Automatic Control*, 16(6), 824-145.

85. Lee, J. D., & Moray, N. (1992). Trust, Control Strategies and Allocation of Function in Human-machine Systems. *Ergonomics*, 35(10), 1243-1270.
86. Lee, J. D., & Moray, N. (1994). Trust, Self-confidence, and Operators' Adaptation to Automation. *International Journal of Human-Computer Studies*, 40(1), 153-184.
87. Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50-80.
88. Levison, W. H., Baron S., & Kleinman, D. L. (1969). A Model for Human Controller Remnant. *IEEE Transactions on Man-Machine Systems*, 10(4), 101-108.
89. Lowenthal, C. S. (2012). *Evaluation of Sleepiness in Space Robotics Task Performance and Discussing Sleep with High School Students in a Museum* (Master's dissertation). Retrieved from Dspace@MIT database.
90. Lowy, J. (2011, Aug. 31). Automation Linked to Loss of Pilot Skills, Leading to Crashes. *Denver Post*. Retrieved from http://www.denverpost.com/ci_18792981
91. Manor, B. R., & Gordon, E. (2003). Defining the Temporal Threshold for Ocular Fixation in Free-viewing Visuocognitive Tasks. *Journal of Neuroscience Methods*, 128(1), 85-93.
92. Marsden, P. & Kirby, M. (2004). Allocation of Functions. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods* (338-346). Boca Raton, FL: CRC Press.
93. McRuer, D. T., Graham, D., Krendel, E. S., & Reisener, W., Jr. (1965). *Human Pilot Dynamics in Compensatory Systems: Theory, Models and Experiments with Controlled Element and Forcing Function Variations.* (Technical report AFFDL-TR-65-15.) Wright-Patterson Air Force Base, OH: Air Force Flight Dynamics Laboratory.
94. McRuer, D. T. & Jex, H. R. (1967). A Review of Quasi-Linear Pilot Models. *IEEE Transactions on Human Factors in Electronics* 8(3), 231-249.
95. McRuer, D. T. & Krendel, E. S. (1974). *Mathematical Models of Human Pilot Behavior.* (Technical report AGARD-AG-188). Neuilly sur Seine, France: NATO Advisory Group for Aerospace Research and Development.
96. Merfeld, D. M., Young, L. R., Oman, C. M., & Shelhamer, M. J. (1993). A multidimensional model of the effect of gravity on the spatial orientation of the monkey. *Journal of Vestibular Research: Equilibrium & Orientation* 3(2), 141-161.
97. Miller, C. A. (2005). Levels of Automation in the Brave New World: Adaptive Autonomy, Virtual Presence and Swarms-Oh My! *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (901-905). Thousand Oaks, CA: SAGE Publications.
98. Miller, C. A., & Dorneich, M. C. (2006). From Associate Systems to Augmented Cognition: 25 Years of User Adaptation in High Criticality Systems. In D. D. Schmorrow, K. M. Stanney, and L. M. Reeves (Eds.), *Foundations of Augmented Cognition* (2nd ed.) (344-353). Arlington, VA: Strategic Analysis.
99. Miller, C.A., Guerlain, S., & Hannen, M.D. (1999). The Rotorcraft Pilot's Associate Cockpit Information Manager: Acceptable Behavior from a New Crew Member? *Proceedings of the*

American Helicopter Society 55th Annual Forum (1321-3332). Alexandria, VA: American Helicopter Society.

100. Miller, C. A., & Funk, H. B. (2001). Associates with Etiquette: Meta-Communication to Make Human-Automation Interaction more Natural, Productive and Polite. 8th European Conference on Cognitive Science Approaches to Process Control, Munich, Germany.
101. Miller, C. A., & Parasuraman R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors*, 49(1), 57-75.
102. Miller, C. A., Shaw, T., Emfield, A., Hamell, J., de Visser, E., Parasuraman, R., & Musliner, D. (2011). Delegating to Automation. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting* (95-99). Thousand Oaks, CA: SAGE Publications.
103. Moore, K., & Gugerty, L. (2010). Development of a Novel Measure of Situation Awareness: The Case for Eye Movement Analysis. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (1650-1654). Thousand Oaks, CA: SAGE Publications.
104. National Aeronautics and Space Administration. (2008). *Constellation design reference missions and operational concepts, Revision B, Change 003.* (Document no. CxP 70007). Washington, DC: National Aeronautics and Space Administration.
105. National Aeronautics and Space Administration. (2014). *Human Research Roadmap*. Retrieved from http://www.humanresearchroadmap.nasa.gov
106. National Transportation Safety Board. (2014). *Descent Below Visual Glidepath and Impact With Seawall: Asiana Airlines Flight 214.* (Accident report NTSB/AAR-14/01). Washington, DC: National Transportation Safety Board.
107. Newman, R. L. (1995). *Head-Up Displays: Designing the way ahead.* Aldershot, England: Ashgate Publishing Limited.
108. O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload Assessment Methodology. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance, Volume II: Cognitive Processes and Performance.* (42/1-42/9). New York, NY: Wiley Interscience.
109. Oman, C. M. (1982). A Heuristic Mathematical Model for the Dynamics of Sensory Conflict and Motion Sickness. *Acta Otolaryngol Supplement*, 94(S392), 4-44.
110. Oppermann, R. (1994). *Adaptive User Support.* Hillsdale, NJ: Erlbaum.
111. Parasuraman, R., Cosenzo, K.A., and de Visser, E. (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.
112. Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-induced "Complacency." *The International Journal of Aviation Psychology*, 3(1), 1-23.
113. Parasuraman, R., Mouloua M., & Hilburn, B. (1999). Adaptive Aiding and Adaptive Task Allocation Enhance Human-Machine Interaction. *Proceedings of the Third Conference on Automation Technology and Human Performance* (119-123). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

114. Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of Adaptive Task Allocation on Monitoring of Automated Systems. *Human Factors*, 38(4), 665-679.
115. Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1993). Adaptive Function Allocation Reduces Performance Cost of Static Automation. In J. G. Morrison (Ed.), *The Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program: Interim Research and Guidelines for the Application of Adaptive Automation.* (Technical report NADC-91028-60). (37-42). Warminster, PA: Naval Air Warfare Center – Aircraft Division.
116. Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230-253.
117. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 30(3), 286-297.
118. Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014a). Modeling Human-Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*, 8(1), 33-51.
119. Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014b). Measuring Human-Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*, 8(1), 52-77.
120. Proctor, R. W., & Van Zandt, T. (2008). *Human Factors in Simple and Complex Systems*. (2nd ed.) Boca Raton, FL: CRC Press.
121. Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2010). Single Operator, Multiple Robots: an Eye Movement based Theoretic Model of Operator Situation Awareness. *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (235-242). Piscataway, NJ: IEEE Press.
122. Rencken, W. D. & Durrant-Whyte, H. F. (1993). A Quantitative Model for Adaptive Task Allocation in Human-computer Interfaces. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 23(4), 1072-1090.
123. Robertson, G. (2000). Flight Demonstration of an Associate System: A Rotorcraft Pilot's Associate Example. *Proceedings of the American Helicopter Society 56th Annual Forum* (431-445). Alexandria, VA: American Helicopter Society.
124. Roscoe, A. H., & Ellis, G. A. (1990). *A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use.* (Technical Report TR 90019). Farnborough, UK: Royal Aerospace Establishment.
125. Rouse, W. B., Geddes, N. D., & Curry, R. E. (1987). An Architecture for Intelligent Interfaces: Outline of an Approach to Supporting Operators of Complex Systems. *Human-Computer Interaction* 3(2), 87-122.
126. Sarter, N. B. & Woods, D. D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors*, 37(1), 5-19.
127. Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation Surprises. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.). Hoboken, NJ: Wiley.

128. Sauer, J., Kao, C. S., Wastell, D., and Nickel, P. (2011). Explicit Control of Adaptive Automation under Different Levels of Environmental Stress. *Ergonomics*, 54(8), 755-766.
129. Scallen, S. F., & Hancock, P. A. (2001). Implementing Adaptive Function Allocation. *The International Journal of Aviation Psychology*, 11(2), 197-221.
130. Scerbo, M. W. (1996). Theoretical Perspectives on Adaptive Automation. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (37-63). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
131. Scerbo, M. W. (2007). Adaptive Automation. In R. Parasuraman and M. Rizzo (Eds.). *Neuroergonomics: The Brain at Work* (239-252). New York, NY: Oxford University Press.
132. Segal, L. D. (1994). Actions Speak Louder than Words: How Pilots use Nonverbal Information for Crew Communications. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (21-25). Santa Monica, CA: Human Factors and Ergonomics Society.
133. Sharit, J. (1997). Allocation of Functions. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (301-339). New York, NY: Wiley.
134. Shaw, T., Emfield, A., Garcia, A., de Visser, E., Miller, C., Parasuraman, R., & Fern, L. (2010). Evaluating the Benefits and Potential Costs of Automation Delegation for Supervisory Control of Multiple UAVs. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (1498-1502). Thousand Oaks, CA: SAGE Publications.
135. Sheridan, T. B. (1970). On how often the supervisor should sample. *IEEE Transactions on Systems, Science, and Cybernetics*, 6(2), 140-145.
136. Sheridan, T. B. (1992). *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press.
137. Sheridan, T. B. (2011). Adaptive Automation, Level of Automation, Allocation Authority, Supervisory Control, and Adaptive Control: Distinctions and Modes of Adaptation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 41(4), 662-667.
138. Sheridan, T. B., & Ferrell, W. R. (1974). *Man-Machine Systems*. Cambridge, MA: MIT Press.
139. Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. (Technical report). Cambridge, MA: MIT Man-Machine Systems Lab.
140. Sherry, R. R., & Ritter, F. E. (2002). *Dynamic Task Allocation: Issues for Implementing Adaptive Intelligent Automation*. (Report no. ACS 2002-2). University Park, PA: The Pennsylvania State University.
141. Smith, P. J., & Geddes, N. D. (2003). A Cognitive Systems Engineering Approach to the Design of Decision Support Systems. In A. Sears and J. Jacko (Eds.), *Handbook of Human-Computer Interaction* (656-675). Mahwah, NJ: Lawrence Erlbaum.
142. St. John, M., Kobus, K. L., Morrison, J. G., & Schmorrow, D. (2004). Overview of the DARPA Augmented Cognition Technical Integration Experiment. *International Journal of Human-Computer Interaction*, 17(2), 131-149.

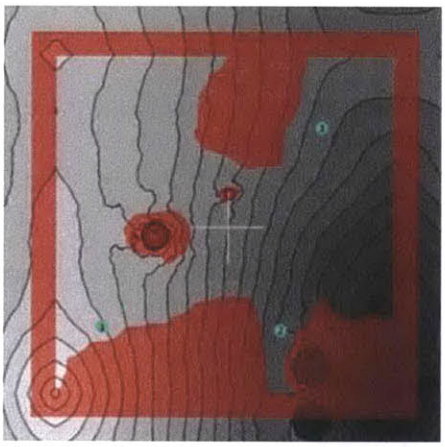
143. Stanley, W. D., Dougherty, G. R., & Dougherty, R. (1983). *Digital Signal Processing* (2nd ed.). Reston, VA: Reston Publishing Company, Inc.
144. Stanton, N. A. (2006). Hierarchical Task Analysis: Developments, Applications, and Extensions. *Applied Ergonomics*, 37(1), 55-79.
145. Steelman K. S., McCarley, J. S., & Wickens, C. D. (2011). Modeling the Control of Attention in Visual Workspaces. *Human Factors*, 53(2), 142-153.
146. Steelman-Allen, K. S. (2011). *Modeling the Control of Visual Attention in Complex Workspaces* (Doctoral dissertation). Retrieved from University of Illinois IDEALS database.
147. Stimpson, A. J. (2011). *Design and Evaluation of an Achievability Contour Display for Piloted Lunar Landing* (Master's dissertation). Retrieved from DSpace@MIT database.
148. Stock, S., Putnam, J., & Carroll, J. (2013, Aug. 20). Commercial Pilots: Addicted to Automation? *NBC Bay Area.* Retrieved from http://www.nbcbayarea.com/investigations/Commercial-Pilots-Addicted-to-Automation--221727971.html
149. Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design. In *Situational Awareness in Aerospace Operations* (3/1-3/17). (Technical report AGARD-CP-478). Neuilly Sur Seine, France: NATO-AGARD.
150. Tobii Technology. (2005). *How to Set up the Tobii x50.* Stockholm, Sweden: Tobii Technology.
151. Tobii Technology. (2006). *User Manual: Tobii Eye Tracker and ClearView Analysis Software.* Stockholm, Sweden: Tobii Technology.
152. Tyler, S. W., Neukom, C., Logan, M., & Shively, J. (1998). The MIDAS Human Performance Model. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (320-324). Thousand Oaks, CA: SAGE Publications.
153. Walther, D., & Koch, C. (2006). Modeling Attention to Salient Proto-objects. *Neural Networks*, 19, 1395-1407.
154. Warwick, G. (2009, April 6). Beyond Brownout. *Aviation Week & Space Technology*, 171, 44-45.
155. Welch, G., & Bishop, G. (2006). *An Introduction to the Kalman Filter.* (Technical Report TR 95-041.) Chapel Hill, NC: University of North Carolina at Chapel Hill.
156. Wickens, C. D. (1980). The Structure of Attentional Resources. In R. S. Nickerson (Ed.) *Attention and Performance VII.* (239-257). Hillsdale, NJ: Lawrence Erlbaum.
157. Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors*, 45(3), 360-380.
158. Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
159. Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009). Identifying Black Swans in NextGen: Predicting Human Performance in Off-Nominal Conditions. *Human Factors*, 51(5), 638-651.

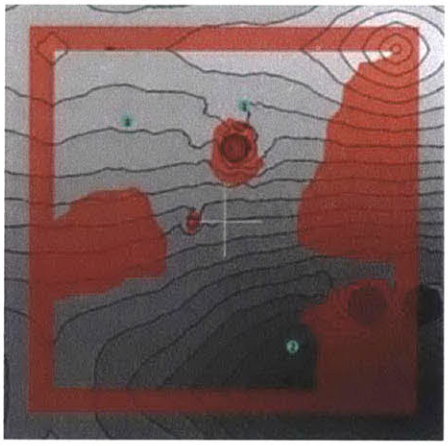
160. Wickens, C. D., McCarley, J. S., Alexander, A. L., Thoman, L. C., Ambinder, M., & Zheng, S. (2005). *Attention-Situation Awareness (A-SA) Model of Pilot Error*. (Technical Report AHFD-04-15/NASA-04-5). Moffett Field, CA: NASA Ames Research Center.
161. Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft*. (Technical report 117528). Moffett Field, CA: NASA Ames Research Center.
162. Wiener, E. L., & Curry, R. E. (1980). Flight-deck Automation: Promises and Problems. *Ergonomics*, 23(10), 995-1011.
163. Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots during Flight Using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology*, 12(1), 3-18.
164. Wood, S. (2004). *Flight Crew Reliance on Automation*. (CAA PAPER 2004/10). West Sussex, UK: Civil Aviation Authority.
165. Woods, D. D. & Sarter, N. B. (1998). *Learning from Automation Surprises and "Going Sour" Accidents: Progress on Human-Centered Automation*. (Report no. NASA-CR-1998 -207061). Washington, D.C.: National Aeronautics and Space Administration.
166. Wortelen, B., Baumann, M., & Lüdtke, A. (2013). Dynamic Simulation and Prediction of Drivers' Attention Distribution. *Transportation Research, Part F: Traffic Psychology and Behaviour*, 21, 278-294.
167. Wortelen, B., Lüdtke, A., & Baumann, M. (2013). Integrated Simulation of Attention Distribution and Driving Behavior. *Proceedings of the 22nd Annual Conference on Behavior Representation in Modeling and Simulation* (69-76).
168. Wortelen, B., Lüdtke, A., & Baumann, M. (2013). Simulating Attention Distribution of a Driver Model: How to Relate Expectancy and Task Value? *Proceedings of the 12th International Conference on Cognitive Modeling* (269-274).
169. Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459-482.
170. Young, L. R. (1973). Human Control Capabilities. In J. F. Parker, Jr. and V. R. West (Eds.), *Bioastronautics Data Book* (2nd ed.) (751-806). Washington, DC: National Aeronautics and Space Administration.
171. Yu, C., Wang, E. M., Li, W., & Braithwaite, G. (2014). Pilots' Visual Scan Patterns and Situation Awareness in Flight Operations. *Aviation, Space, and Environmental Medicine*, 85(7), 708-714.

Appendix A. Additional Details on Experimental Methods

Digital elevation maps

The following three digital elevation maps (DEMs) were used in the experimental data-collection trials. As is stated in Section 4.2.1, the shades of grey and contour lines on the DEM indicated lunar terrain of different altitudes. Subjects were told the red square symbolized a prior radar "scan area," on which hazards and LAPs had been identified. Subjects could overfly the red hazard areas but not land in them. All trials began targeting the center of the scan area as the LAP. In Figure 32, the white crosshairs representing the vehicle position is over the scan area center. In some trials, the LAP was automatically redesignated to one of the alternate LAPs (marked with cyan circles).

Map 0: 

Map 1: 

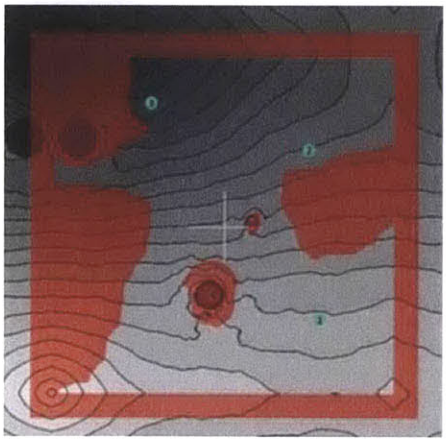
Map 3: 

Figure 32. Digital elevation maps used in the experimental data-collection trials, cropped to only show the square, red "scan area."

Subject demographics

Demographic information for the twelve experimental subjects is shown in Table 14. Only one subject had flying experience, and he held a private pilot license (PPL) with 260 flight hours. Five other subjects had experience with computer flight simulators, to varying degrees. While flight experience varied between subjects, all were trained to a similar level of performance before the data-collection trials.

Table 14. Demographics of experimental subjects.

