SUPPORTING MULTITASKING: EVALUATION OF NOVEL INPUT AND OUTPUT CHARACTERISTICS TO SUPPORT PRIMARY AND SECONDARY TASK PERFORMANCE USING SITUATION AWARENESS AND MENTAL WORKLOAD

A Dissertation

by

KATHRYN GENEVIEVE TIPPEY

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Thomas K. Ferris
Committee Members,	Mark A. Lawley
	Mark S. Avnet
	Steven M. Smith
Head of Department,	César Malavé

August 2016

Major Subject: Industrial Engineering

ABSTRACT

Driving and flying are examples of tasks in dynamic process control environments where performing secondary tasks poses risks to the user, potentially resulting in detrimental effects on situation awareness (SA) and mental workload (MWL). The predominant theories of multitasking emphasize the potential of using alternate information processing channels to better support multitasking, and hence prompt the exploration of novel interface characteristics that may reduce cognitive and structural interference. This dissertation uses measures of SA and MWL to evaluate the potential of input and output display characteristics found in novel technologies, such as Google Glass, Pebble Smartwatch, and Windows Surface, to support multitasking performance. Three experiments were performed to evaluate the following input and output characteristics: voice input (Experiments 1 and 3), head-up display (Experiment 1), size of display (Experiments 2 and 3), use of discrete tactile signals (Experiment 2), and use of continuously informing tactile signals (Experiment 3). Experiments 1 and 2 altered only interface characteristics associated with secondary task performance, while Experiment 1 also altered interface characteristics associated with primary task performance in the attempt to more effectively redistribute MWL.

While Experiments 1 (Texting and Driving with Google Glass) and 2 (Weather Technology Characteristics in General Aviation Cockpits) indicate the potential SA and MWL benefits of using voice input and larger displays for secondary tasks in multitasking settings, Experiment 3 (Supporting Emergency Vehicle Mobile Command Terminal Use While Driving) sheds light on the limitations of these benefits with increasing task complexity. Experiment 1 showed that combining a head-up display with voice input

provided additional marginal SA and MWL benefits. Experiment 2 also suggested SA and MWL benefits when using discrete tactile signals to aid in indicating the need for secondary task attentional shifts. Experiment 3 furthered this exploration of tactile signaling by presenting continuously informing vibrations relating to the primary task, exhibiting both the potential benefits of providing continuous information and the potential drawbacks of overreliance on such displays. These findings have the potential to fundamentally change the way users interact with technology by informing the development of and policies surrounding new products using these features.

DEDICATION

Dedicated in loving memory of my loyal, faithful, and supportive canine companions, Jeannie and Simba.

ACKNOWLEDGMENTS

Thanks to all of my family and friends for their support and encouragement through my graduate career, especially my mom Jacalyn Tippey, who has been one of my greatest sources of support. Special thanks to Andrew Tippey for his programming assistance. Special thanks also to Swaroop Dinakar and Elayaraj Sivaraj for their collaborative efforts and to my other colleagues in the Human Factors & Cognitive Systems Laboratory at Texas A&M.

Thank you to my committee for your patience and guidance. Thank you also to all Texas A&M students who participated in these studies and to the staff at the Federal Aviation Administration (FAA) William J. Hughes Technical Center (WJHTC) simulation facility.

This research was supported, in part, by the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) (Grant number: DGE-1252521), and in part through the Federal Aviation Administration (FAA) Weather Technology in the Cockpit (WTIC) project.

NOMENCLATURE

FAA	Federal Aviation Administration
ITS	Intelligent Transportation Systems
MCT	Mobile Computer Terminals
PED	Personal Electronic Device
MRT	Multiple Resource Theory
SA	Situation Awareness
MWL	Mental Workload
VFR	Visual Flight Rules
IFR	Instrument Flight Rules
VMC	Visual Meteorological Conditions
IMC	Instrument Meteorological Conditions
GA	General Aviation
SAP	Situation Awareness Probe
ADAP	Alert Decision-Action Point

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

Page

1. INTRODUCTION

The overarching goal of this dissertation is to address how to better support operators in multitasking environments through the use of novel interface design characteristics. The metrics used for this evaluation are centered around assessing situation awareness (SA) and mental workload (MWL). This research directly applies to the development of interfaces to increase both operational performance and safety in multitasking settings.