Subject Number	Gender	Age	Handedness	Contacts Prescription	Pilot Experience	Computer-based Flight Simulator Experience
1	M	24	R	None	No	No
2	M	22	R	None	No	No
3	F	23	R	None	No	No
4	M	31	R	None	PPL	Yes
5	M	21	R	None	No	Yes
6	M	28	R	None	No	No
7	F	22	R	None	No	Yes
8	M	20	R	-5 in both	No	Yes
9	F	21	R	-1.5 in right, -1.75 in left	No	Yes
10	F	22	R	-4 in both	No	No
11	F	24	R	None	No	No
12	M	28	R	None	No	Yes

Test matrices

In the following test matrices, VC 1 represents the pre-mode transition control mode and VC 2 represents the post-MT control mode. In each trial, a landing point redesignation (LPR) was either present (LPR=1) or not (LPR=0).

Table 15. Test matrix for training protocol.

TRAINING							
Trial	**Map**	**Offset**	**VC 1**	**VC 2**	**LPR**	**Failure**	**Failure Time**
1	12	R	Auto	Auto	0	None	
2	13	L	Auto	TA	1	None	
3	14	R	Auto	TA+RoD	1	None	
4	15	L	TA	Auto	1	None	
5	12	R	TA	TA+RoD	1	None	
6	13	L	TA+RoD	Auto	1	None	
7	14	R	TA+RoD	TA	1	None	
8	15	L	TA	Auto	0	Fuel	61.41
9	12	R	TA+RoD	TA	1	Fuel	54.17
10	13	L	TA	Auto	0	Radar	57.54
11	14	R	Auto	TA+RoD	1	Radar	53.79
12	15	L	TA	Auto	0	Thrust	53.67
13	12	R	TA	TA+RoD	1	Thrust	54.02
14	13	L	TA+RoD	Auto	0	Fuel	58.90
15	14	R	TA+RoD	TA	1	Radar	56.76
16	15	L	TA	TA+RoD	1	Thrust	63.48
17	12	R	Auto	TA	0	None	
18	13	L	TA	Auto	0	Radar	63.70
19	14	R	TA	TA+RoD	1	Fuel	60.67
20	15	L	TA+RoD	Auto	0	None	
21	12	R	TA+RoD	TA	1	Thrust	57.72
22	13	L	Auto	TA+RoD	1	Radar	54.36
23	14	R	Auto	TA	0	Fuel	52.34
24	15	L	Auto	TA+RoD	1	None	
25	12	R	TA	Auto	0	Thrust	52.51

Table 16. Test matrix for data-collection trials.

EXPERIMENT							
Trial	**Map**	**Offset**	**VC1**	**VC2**	**LPR**	**Failure**	**Failure Time**
1	0	R	TA+RoD	Auto	0	Radar	61.02
2	0	R	Auto	TA	0	Thrust	50.30
3	1	L	Auto	TA+RoD	0	Fuel	53.19
4	1	R	TA	Auto	1	None	
5	3	L	TA	TA+RoD	1	Radar	59.14
6	1	L	TA+RoD	TA	0	Thrust	60.82
7	3	L	Auto	TA+RoD	1	None	
8	0	L	Auto	TA	1	Fuel	63.67
9	1	R	TA+RoD	Auto	1	Thrust	54.30
10	1	L	TA	TA+RoD	0	Radar	62.30
11	3	L	TA+RoD	TA	1	Fuel	54.17
12	3	L	TA	Auto	1	Thrust	60.05
13	1	L	Auto	TA	0	None	
14	3	R	TA+RoD	Auto	0	Fuel	58.56
15	1	R	Auto	TA+RoD	1	Thrust	56.04
16	1	L	TA	Auto	0	Radar	57.45
17	0	L	TA	TA+RoD	1	None	
18	1	R	TA+RoD	TA	1	Radar	54.60
BREAK							
19	1	L	TA+RoD	Auto	0	Fuel	61.38
20	0	L	Auto	TA+RoD	1	Thrust	60.98
21	3	R	TA	Auto	0	Radar	51.78
22	3	R	Auto	TA	0	None	
23	0	R	TA	TA+RoD	0	Fuel	58.72
24	3	L	TA+RoD	Auto	1	None	
25	3	R	TA+RoD	TA	0	Thrust	58.54
26	0	R	Auto	TA+RoD	0	Radar	53.79
27	1	R	Auto	TA	1	Fuel	58.53
28	0	L	TA	Auto	1	None	
29	0	L	TA+RoD	TA	1	Radar	59.98
30	3	R	TA	TA+RoD	0	Thrust	54.02
31	0	R	TA+RoD	TA	0	None	
32	0	R	TA	Auto	0	Fuel	52.12
33	1	R	TA	TA+RoD	1	None	
34	0	L	TA+RoD	Auto	1	Thrust	59.71
35	3	R	Auto	TA+RoD	0	Fuel	60.86
36	3	L	Auto	TA	1	Radar	60.56

Setup and Calibration of the Tobii x50 Eye Tracker

The Tobii x50 eye tracker was mounted below the simulator displays, as seen in Figure 33:

Figure 33. Location of the Tobii x50 eye tracker in the Draper Laboratory fixed-base lunar landing simulator.

The specific locations and orientations of the eye tracker and subject are detailed in Table 17 and Figure 34.

Table 17. Details of the eye tracker and subject positioning in the lunar landing simulator.

Parameter	Value	
Distance from eye tracker to screen (D, Figure 34)	3.5 in	(9 cm)
Angle between eye tracker and screen (θ, Figure 34)	3.5°	
Distance from eye tracker to the floor	36.75 in	(93 cm)
Distance from subjects' eyes to the floor	45.9±0.4 in	(117±1 cm)
Distance from eye tracker to subjects' eyes	22.4±0.7 in	(57±2 cm)

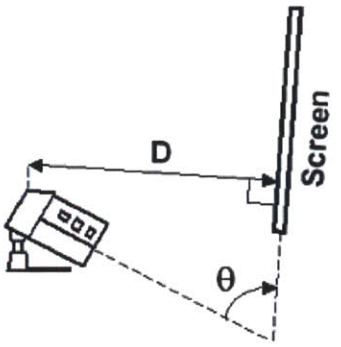


Figure 34. Representation of the parameters that specify the eye tracker's positioning relative to the simulator displays (Tobii Technology, 2005).

This placement of the eye tracker gave an accurate reading of subjects' point-of-regard on the simulator displays, and did not interfere with any of the experimental tasks. The operational range of the eye tracker was only 19.1 in (49 cm) wide, which is 3/4 the width of the two simulator displays combined (25.75 in (65 cm)). As a result, the instruments were all moved towards the center of the combined display to reside within the eye tracker's operational range.

As is stated in Section 4.2.3, there were two calibration procedures performed. The first was the x50 eye tracker's native 9-point calibration routine. The second was developed by the experimenters to compensate for substantial, yet consistent, error left from this native calibration. In this secondary calibration, 9 paper circles were placed on the simulator displays, as shown in Figure 35:

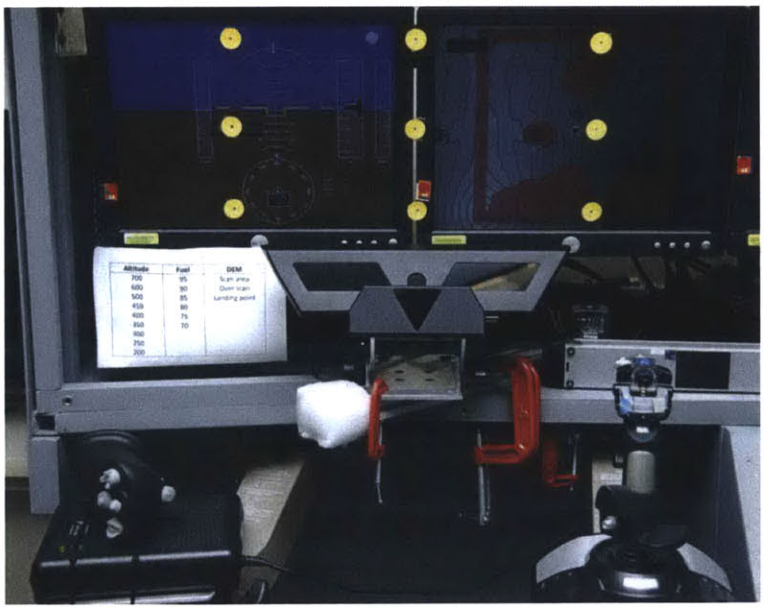


Figure 35. The orange circles on the simulator displays were used in the secondary calibration. These calibration points were taped to the display, and could be removed and re-applied.

Subjects were asked to look at each of these 9 calibration points 5 times, and the average x and y error about each point was determined. These nine "corrections" were applied *post-hoc* to each subsequent fixation, weighted by the inverse square distance between the fixation and the particular calibration point.

After both calibration routines were performed, subjects performed a validation step subjects were instructed to fixate on a series of 11 different instruments in order (Figure 36). Subjects' points-of-regard on the simulator displays were recorded, and both calibrations were applied to the data. Then, the data was inspected to make sure the fixations fell within the AOI they were actually looking at. This validated the attention data collection procedures, and cleared the subject to participate in the remainder of the experiment.

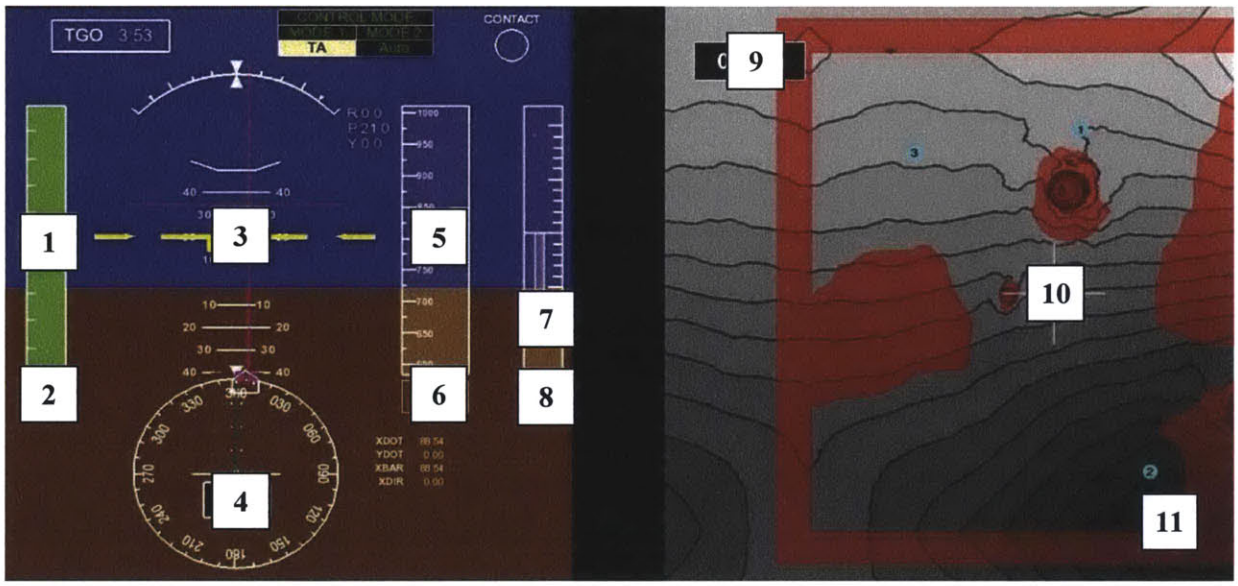


Figure 36. In the eye tracker validation, subjects looked at these 11 instruments in the specified order.

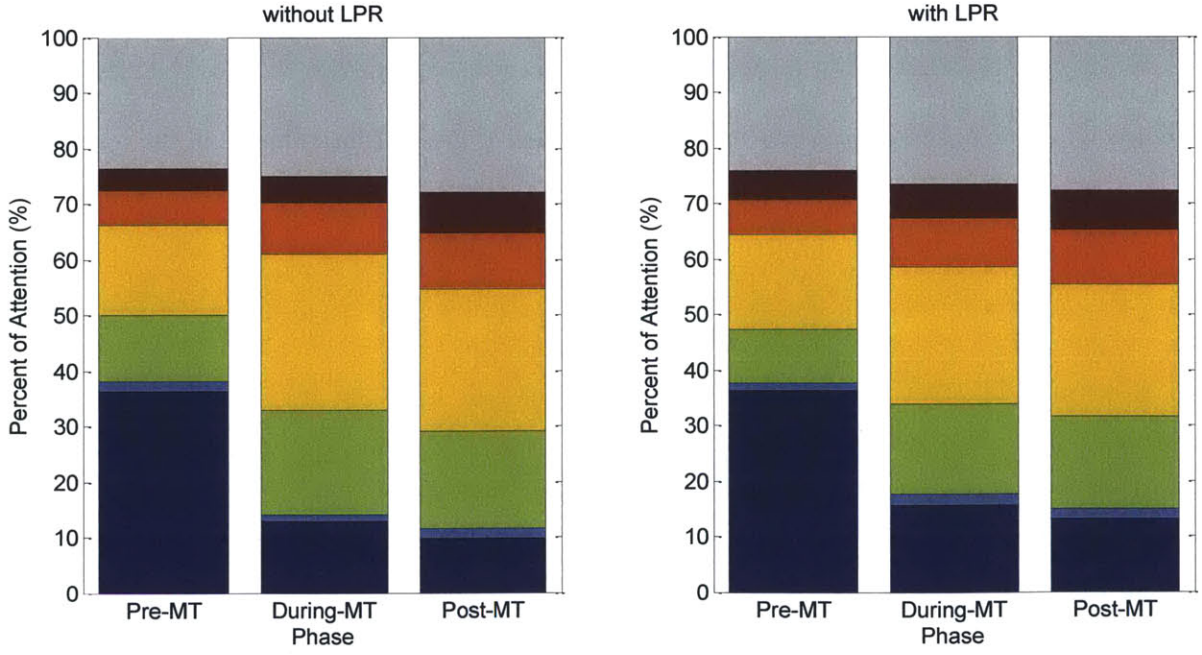
Appendix B. Additional Experimental Results

Attention Budgets for All Mode Transitions

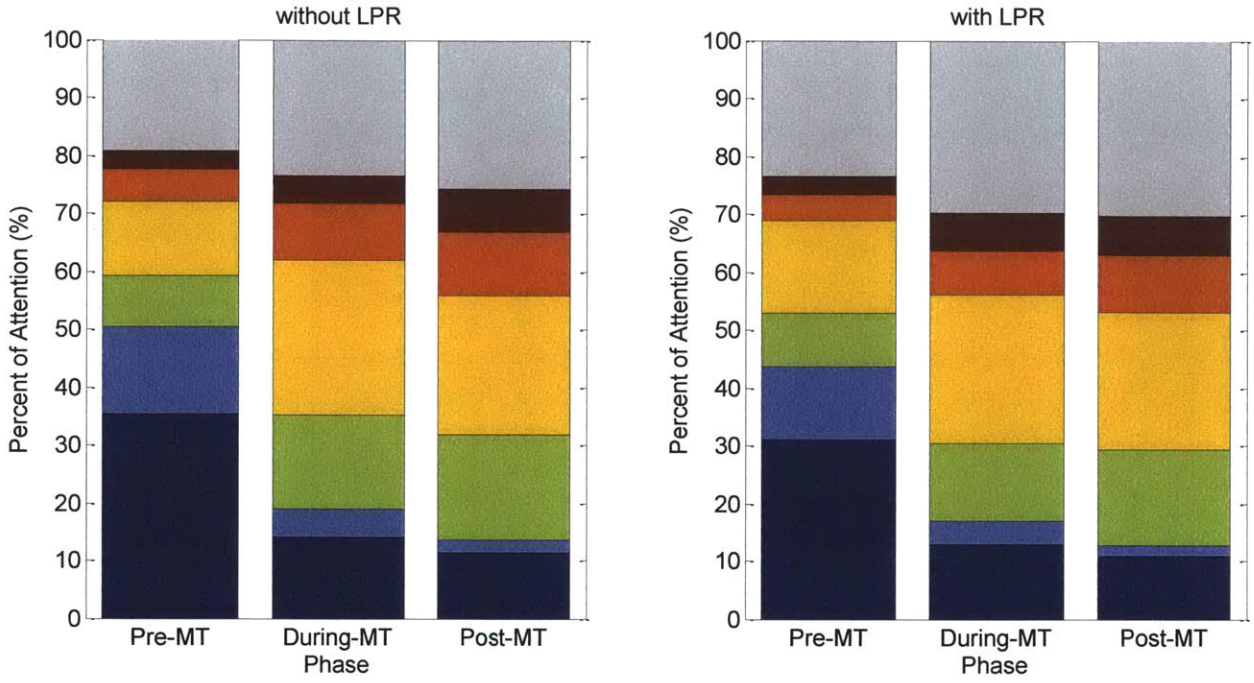
The following section shows the attention budget (similar to Figure 6) for all six mode transitions, with and without landing point redesignations (LPR). The legend for all figures is the same, and is shown in Figure 37. The order of the instruments listed is the same order as the instruments in the stacked bars, from top to bottom.

Figure 37. Legend for attention budget figures.

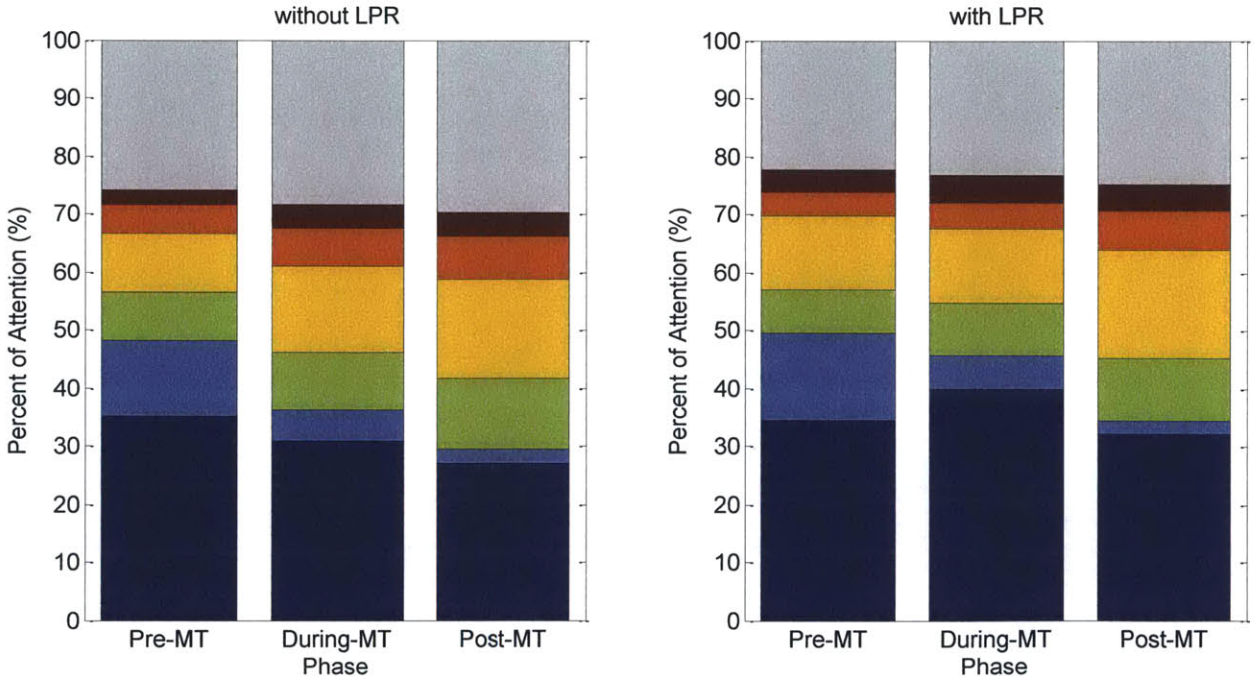
TA→Auto



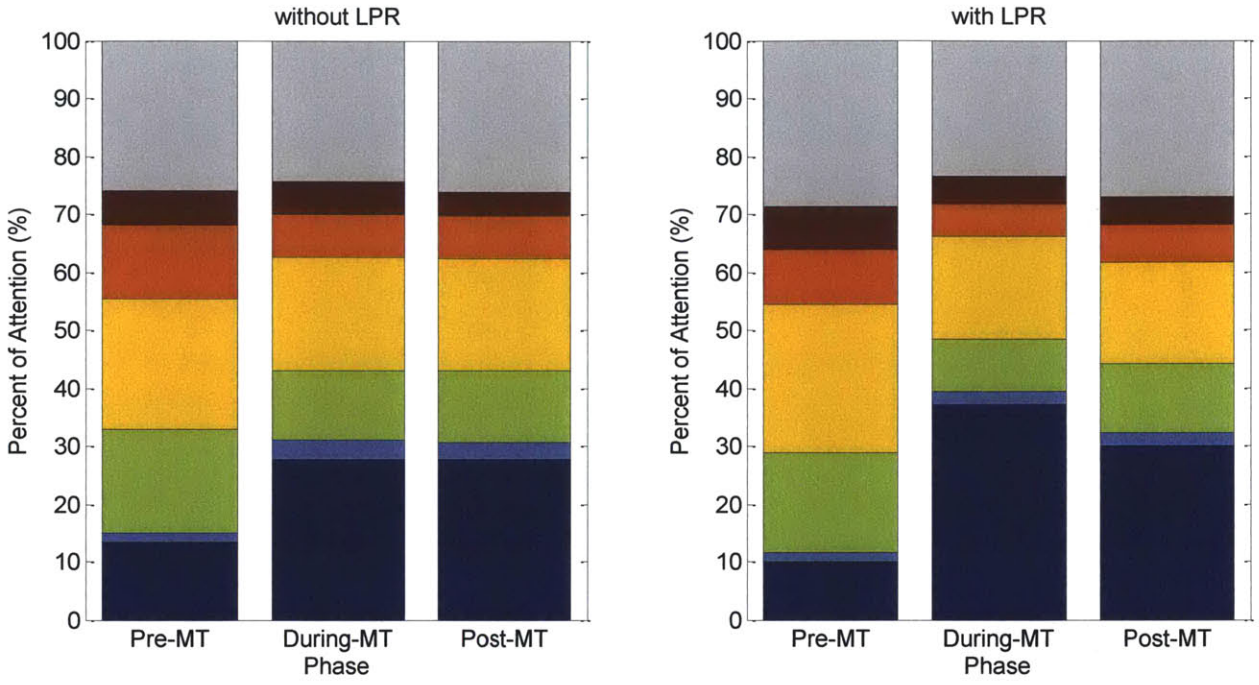
TA+RoD→Auto



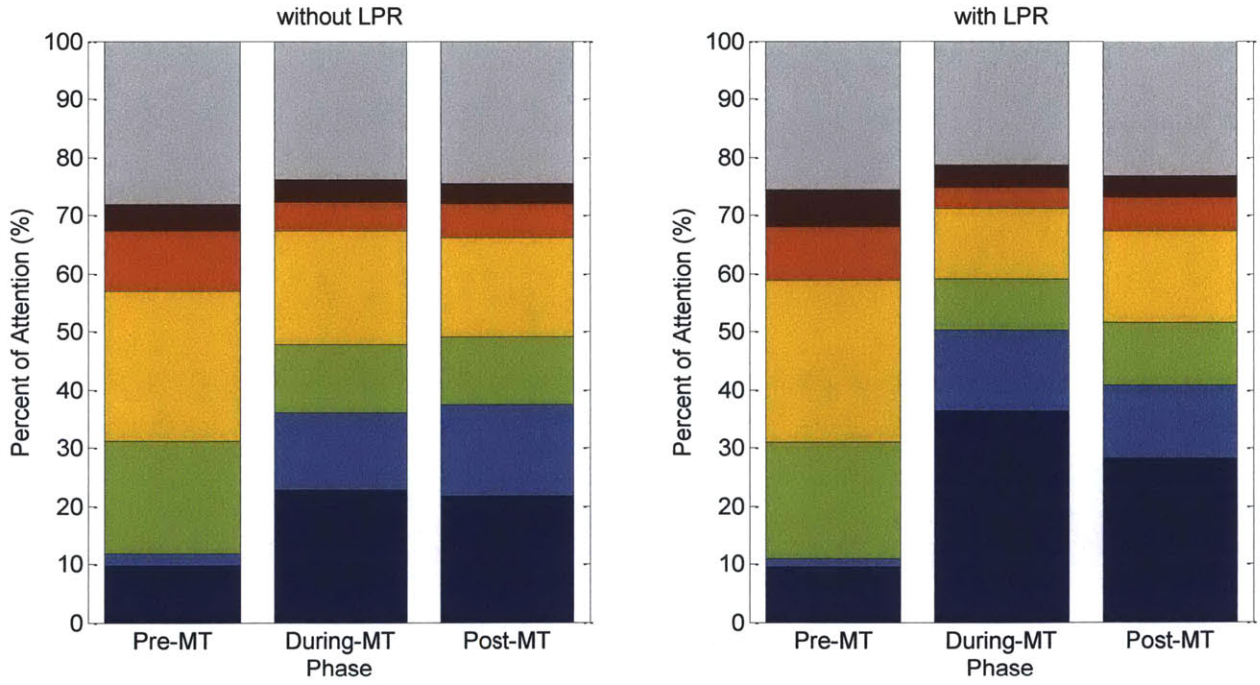
TA+RoD→TA



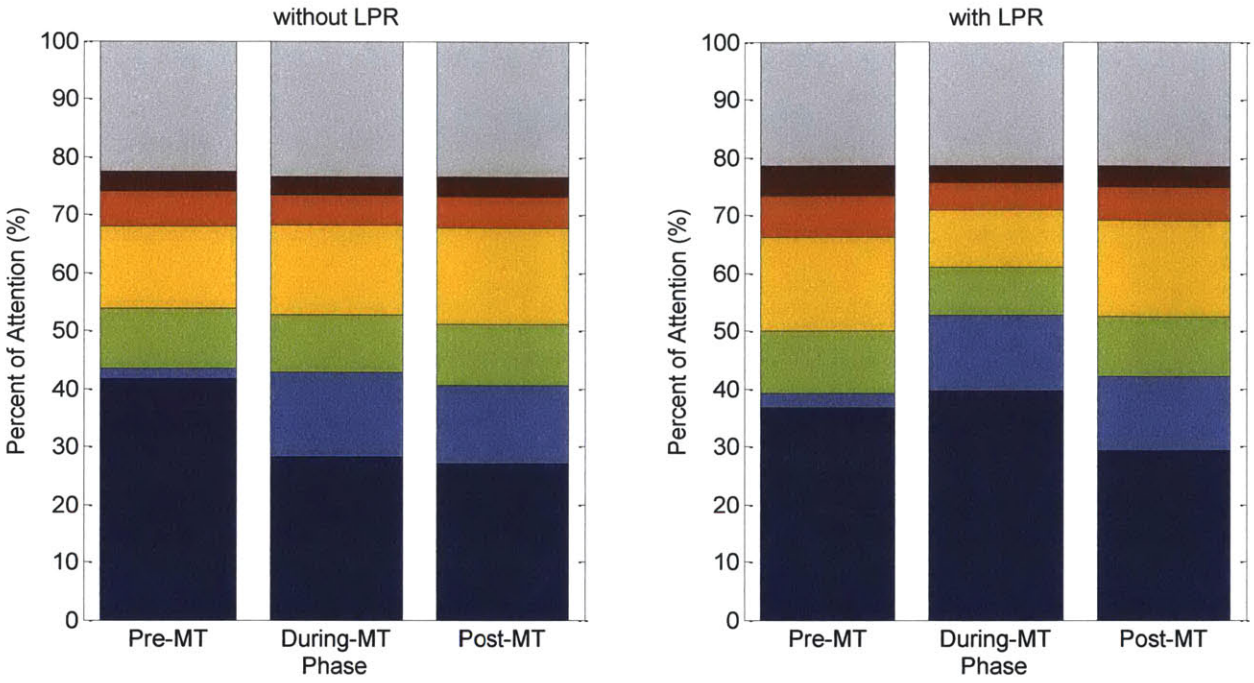
Auto→TA



Auto→TA+RoD



TA→TA+RoD



Number of Fixations and Dwell Durations for All Mode Transitions

The following section shows the number of fixations and dwell duration (similar to Figure 10) for all six mode transitions, with and without landing point redesignations (LPR). The number of fixations is measured on the y-axis, and the dwell duration is indicated by the size of the data points. The legend for all figures is the same, and is shown in Figure 38.

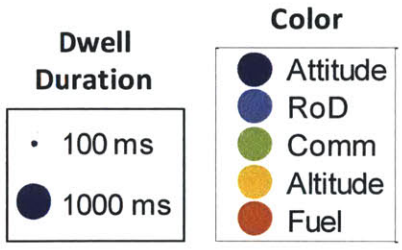
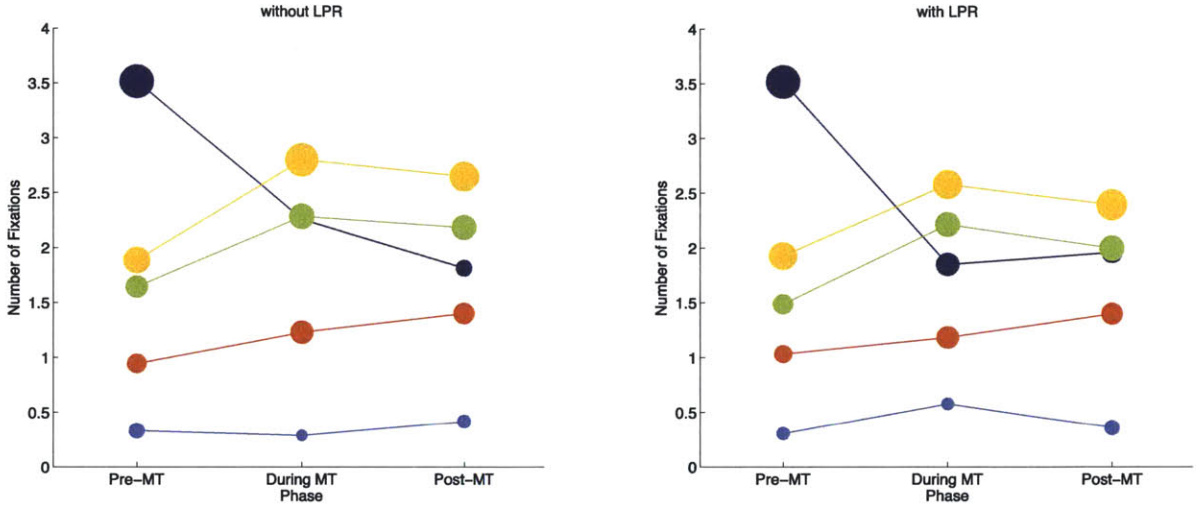
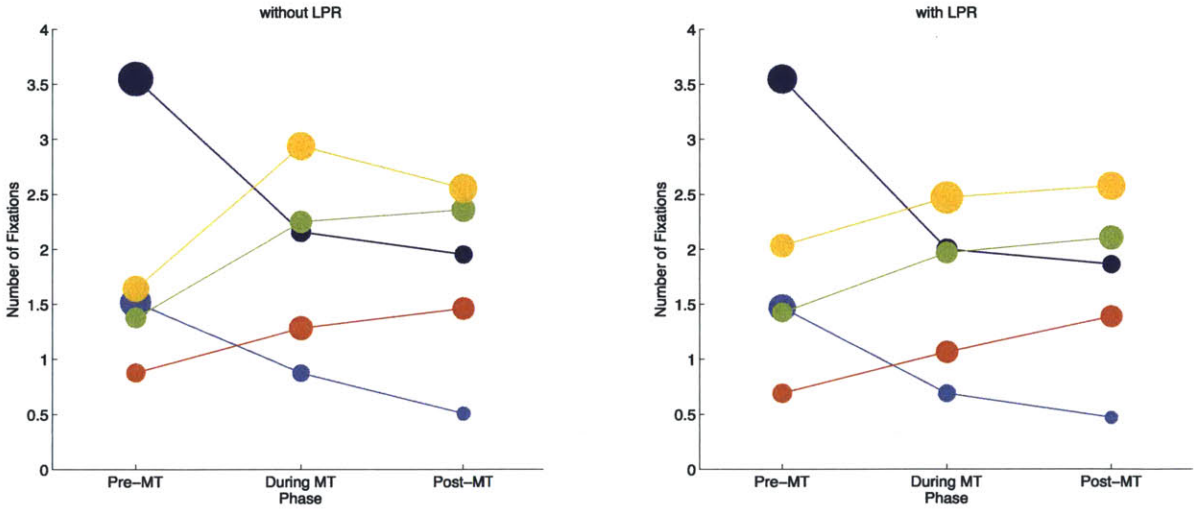


Figure 38. Legend for plots of fixation and dwell duration. The dwell duration legend is to-scale with the rest of the figures presented.

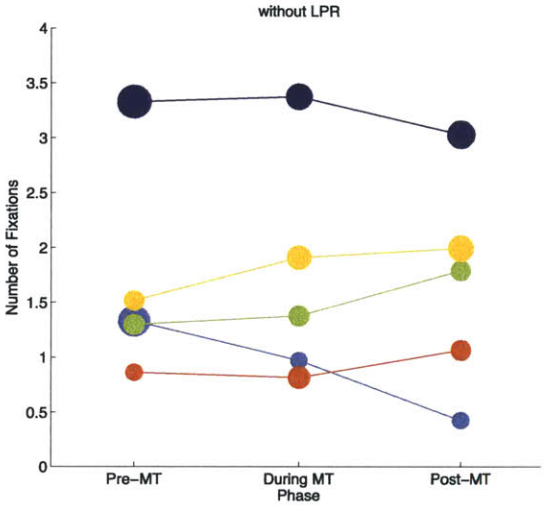
TA→Auto

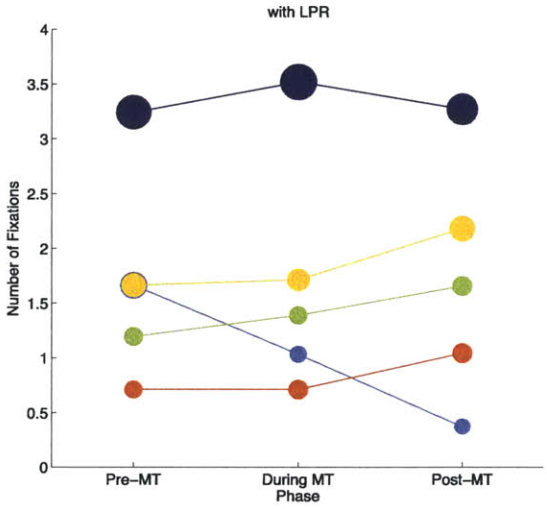


TA+RoD→Auto

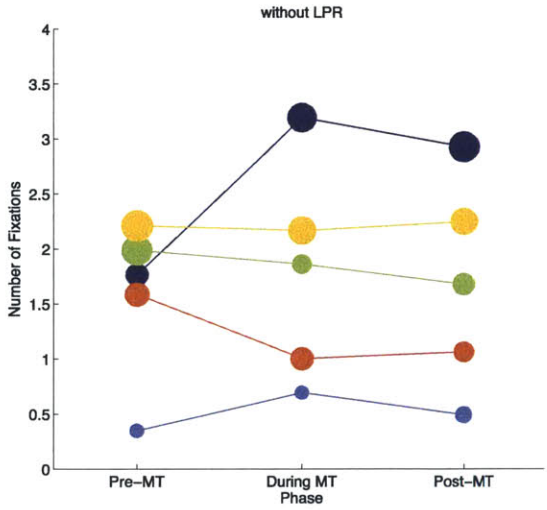


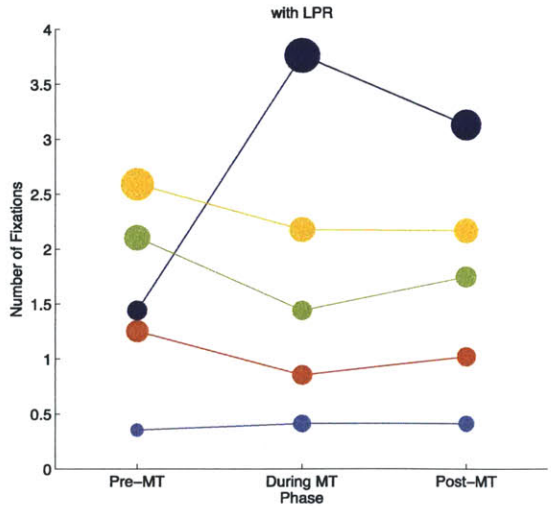
TA+RoD→TA



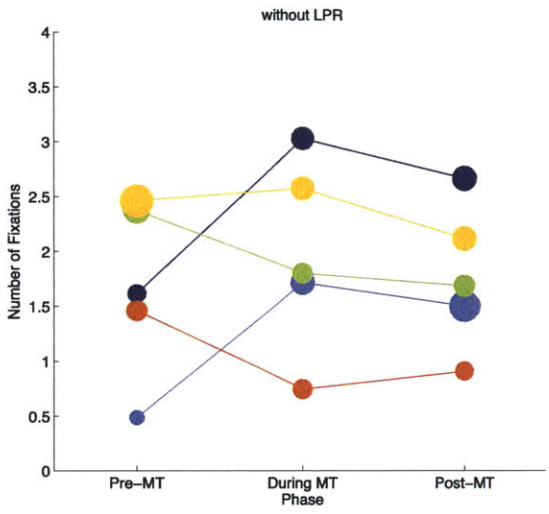


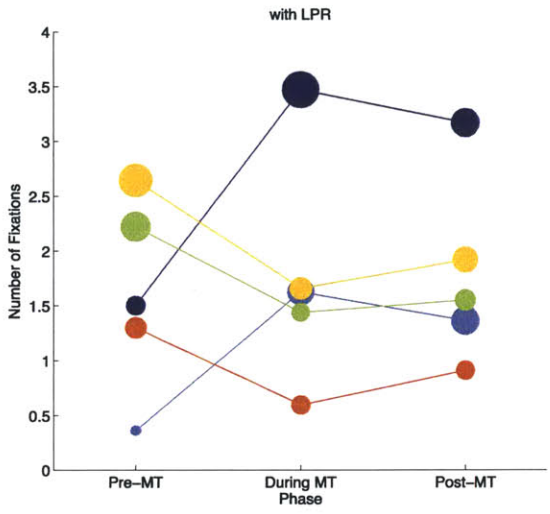
Auto→TA



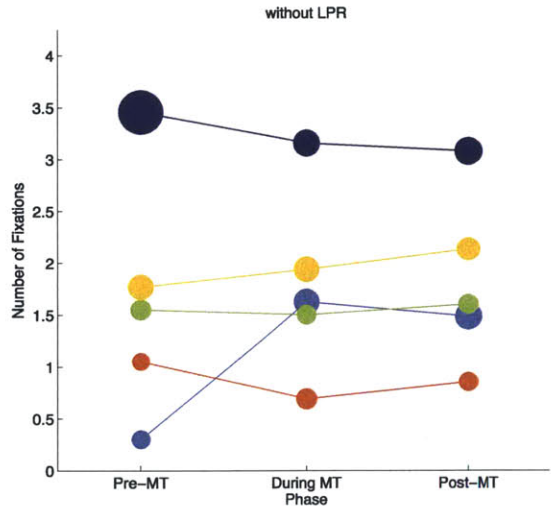


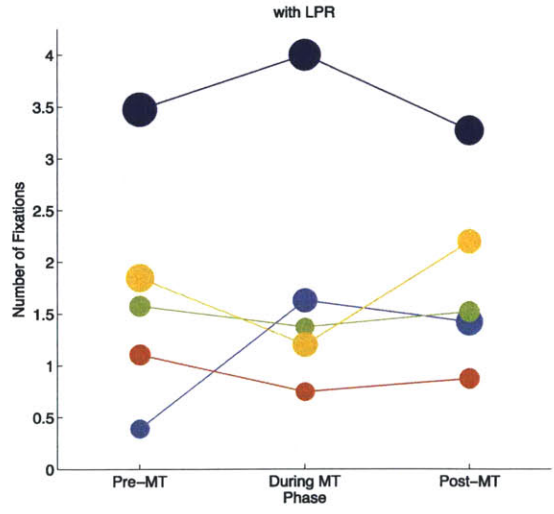
Auto→TA+RoD





TA→TA+RoD

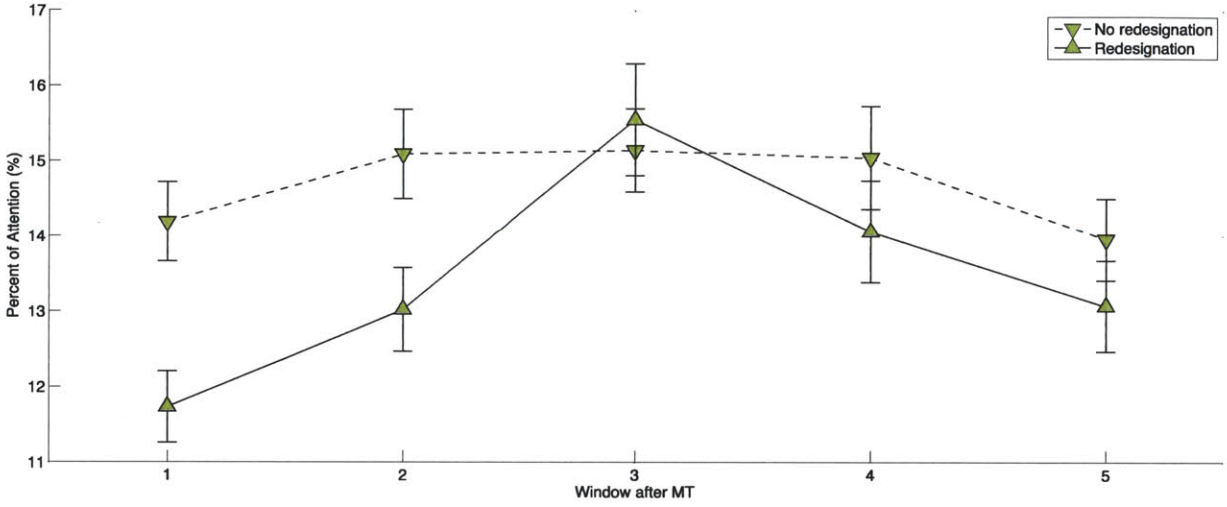




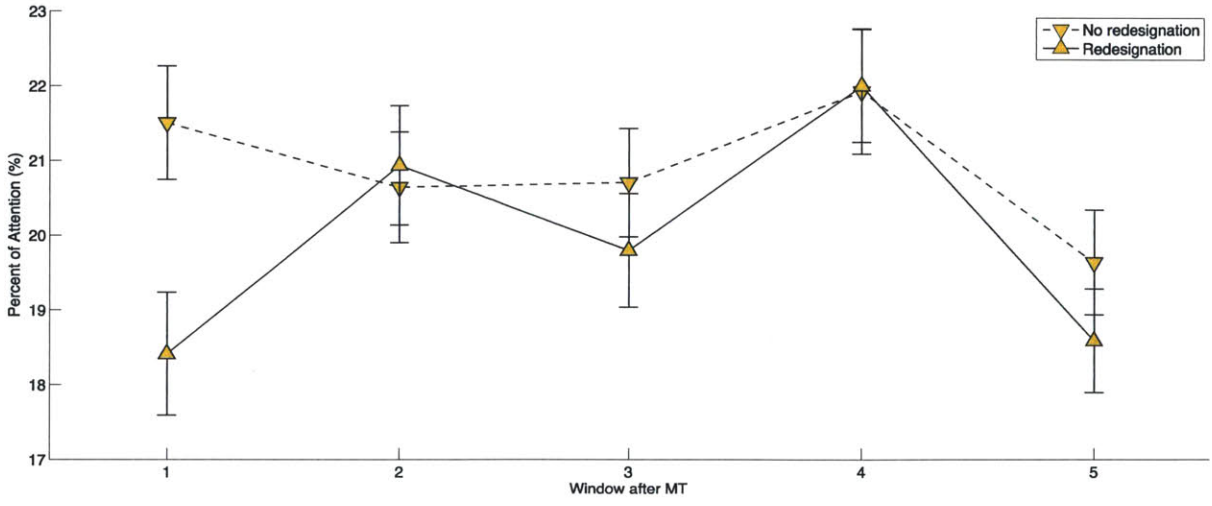
Transient Effects on the Secondary Instruments Post-Mode Transition

This section shows the post-mode transition transient behaviors (similar to Figure 14) in the percent of attention on the comm light, altitude indicator, and fuel indicator. The data represents the four mode transitions ending in manual control modes (TA or TA+RoD), and is split by whether or not there was a landing point redesignations (LPR).

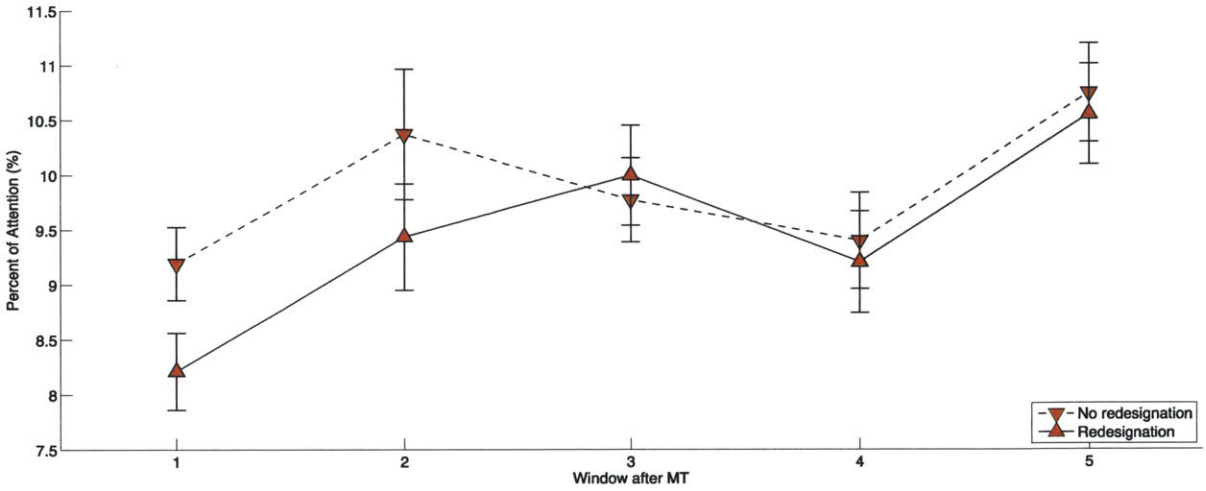
Comm Light



Attitude Indicator



Fuel Indicator



Appendix C. Experiment Training Slides

After subjects performed the initial eye tracker calibrations and were approved to participate in the study, the experimenters walked them through the following training slides (prepared by Kaderka, 2014). After this initial presentation, subjects were given further hands-on training in the simulator. This session that lasted approximately 1 hour. During this training session, they were taught how to fly the lunar lander, complete the secondary and tertiary tasks, and detect and diagnose failures. The experimenters gave comments and suggestions to the subjects and answered questions.

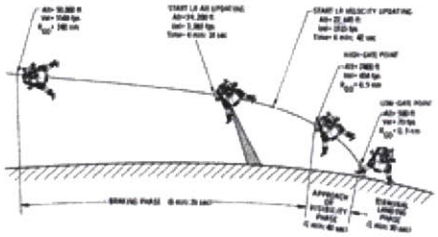
Lunar Landing Simulator Training Slides

Aaron Johnson
Justin Kaderka
Draper: Kevin Duda
MIT: Chuck Oman

1

Introduction

- A typical lunar trajectory has 3 phases
  - Braking phase (deceleration out of orbit)
  - Approach phase (to establish visual contact with the surface)
  - Terminal descent phase (pilot directs vehicle down to the surface)
- This experiment focuses on the terminal descent phase of landing



2

Introduction

- Goal of today's experiment:
  - Investigate the instrument scanning behavior of pilots in simulated lunar landing scenarios

3

The Scenario

- You'll be flying a lunar lander in a simulated terminal descent
- Several displays will be available to assist you, and you will make use of several control modes
- A flight director will also assist your landing efforts
- There will be a total of about 70 Simulations (about 70 seconds each)

4

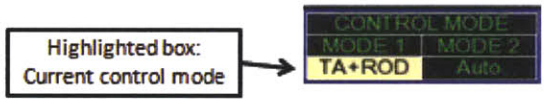
The Scenario continued...

- You begin in one of three control modes (**Mode 1**), which may or may not require you to fly the vehicle.
- After 20s, you will be required to make a mode transition to one of three control modes (**Mode 2**), which may or may not require you to fly the vehicle.
- You will need to refer to the mode annunciator to see your current control mode.
- During the trial you are asked to:
  1) Null guidance errors
  2) Monitor system states and detect and diagnose any system failures that may occur.

5

Control Modes

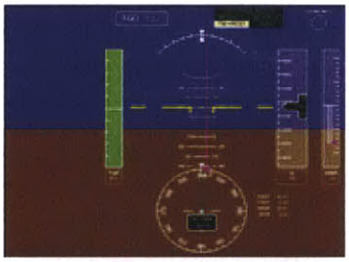
- Three control modes
  - Autopilot [**Auto**]
  - Triple axis control [**TA**]
    - Rate command attitude hold (RCAH)
      - Pitch, roll, and yaw
    - Joystick inputs command angular velocities
    - Hands off joystick – spacecraft holds attitude
    - Rate of descent controlled by autopilot
  - Triple axis and rate of descent control [**TA+ROD**]
    - RCAH (pitch, roll, and yaw) identical to previous mode
    - Incremental rate of descent – 1ft/s intervals
- Mode Annunciator Panel (shown in next slides)



6

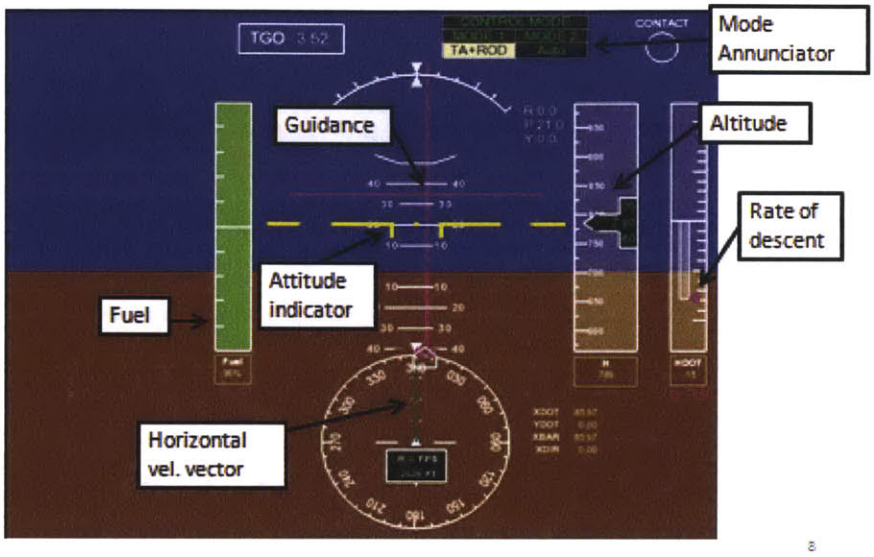
Displays

- Two displays will be provided
  - Primary flight display (PFD)
    - Provides information about vehicle states, such as attitude, horizontal and vertical velocity
    - Provides flight director cues
    - Uses a simulated horizon; does not show out-the-window information
  - Horizontal Situation Display (HSD)
    - Provides information about hazards and recommended landing aimpoints
    - Provides information about the amount of fuel remaining through the use of achievability contours
    - Top down (bird's eye) view of terrain



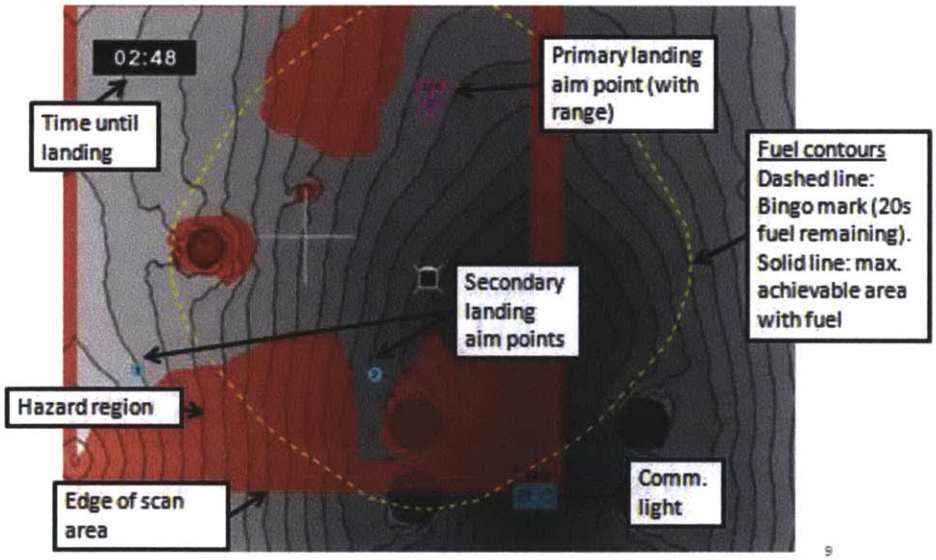
7

Displays – Primary Flight Display



8

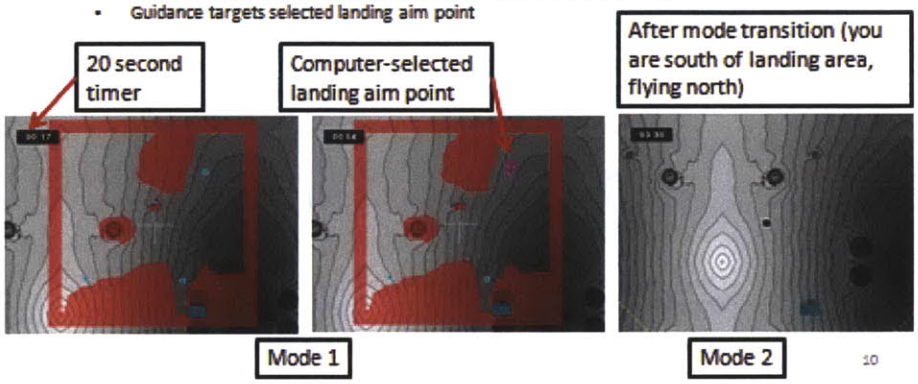
Displays – Horizontal Situation Display (1/2)



9

Displays – Horizontal Situation Display (2/2)

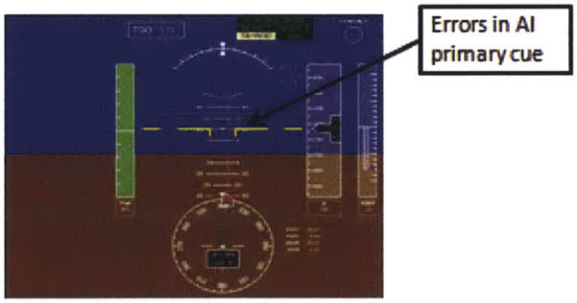
- Mode 1:
  - HSD is static map showing possible landing aim points
  - Guidance targeting cross-hairs
- Once timer reads 5s, the computer selected landing aim point is highlighted (cross-hairs can also be highlighted)
- Once timer reads 0s, hit the mode transition button to transition to a different control mode
- Mode 2:
  - HSD is moving bird's eye map (spacecraft is fixed in center of screen)
  - Guidance targets selected landing aim point



10

Failure 1/3 – Thruster stuck on

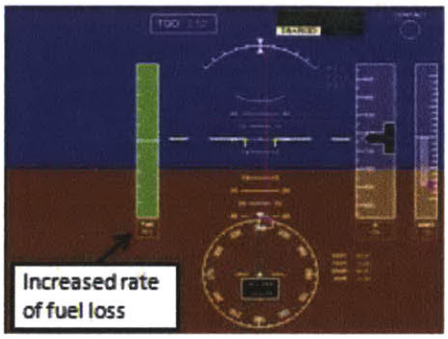
- A thruster may fail on (firing)
  - Effect in **manual control** – thruster stuck on results in **constant pitch and roll inputs** (pitch down, roll left)
  - Effect in **autopilot control** – autopilot system allows pitch and roll errors to accumulate within a dead band of 5 deg., then nulls errors, then allows errors to accumulate again. Results in a **cycling effect of attitude indicator**.
- Primary cues – unexpected movements in attitude indicator (two axes)
- Secondary cue – An increased number of inputs to joystick in manual control

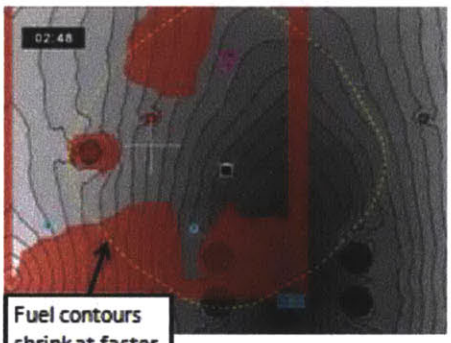


11

Failure 2/3 – Fuel Leak

- Fuel leak may occur and cause a decrease in amount of descent engine fuel available
- Primarily cue – fuel contours shrink at a faster rate
- Secondary cue – percent fuel remaining in fuel gauge decreases with faster rate





12

Failure 3/3 – Faulty landing radar

- Landing radar malfunctions and causes faulty signals (noise) in altitude, rate of descent, and horizontal velocity
  - Recognized by errors between guidance and attitude, errors in position velocity vector
- Primary cue – Noise in guidance needles (in both axes)
- Secondary cues – Increased noise level in altitude, rate of descent, and noise in horizontal velocity

13

Failure Summary

- Thruster stuck on
  - Primary cue – errors in attitude indicator (two axes) and motion cues (if present)
- Fuel leak
  - Primary cue – fuel contours shrink faster
- Faulty landing radar
  - Primary cue – noise in guidance needles (in both axes), and noise in horizontal position velocity

14

Controllers

Throttle

Joystick

15

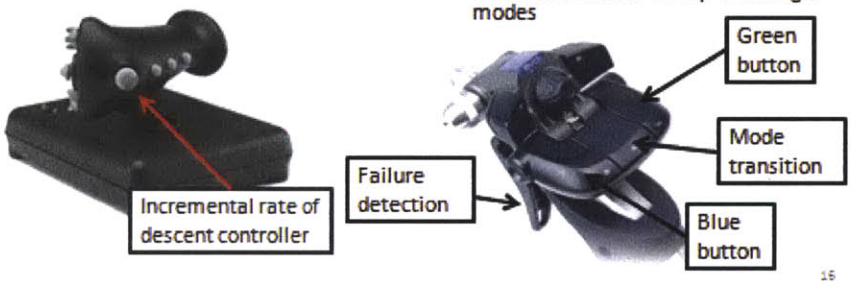
Controllers

Throttle

- Press up/down on button to decrease/increase rate of descent by 1 ft/s

Joystick

- Fore-aft motion of the stick commands pitch, and left-right motion commands roll
- Use the trigger to detect failures and another trigger pull to diagnose failures
- Use green and blue buttons on top of joystick to answer alert requests
- Use center button on top to change modes



16

Other Communications

- At various intervals, ground control will request your attention (as designated by a lit "Comm" light on the landing area display)
- The "Comm" light will be transparent (neutral) if you have attended to all requests, and will appear lit (blue or green) if your attention is needed
- It is your responsibility to attend to these requests
  - If the left light is lit, press the blue button on the joystick
  - If the right light is lit, press the green button on the joystick
  - The light will turn off once you press the correct button
- Try to attend to the requests as quickly as you can
- However, do NOT compromise your main task (detection of any system failures and piloting the spacecraft)
  - Only address the "Comm" signal if it will not hinder your flying performance / failure detection

Neutral

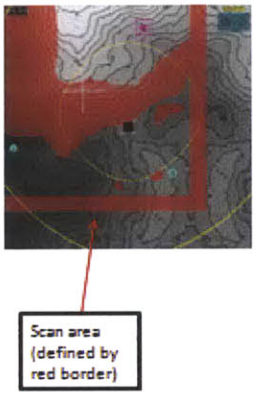
Blue

Green

17

Situational Awareness

- Three types of callouts:
  - Altitude, fuel, and landing area display landmarks
- Passing Altitude Callout
  - Every 100 ft until 500ft, every 50 ft afterwards
- Fuel Percent Change Callout
  - Every 5% increments (e.g. 95%, 90%, etc)
- Landing area display landmarks
  - "Scan area" - Scan area in view
  - "Over scan" - Crossing into scan area (lander is over the edge of the scan area – hazard zone)
  - "Landing target" - Selected LAP in view
- DO NOT SACRIFICE THE FLYING OR COMM TASKS!



18

Goals

- Your primary goals:
  - Monitor system states and detect and diagnose any system failures
  - Null the guidance errors to ensure an accurate and safe landing
    - Roll and Pitch within 1 degree
    - Descent rate: within 0.5 ft/s
- Your secondary goal:
  - Minimize response time to communication signals
- Your tertiary goal:
  - Make verbal callouts during trial
- If you cannot attend to all tasks, shed callouts first, and if necessary, then shed the comm. light task
- During the trials, the simulation will terminate before touchdown

19

Today's Schedule

- Consent / Briefing (30 min.)
- Eye tracker calibration (15 min.)
- Training trials (45 min.)
- Data collection trials (70 min.)

20

Appendix D. Experiment Consent Form

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

Pilot performance in different control modes during nominal and off-nominal simulated lunar landings

You are asked to participate in a research study conducted by Charles M. Oman, Ph.D. from the Department of Aeronautics and Astronautics Man-Vehicle Laboratory at the Massachusetts Institute of Technology (M.I.T.), Kevin R. Duda, Ph.D. from The Charles Stark Draper Laboratory, Inc., and Justin Kaderka and Aaron Johnson from The Charles Stark Draper Laboratory, Inc. and the Man-Vehicle Laboratory at M.I.T. You were selected as a possible participant in this study because NASA and the National Space Biomedical Research Institute are interested in understanding how to best design the human-machine interface used to control the lunar lander for future lunar missions. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this research is completely VOLUNTARY. If you choose to participate you may subsequently withdraw from the study at any time without penalty or consequences of any kind. If you choose not to participate, that will not affect your relationship with M.I.T. or Draper Laboratory or your right to health care or other services to which you are otherwise entitled.

- **PURPOSE OF THE STUDY**

The purpose of this experiment is to measure the performance of participants flying an Apollo-like lunar lander simulator during candidate approach and landing trajectories while interacting with various levels of automation within the flight control system. The research goals are to evaluate pilots' supervisory control performance during expected or unexpected vehicle control mode transitions and under different task allocations, as well as to evaluate the pilots' ability to detect and identify spacecraft failures while in different control modes. The results will help further understand human-automation interactions in complex space systems, such as a lunar lander.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to do the following things:

At the start of your session, you will be asked to answer survey questions about your piloting and video game experience, knowledge with respect to spacecraft systems, as well as standard biographical information. Prior to starting the simulation, you will be informed of the task you will be asked to complete and informed of how to interact with the simulation (i.e. what the

display symbology means and how to control the simulated vehicle with the joysticks). This will be done with both a presentation and the experimenter demonstrating the simulation to you. You are encouraged to ask any questions that you may have regarding the protocol.

Before the simulation begins, you will be asked to look at a regularly-spaced grid of dots in order to calibrate an eye tracker. This equipment, mounted away from you and below the displays in the simulator, will record your visual attention during the training and experimental trials. After the eye tracker calibration, the experimental training session will commence and you will learn to operate the lunar lander. The trial may begin with landing point selection in which you will select a landing aim point from several candidate points using a joystick button or mouse and a visual display of terrain information. The lunar lander will fly in one of four control modes (three of which require manual inputs from you using a joystick and throttle controller) during this initial phase in which you may or may not be required to select a landing point. After the initial flying phase, the control mode may change unexpectedly or you may initiate a control mode change from the current mode to one of the four control modes. In the manual flight control modes you will be responsible for flying the vehicle during final descent using the joystick controllers and flight displays. Your primary task is to fly the simulated lunar lander and maintain control throughout the trial. Your required interaction to fly the simulator will change according to the active manual control mode, which is indicated on one of the flight displays. At any point in the trial you may be asked to identify system failures (manifesting on the flight displays) and diagnose which of the system failures has occurred. Or, a failure may be identified for you, and you may be asked to perform a new failure response task. In addition, the simulation may be paused during the trial and you will be asked to estimate a system state, such as pitch, roll, rate of descent, etc. During each trial you will also be asked to respond to a secondary task, which will require you to differentiate between one of two illuminated colors on a flight display. A tertiary task may require you to make verbal callouts of system states throughout the trial (e.g., altitude, altitude rate, estimated time until touchdown). Each of these system states will be available to you on one of the flight displays. After each trial, you may be asked to complete standardized tests (e.g. NASA TLX, Cooper Harper, SART, Modified Bedford) or answer questions about the final system state to assess your workload and situational awareness. Once you are proficient at flying the lunar lander, and the training session has been completed, the test session will commence and you will complete trials similar to the training session.

Your total experiment time will last no more than 3 hours per day over one or two days. Breaks will be offered at least every 30 minutes, but you may request a break at any time during the experiment. The entire session will take place in Draper Laboratory's fixed-base lunar lander cockpit simulator.

- **POTENTIAL RISKS AND DISCOMFORTS**

    - Boredom due to the large number of repetitive trials.
    - Fatigue from operating the joystick and attending to the displays and tasks
    - Symptoms of simulator sickness due to visual motion in the displays.

You will be given short breaks between trials to reduce the risks of boredom, fatigue and motion sickness. You may request a break at any time during the experiment if you begin to feel any discomfort.

- **ANTICIPATED BENEFITS TO SUBJECTS**

There are no benefits to you aside from becoming familiar with the potential human-automation interactions, displays, and tasks associated with future lunar landings.

- **ANTICIPATED BENEFITS TO SOCIETY**

NASA may benefit from the results of these experiments by being able to design appropriate human-machine interfaces for space missions of this type.

- **PAYMENT FOR PARTICIPATION**

You will receive $10 per hour for your participation. Payment is prorated on the basis of time spent if you decide to withdraw.

- **PRIVACY AND CONFIDENTIALITY**

The only people who will know that you are a research subject are members of the research team. No information about you, or provided by you during the research will be disclosed to others without your written permission, except: if necessary to protect your rights or welfare, or if required by law.

No personal information will be collected in this experiment. All data collected in this experiment will be coded to prevent the identification of the data with a specific person. When the results of the research are published or discussed at conferences, no information will be included that would reveal your identity. The data will be archived when the project is completed and papers published (approximately during the 2013-14 academic year). No identifying information will be kept with the data

- **WITHDRAWAL OF PARTICIPATION BY THE INVESTIGATOR**

The investigator may withdraw you from participating in this research if circumstances arise which warrant doing so. If you experience any of the following side effects (simulator sickness or fatigue) or if you become ill during the research, you may have to drop out, even if you would like to continue. The investigator, Charles Oman, will make the decision and let you know if it

is not possible for you to continue. The decision may be made either to protect your health and safety, or because it is part of the research plan that people who develop certain conditions may not continue to participate.

If you must drop out because the investigator asks you to or because you have decided on your own to withdraw, you will be paid $10.

- **NEW FINDINGS**

During the course of the study, you will be informed of any significant new findings (either good or bad), such as changes in the risks or benefits resulting from participation in the research or new alternatives to participation, that might cause you to change your mind about continuing in the study. If new information is provided to you, your consent to continue participating in this study will be re-obtained.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact:

Principal Investigator: Charles M. Oman, Ph.D., (617) 253-7508, coman@mit.edu
Co-Investigator: Kevin R. Duda, Ph.D., (617) 258-4385, kduda@draper.com
Research Assistant: Justin Kaderka, (617) 258-2098, jkaderka@draper.com
Research Assistant: Aaron Johnson, (617) 258-4751, awjohnson@draper.com

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of

Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I have read (or someone has read to me) the information provided above. I have been given an opportunity to ask questions and all of my questions have been answered to my satisfaction. I have been given a copy of this form.

BY SIGNING THIS FORM, I WILLINGLY AGREE TO PARTICIPATE IN THE RESEARCH IT DESCRIBES.

_____

Name of Subject

_____

Name of Legal Representative (if applicable)

_____ _____

Signature of Subject or Legal Representative Date

SIGNATURE OF INVESTIGATOR

I have explained the research to the subject or his/her legal representative, and answered all of his/her questions. I believe that he/she understands the information described in this document and freely consents to participate.

_____

Name of Investigator

_____ _____

Signature of Investigator Date (must be the same as subject's)

Appendix E. Additional Details on Model

Actual Non-linearized Attitude Dynamics

A linearized model of the lunar lander attitude dynamics is used to perceive, estimate, and control the lunar lander attitude:

$$\theta = \frac{8.57}{0.12s^2 + s}\delta_\theta \tag{E.1}$$

$$\phi = \frac{9.68}{0.12s^2 + s}\delta_\phi \tag{E.2}$$

These linearized dynamics show there is a first-order lag from the joystick deflection, δ, to the pitch and roll rates ($\dot{\theta}$ and $\dot{\phi}$, respectively). The State Estimator uses these linearized dynamics in a Kalman filter to estimate the actual vehicle attitude (Section 5.2.3.1). Also, the operator control gain (G) is selected based on an analysis of these linearized dynamics (Section 5.2.3.2). However, the actual dynamics of the system are non-linear. Figure 39 shows a block diagram of the actual, non-linear pitch dynamics:

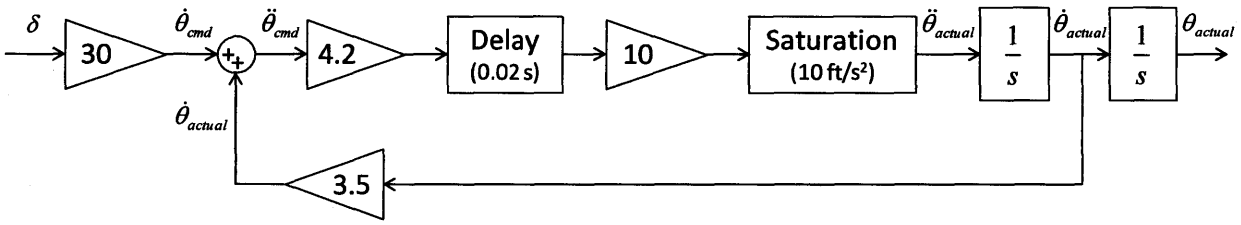


Figure 39. Block diagram of actual, non-linear pitch dynamics.

Combining these blocks, the pitch acceleration ($\ddot{\theta}$) is related to the joystick deflection ($\square$) and the pitch rate ($\dot{\theta}$) by the equation:

$$\ddot{\theta} = 1260\delta - 147\dot{\theta} \tag{E.3}$$

Because the gain on the joystick deflection is much higher than the gain on the pitch rate feedback, a small deflection of the joystick causes a large pitch acceleration. This acceleration is quickly saturated to ±10 °/s^2. With a constant deflection, the pitch acceleration returns to 0 when the pitch rate reaches the steady-state value of $\dot{\theta}_{ss} = 8.57\delta_\theta$. This number is found by setting $\ddot{\theta} = 0$ in Equation (E.3) and solving for $\dot{\theta}$.

The time constant of the linearized system represents the time it takes the pitch rate to reach its steady-state value. Because the steady-state pitch rate is a function of the joystick deflection, the time constant is also a function of the joystick deflection. The average joystick deflection was found to be |0.25|, and this deflection was used to derive the time constant of 0.12.

The actual roll dynamics are similar to the pitch dynamics, except the gain on the rate feedback is 3.1, instead of 3.5 as it is in the pitch dynamics. This makes the steady-state roll rate $\phi_{ss} = 9.68\delta_{\phi}$.

Additional Details on the Attitude State Estimator Kalman Filter

In state-space form, the transfer functions (E.1) and (E.2) take on the following form:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ z_k &= Cx_k + v_k \end{aligned} \tag{E.4}$$

where k is the current time step, x is the vehicle state vector, z is the measurement of the vehicle state, u is the control input, w is the process noise, and v is the measurement (perceptual) noise. Please note that u, w, and v are different than the variables used in the attention model (Chapter 5). The symbols in this appendix have been chosen to conform to commonly-used descriptions of the Kalman filter (Welch and Bishop, 2006).

For the linearized lunar landing dynamics, the matrices A, B, and C (Equation (E.5.)) are constant with time. In these equations, q is the pitch rate, θ is the pitch, r is the roll rate, ϕ is the roll, and Δt is the length of a time step (in s). This state-space representation indicates that the operator controls the pitch and roll rates, but can only observe the pitch and roll.

Because the process noise is expected to be small (Section 5.2.3.2) it is modeled as white noise with a mean of 0 and a very small variance: $p(w) \sim N(0,0.00001^2)$. The 5° quantization of measurement noise (Section 5.2.3.2) is modeled as zero-mean white noise with a variance equal to $5^2/12$ (Stanley et al., 1983). Therefore, the probability distribution of v is $p(v) \sim N(0,5^2/12)$. The process and measurement noises are both constant with time.

$$x = \begin{bmatrix} q \\ \theta \\ r \\ \phi \end{bmatrix}$$

$$A = \begin{bmatrix} 1-8.33\Delta t & 0 & 0 & 0 \\ \Delta t & 1 & 0 & 0 \\ 0 & 0 & 1-8.33\Delta t & 0 \\ 0 & 0 & \Delta t & 1 \end{bmatrix} \tag{E.5}$$

$$B = \begin{bmatrix} 71.42\Delta t \\ 0 \\ 80.67\Delta t \\ 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix}$$

The Kalman filter computes an estimate of the actual state, $\hat{x}_k$, with the following equations:

$$\begin{aligned} \hat{x}_k^- &= A\hat{x}_{k-1} + Bu_{k-1} \\ \hat{x}_k &= \hat{x}_k^- + K_k\left(z_k - C\hat{x}_k^-\right) \end{aligned} \tag{E.6}$$

The top part of Equation (E.6) uses the operator's knowledge of the system dynamics and the control input to propagate the vehicle state estimate from the previous time step forwards to the current time step. The Kalman filter then computes the error between this *a priori* estimate ($\hat{x}_k^-$) and the observation (z_k). The *a priori* estimate is updated by this error, multiplied by the Kalman gain matrix *K*. The Kalman gain matrix is computed by the equation:

$$K_k = \frac{P_k^- C^T}{CP_k^- C^T + R} \tag{E.7}$$

where *R* is the measurement noise covariance ($5^2/12$) and P_k^- is the *a priori* estimate error covariance. P_k^- is computed by the equation

$$P_k^- = E\left[e_k^- e_k^{-T}\right] \tag{E.8}$$

e_k^- is the *a priori* estimate error:

$$e_k^- = x_k - \hat{x}_k^- \tag{E.9}$$

At each time step, the model solves Equations (E.7-9) to compute a new value of K. As the *a priori* estimate error decreases, so do the *a priori* estimate error covariance and the Kalman gain. The error is largest when the attitude indicator is first attended, but it quickly reduces as the fixation continues. The Kalman gain follows the same behavior. Figure 40 shows how the error between the actual and estimated pitch and the Kalman gain both reduce across one example 1-s fixation on the attitude indicator. The Kalman gain adapts on a timescale shorter than the length of the fixation, which results in an error between the actual and estimated pitch that begins large but decreases quickly to <1°.

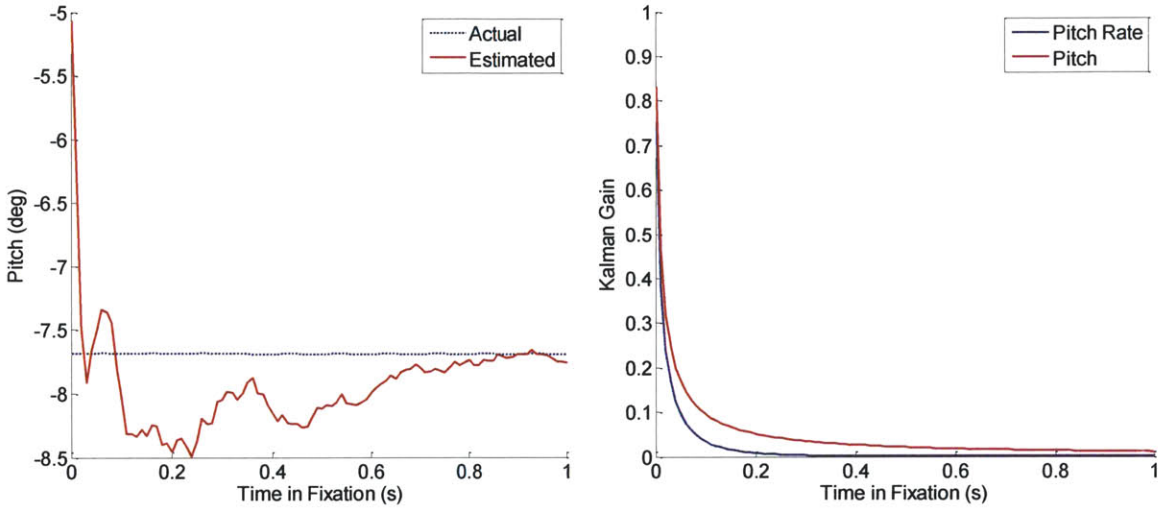


Figure 40. The error between the estimated and actual pitch (left) decreases over a 1-s example fixation on the attitude indicator. As this error decreases, so do the Kalman gains for the pitch rate and pitch (right).

Additional Details on the Effect of Operator Gain on Attitude

A block diagram of the linearized model of vehicle pitch can be seen in Figure 41. The pitch rate ($\dot{\theta}$) is connected to the operator joystick inputs (δ) through a first-order lag with a gain of 8.57 and a time constant of 0.12 s. The pitch rate is then integrated to get the pitch, θ. This linearized model removes the joystick limits of ±1 that were present in the experiment.

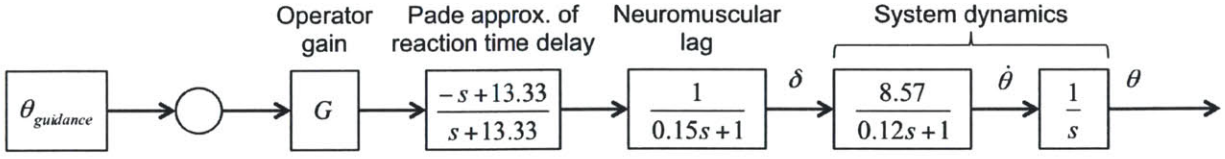


Figure 41. Block diagram of linearized system and operator model

Figure 42 and Figure 43 show the system response to a 7° change in guidance with three different operator gains (G). The magnitude of this step-change is the same that was caused by an LPR in

experimental trials. Figure 42 shows the vehicle pitch, and Figure 43 shows the error between the actual and guidance-prescribed pitch.

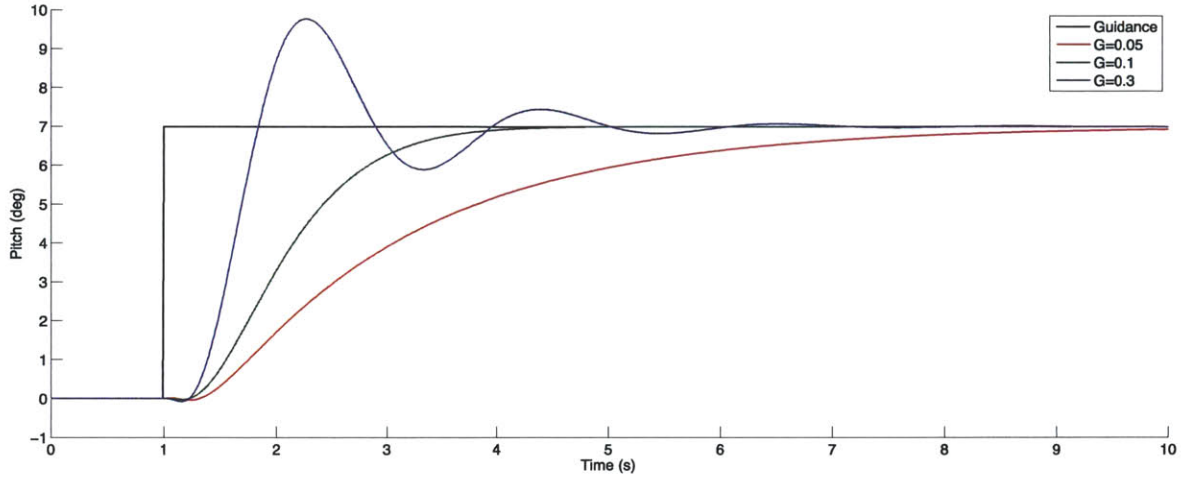


Figure 42. The response of the vehicle pitch to a 7° step change in guidance is a function of the operator gain.

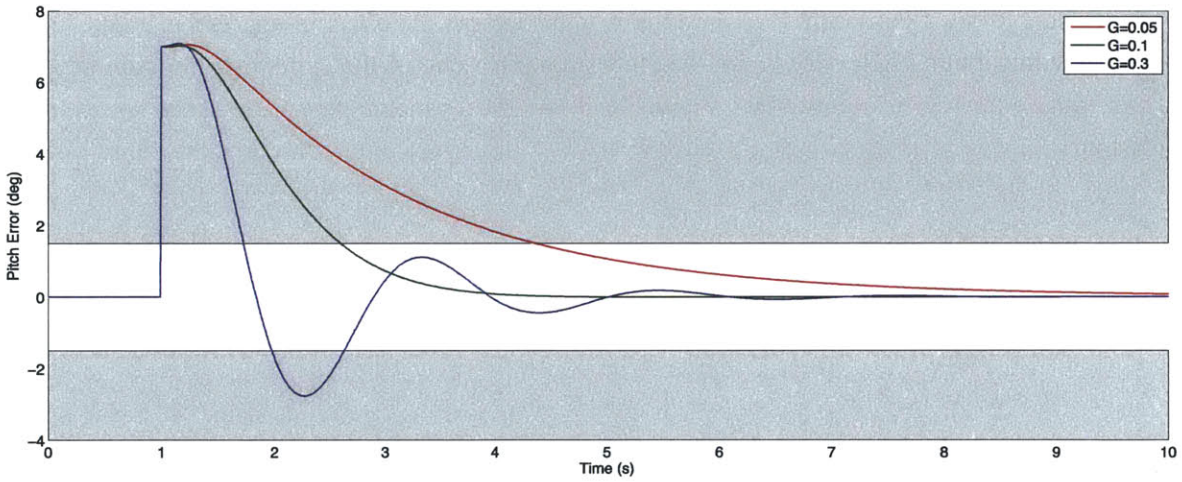


Figure 43. The error between the actual and guidance pitch is a function of the operator gain.

Is it apparent why the G=0.1 gain is preferred over the G=0.05 gain, as it only takes 1.6 s for the error to reach 1.5°, compared to 3.35 s. The effect of overshoot on attention is also demonstrated clearly by the system response when G=0.3. The error first reaches the 1.5° threshold in 0.72 s, which is faster than with the other two gains. However, the error does not reach the threshold *and remain within it* until 1.63 s.

One limitation of this linearized system is that it ignores the physical constraints of the joystick present in the actual simulator (inputs are limited to ±1). As Figure 44 shows, this can cause situations where the operator input is greater than physically allowable.

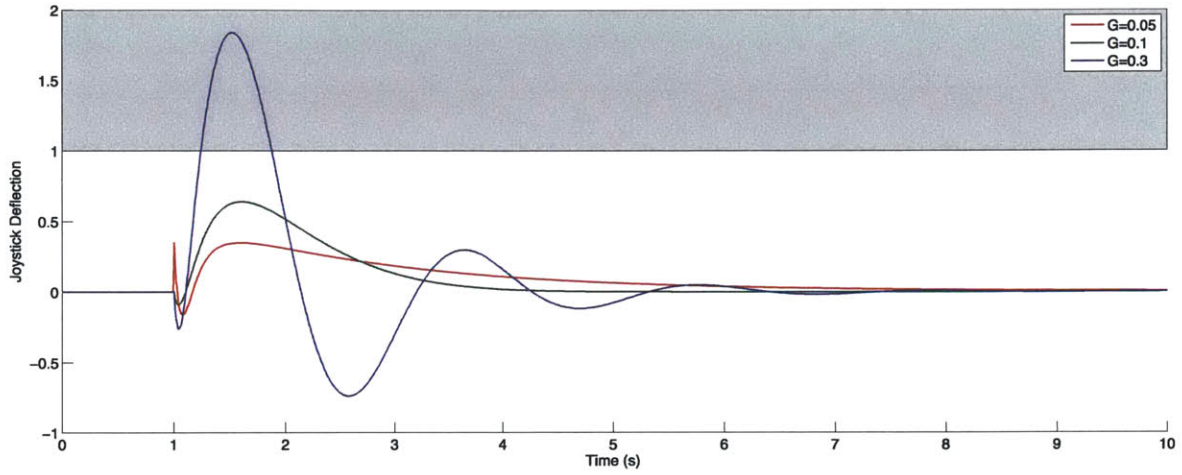


Figure 44. When the operator gain is too high (G=0.3), the joystick inputs exceed the physical limit of ±1.

Adding the joystick constraint back into the systems lessens the effect of aggressive gains, because they limit the maximum pitch rate. Figure 45 shows the pitch rate of the system with a gain of G=0.3, with and without the joystick constraints. Figure 46 shows the joystick inputs made to the system for the same gains.

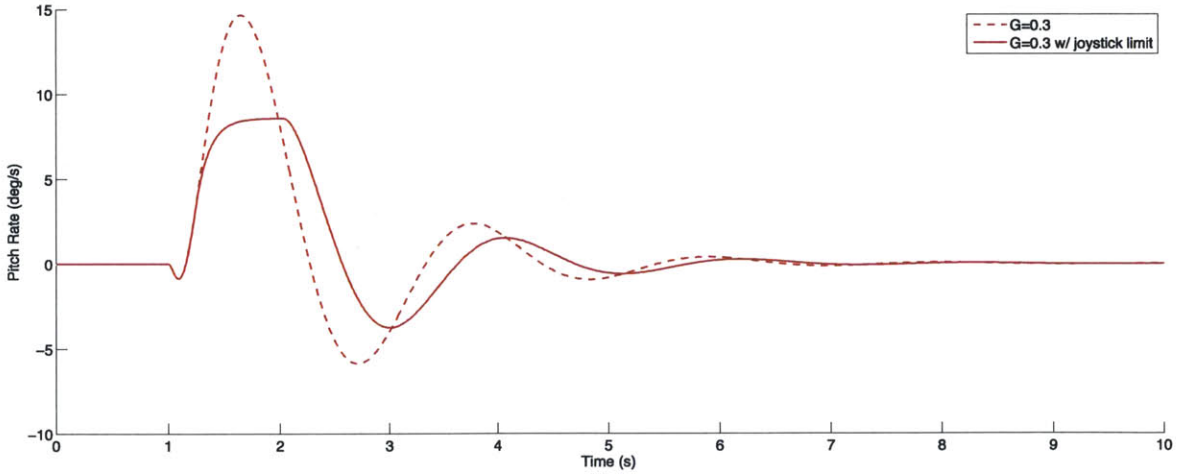


Figure 45. The joystick limits limit the maximum pitch rate of the vehicle.

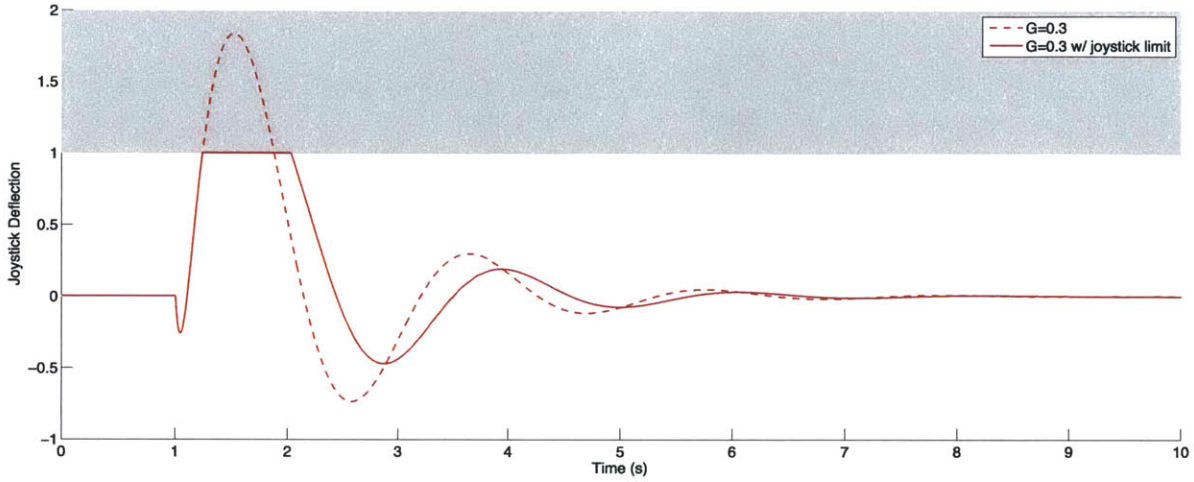


Figure 46. Joystick inputs that give the vehicle behavior in Figure 46 and Figure 47.

These two figures show that the pitch rate of both systems is identical until the joystick deflection reaches the +1 limit. The resulting effect of the joystick constraints on pitch can be seen in Figure 47, which shows the response of the system to a 7° step-change in guidance.

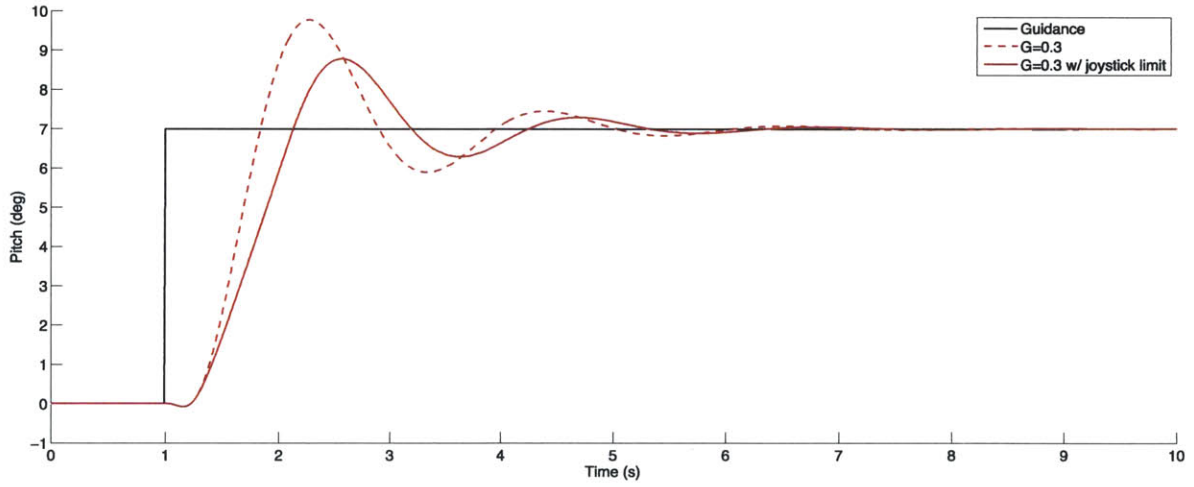


Figure 47. The joystick limits reduce the amount of overshoot with the same operator gain.

It is clear that the overshoot is less with the joystick constraints; however, it is still above the 1.5° threshold. An analysis finds that a gain of G=0.25 will produce a 1.5° overshoot for the 7° step change. This is a 15% increase from the maximum gain of G=0.217 without the joystick limits.

Open-Loop Root-Locus and Bode Plots of the Linearized System

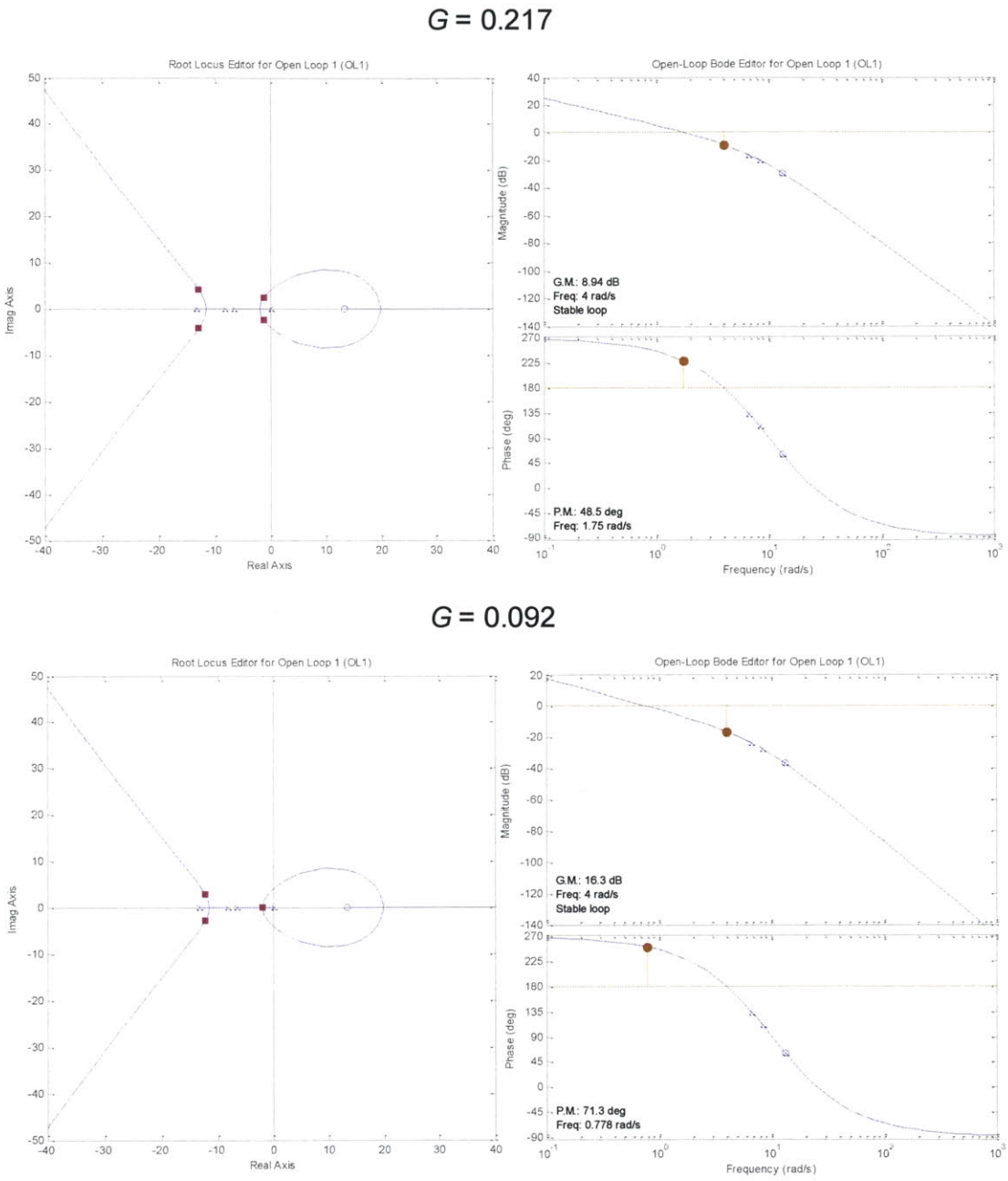


Figure 48. The open-loop root-locus and Bode plots of the linearized system with the upper and lower operator gains as calculated in Section 5.2.3.2. These plots are the output of the MATLAB SISO (single-input/single-output) Design GUI. The magenta squares on the root-locus plots show the location of the closed-loop poles with the indicated gain. The dark yellow vertical lines show the gain and phase margin at the crossover frequency.

Properties of Rate-of-Descent Display Noise

Display noise was added to the RoD signal in the experiment and the model. This noise was the sum of nine sine waves with the properties seen in Table 18, limited to ±0.25 ft/s.

Table 18. Frequency and phase of the sine waves in the RoD noise.

Frequency (Hz)	Phase (°)
0.0035	0
0.06	37
0.1325	74
0.2075	111
0.28	148
0.645	259
0.865	296
1.33	10
1.67	47

Appendix F. Additional Model Results

Attention Budgets for All Mode Transitions

The following section shows the experimental and model-predicted attention budgets (similar to Figure 6) for all six mode transitions, with and without landing point redesignations (LPR). The legend for all figures is the same, and is shown in Figure 49. The order of the instruments listed is the same order as the instruments in the stacked bars, from top to bottom.

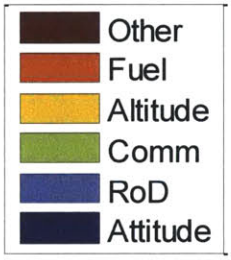
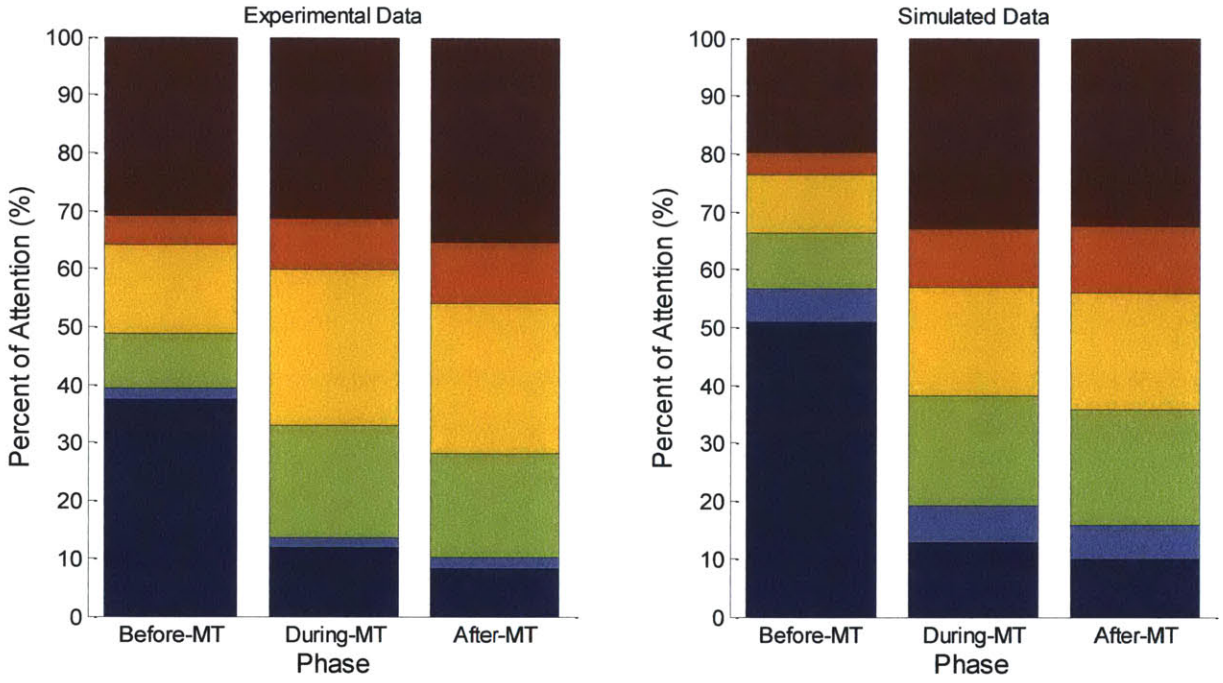


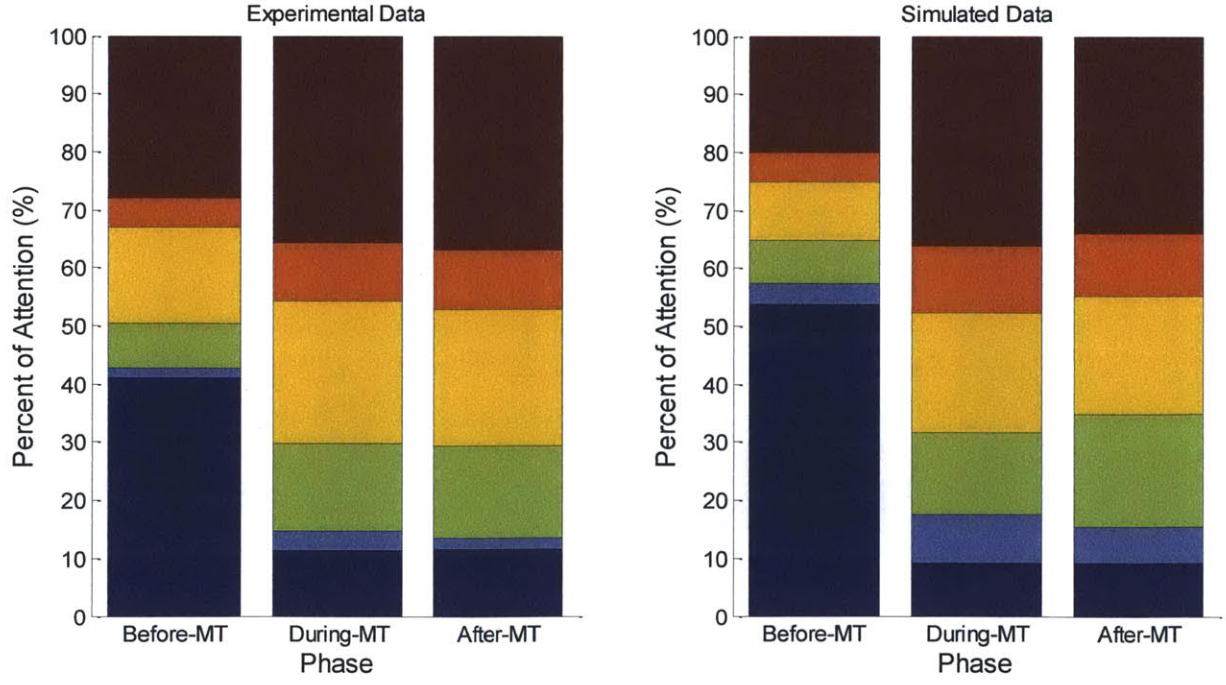
Figure 49. Legend for attention budget figures.

The figures show that while there are differences between the experimental and model-predicted data, the same general trends hold across the mode transitions.

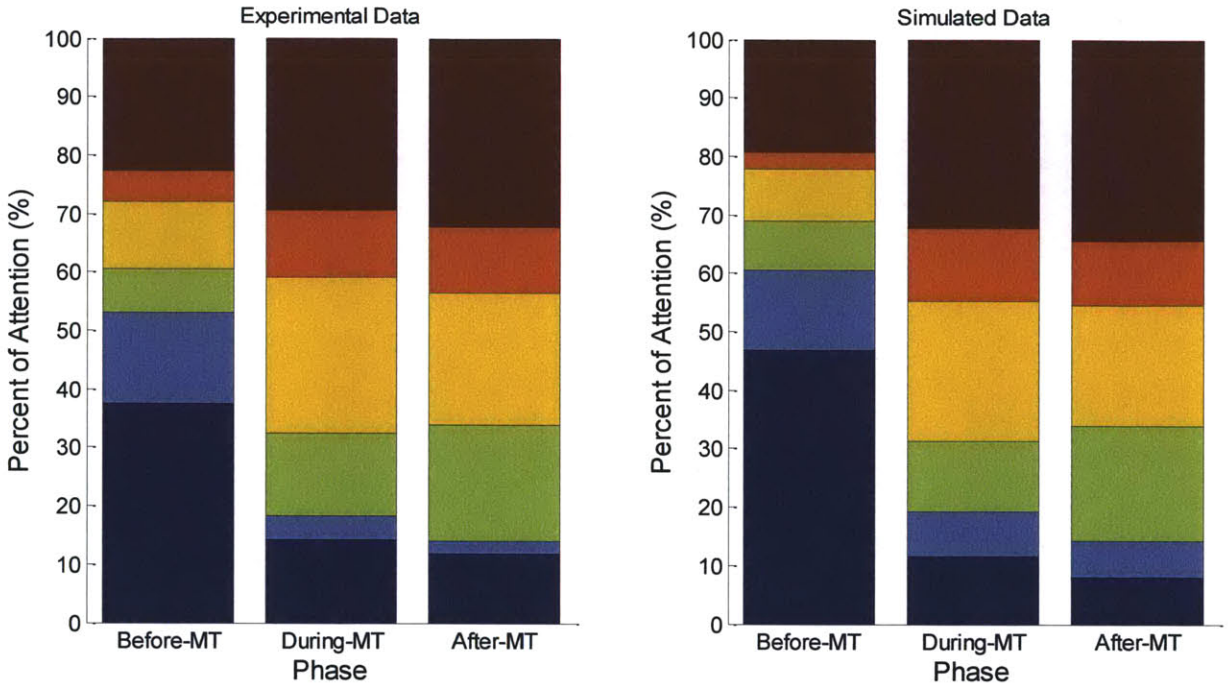
TA→Auto, without LPR



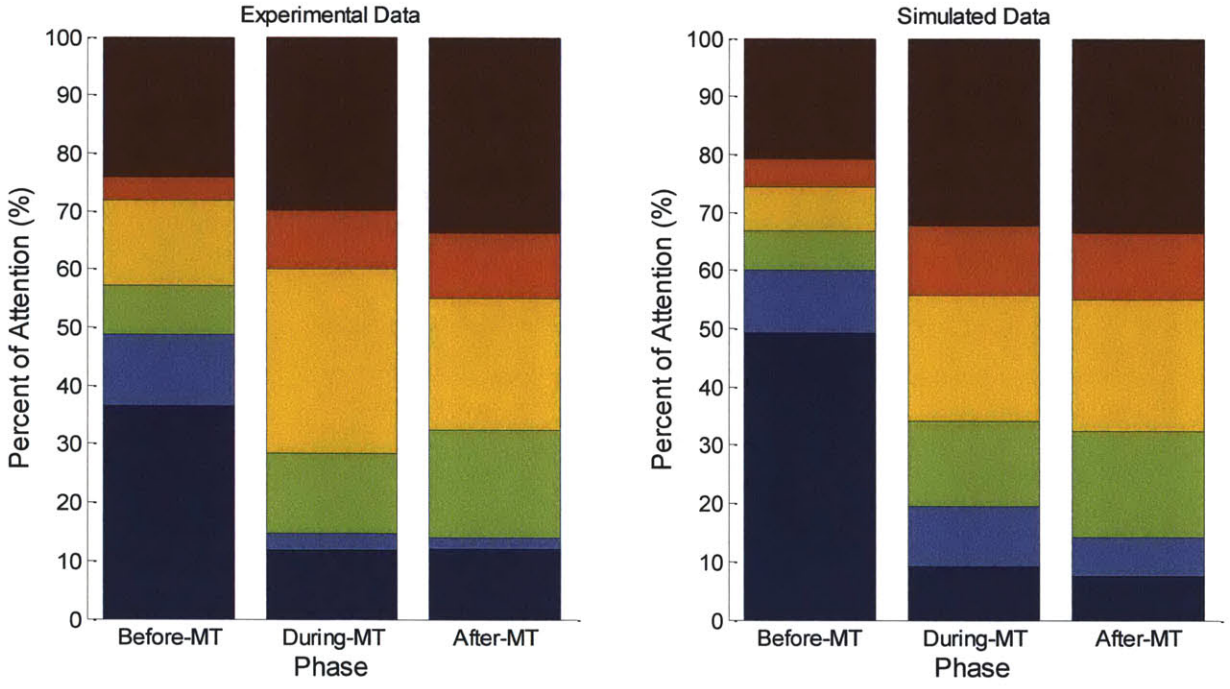
TA→Auto, with LPR



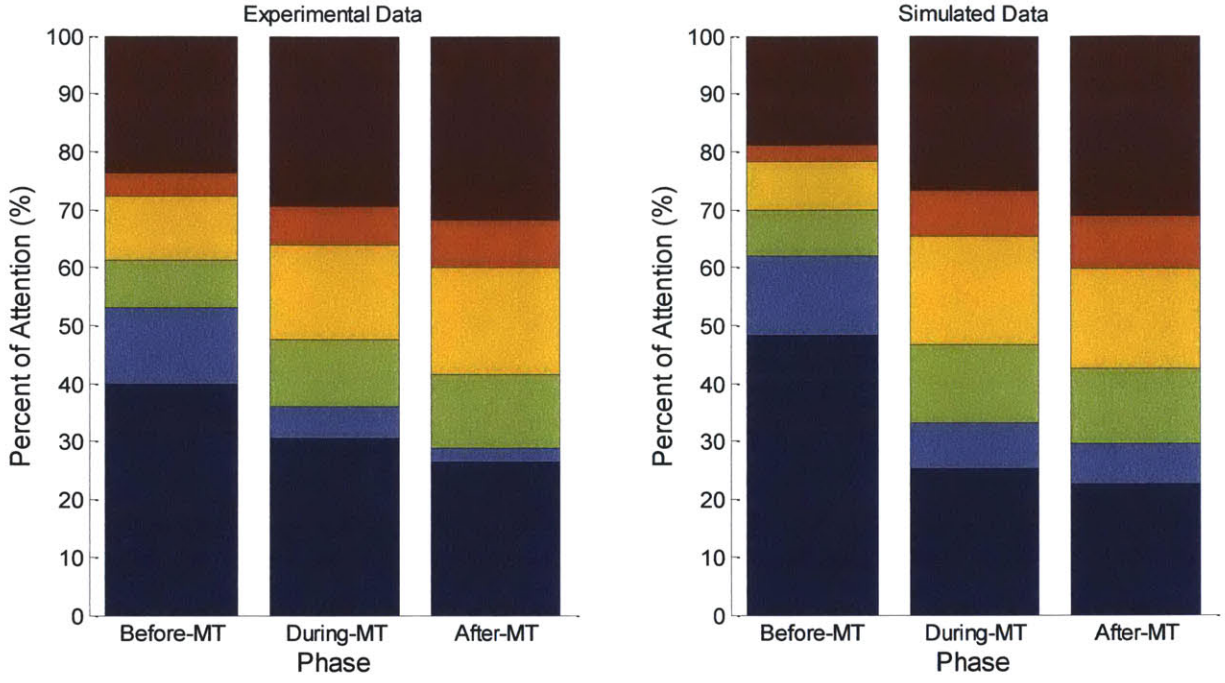
TA+RoD→Auto, without LPR



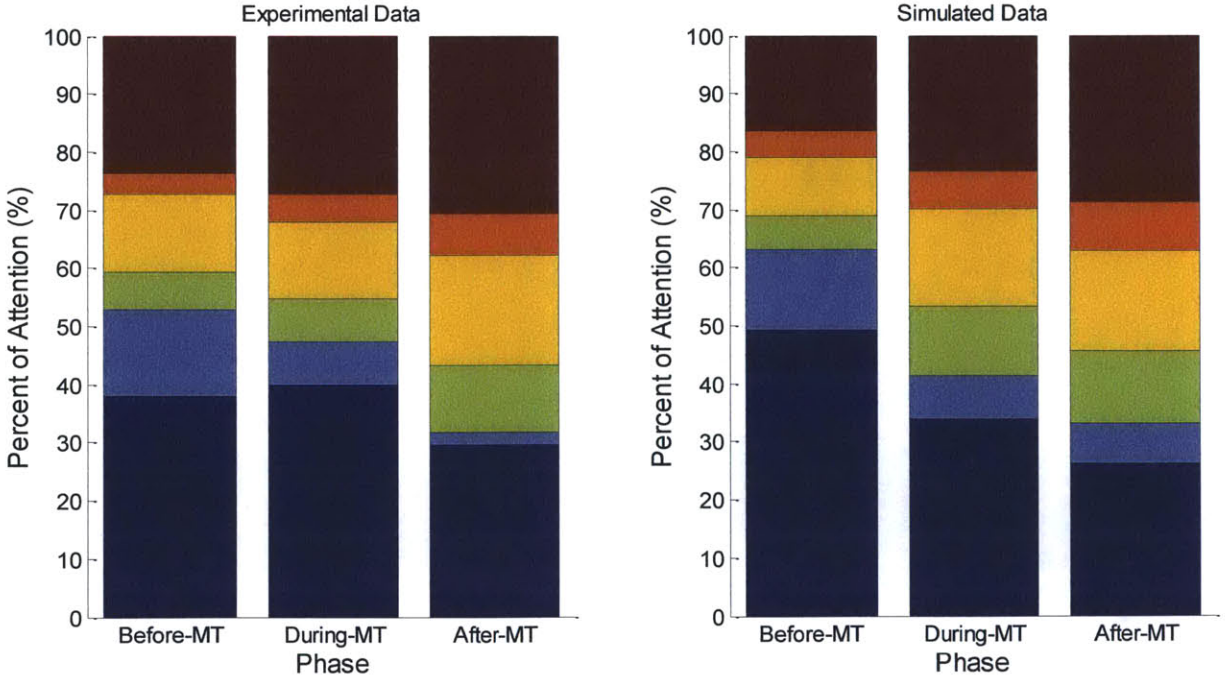
TA+RoD→Auto, with LPR



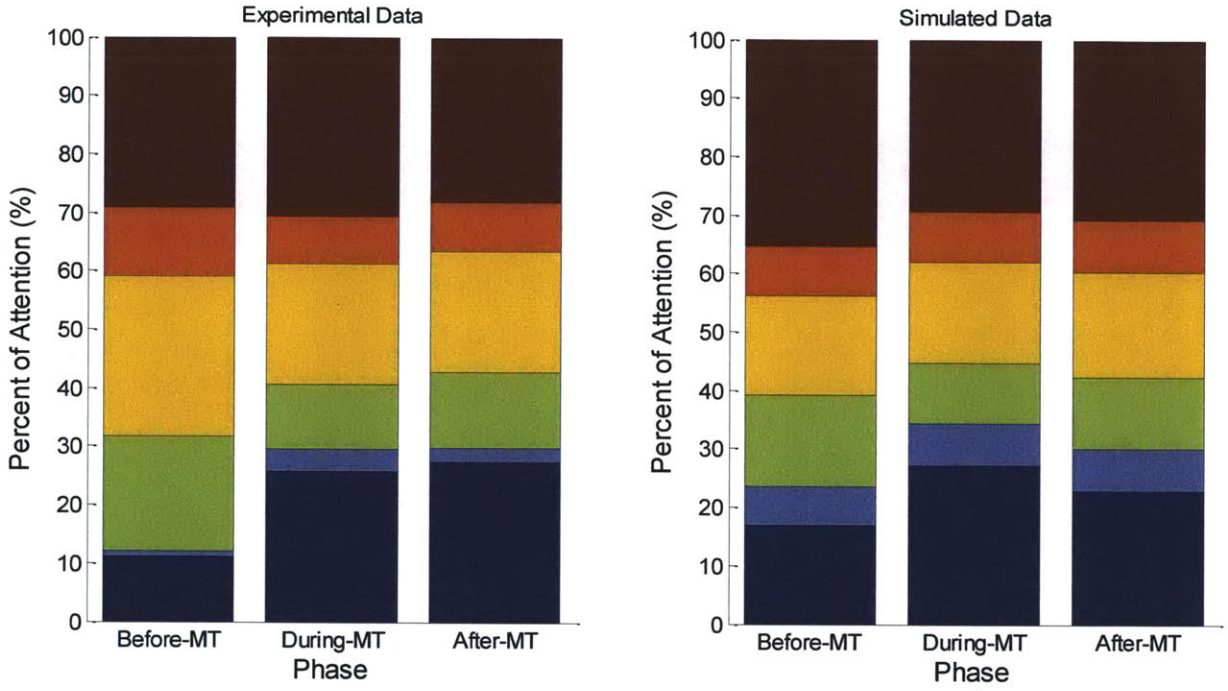
TA+RoD→TA, without LPR



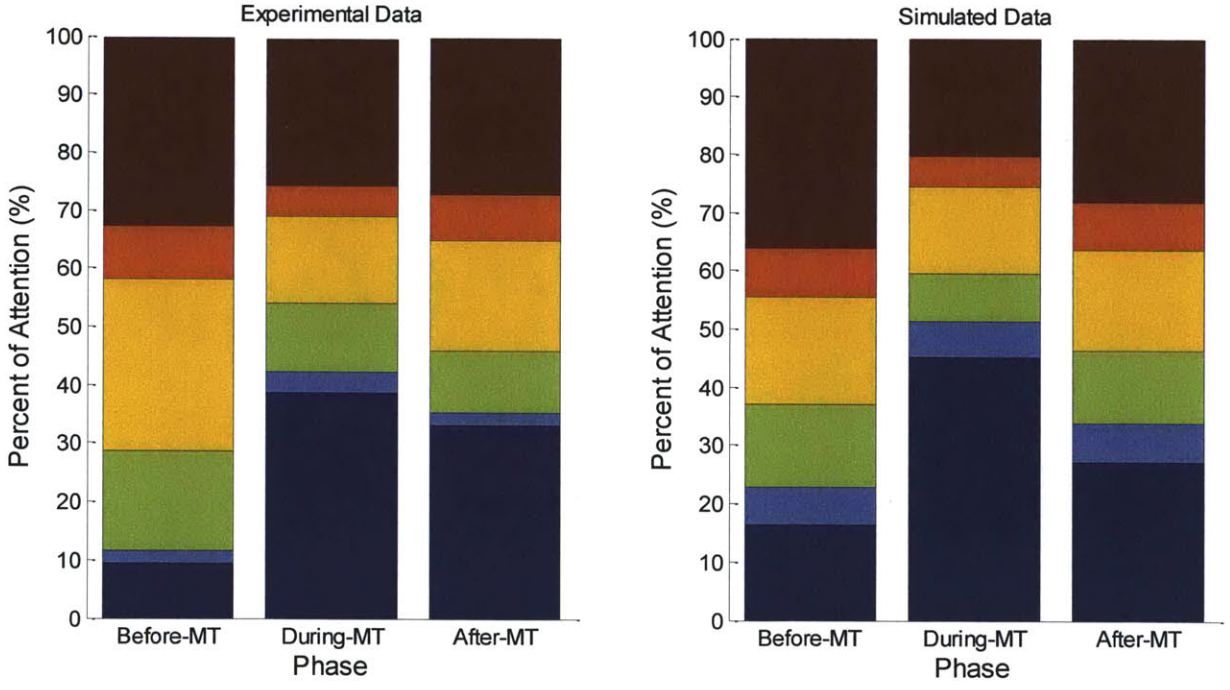
TA+RoD→TA, with LPR



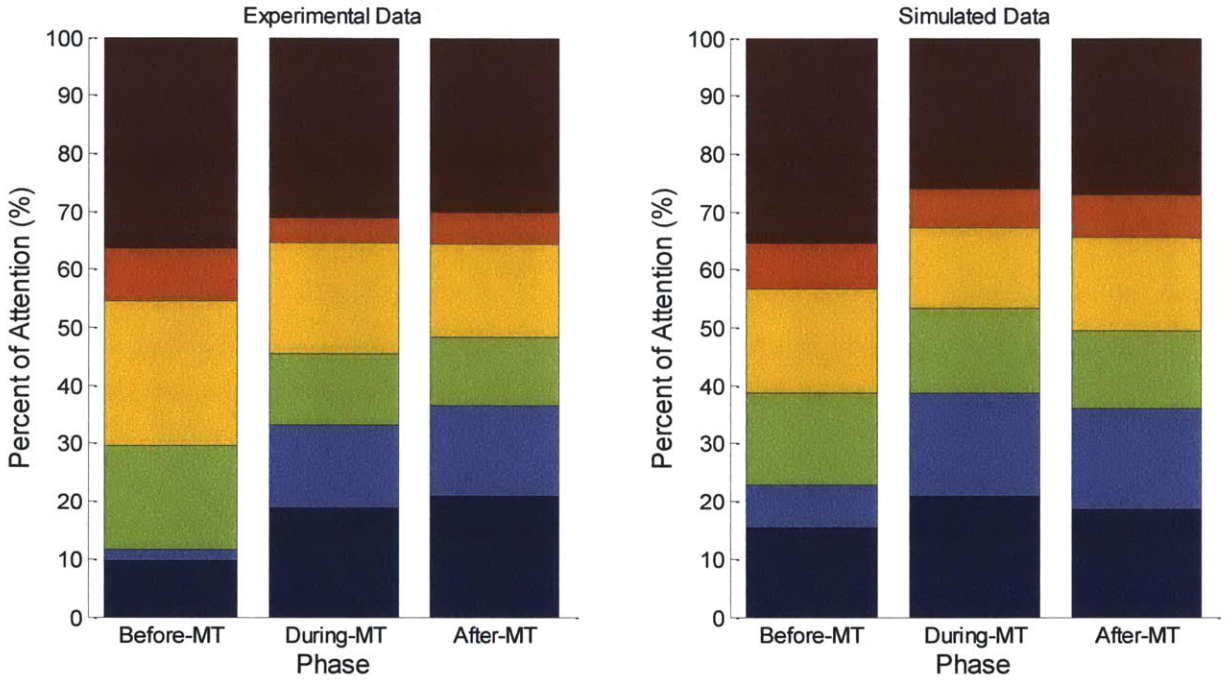
Auto→TA, without LPR



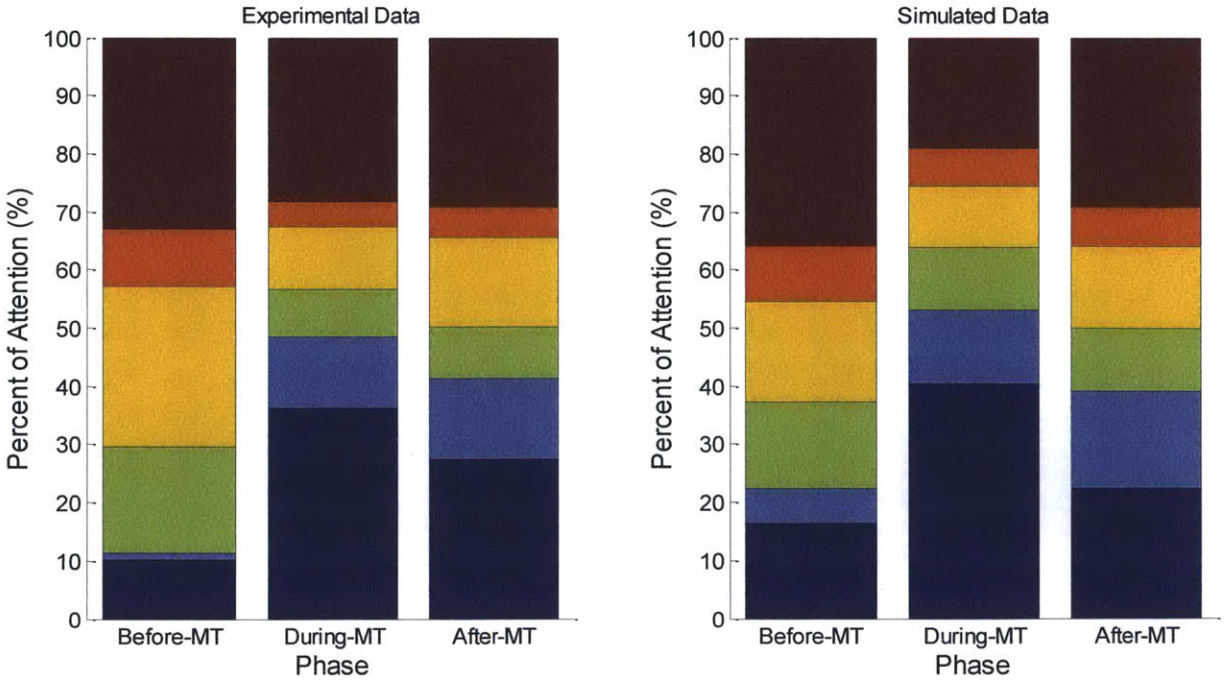
Auto→TA, with LPR



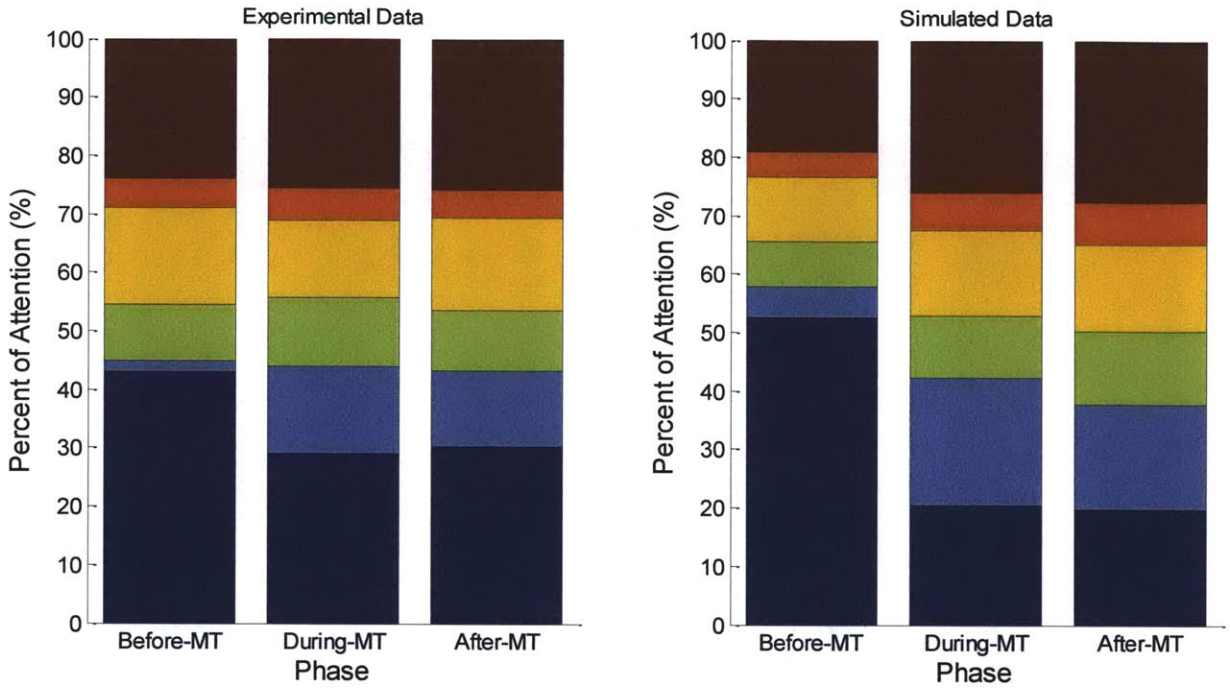
Auto→TA+RoD, without LPR



Auto→TA+RoD, with LPR



TA→TA+RoD, without LPR



TA→TA+RoD, with LPR