1.1 Motivation

The motivation for this dissertation is to enhance the safety of operators when they are engaged in multitasking activities. The primary tasks of driving and flying both present significant multitasking challenges as drivers and pilots must manage operational activities in environments that contains dynamic hazards. However, sometimes operators may find it necessary to engage in secondary tasks while driving or flying, such as for navigation or to aid in maintaining awareness of the weather. This has the potential to result in fatal consequences. In fact, in the United States, over 1,060 people are injured daily in crashes involving distracted drivers (National Highway Traffic Safety Administration (NHTSA), 2010; 2014). Accidents involving the use of portable electronic devices (PEDs) in airplanes were extensive enough to lead the National Transportation Safety Board (NTSB) to issue a safety alert warning pilots to turn off distracting devices before entering the cockpit (NTSB, 2013).

To safely operate an automobile or airplane, users must maintain a high level of situation awareness (SA). If competition exists for the use of the limited mental resources available to maintain this state of SA, this poses a safety risk. This competition often comes in the form of a secondary task. Mental workload (MWL) is a common construct used to evaluated the competition for these mental resources. For example, a driver's internal representation of the car along with their knowledge of objects in the surrounding environment make up the driver's SA. The driver's use of visuo-spatial and manual resources to keep their eyes on the road and hands on the steering wheel make up the driver's MWL. If the driver then engages in a secondary task such as texting, this results in competition for this same pool of limited mental resources because texting also typically requires the use of similar visuo-spatial and manual resources to read and respond to the text messages. Hence, evaluating both the level of SA and level of MWL in these multitasking settings allows researchers to more effectively determine how to best support operators by either reducing or redistributing MWL among multiple modalities and processing channels.

1.2 Contributions

This dissertation is structured in terms of primary and secondary tasks, with secondary tasks being operationally defined as "embedded" or "non-embedded". For instance, in the example noted in the *Motivation* section, the primary task is driving and the secondary task is texting. The secondary task is "non-embedded" because the information from the text message does not typically contribute to the SA level of the primary driving task (though this is dependent on the content of the text message).

Current research on how to best manipulate interface design to support operators focuses on methods that alter the secondary task in order to reduce interference with the primary task. Building on that research, this dissertation evaluates how to better support the secondary task through novel input and output interface characteristics. Additionally, this dissertation also evaluates the strategy of altering the interface for the primary task to reduce interference between the primary and secondary tasks.

1.3 Organization of this Dissertation

This dissertation begins with a literature review discussing the impact of display design on SA and MWL. Next, the proposed research questions are presented, which is followed by a discussion of the approach used to answer the research questions. After this, each of the three experiments in this dissertation is presented separately, including their hypotheses, methods, results, and discussion.

2. LITERATURE REVIEW

Situation awareness (SA) and mental workload (MWL) are both important constructs for supporting multitasking users. Relating back to the texting and driving example in the *Motivation* section, SA is the bigger picture construct that directly relates to the driver's knowledge of the environment and how well the driver is able to update information about the environment to maintain a clear assessment of hazards. Breaking this down, MWL is the smaller picture that aids in the evaluation of resource competition between the primary driving and secondary texting tasks, which in turn impacts SA.

This literature review first defines SA, followed by a discussion about MWL and its relationship with SA. Next, information about multitasking and how to support MWL and SA in multitasking are presented. This is followed by a discussion of the multitasking environments and their challenges (i.e., problem states) as studied in this dissertation. Next, prospective novel display elements (i.e., solution states (i.e., mediating measures)) that may be used to address these multitasking challenges are examined. Finally, potential metrics for evaluating the costs and benefits of these prospective display elements are discussed.

2.1 Situation Awareness (SA)

A general aviation cockpit contains a window giving the pilot and view of the horizon and a control panel with a myriad of instrumentation that the pilot must continuously assess. Building out of this environment, the FAA defines situation awareness (SA) as the, "continuous extraction of environmental information (from the out the window view and the instrument panel), integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events" (Bryne, 2015). To further clarify this definition for pilots, the FAA then states, "Simply put, situational awareness means knowing what is going on around you." (Bryne, 2015).

SA is used in the discussion of three major branches of research (Endsley, 1995): (1) Decision-making (i.e, action choice), (2) Mental models (i.e., internal representation of the system), and (3) Tasks and system factors (i.e., factors that influence the ability to achieve different levels of SA). This dissertation will focus on (1) and (3), with greater emphasis on (3).

The SA construct involves several overarching features that encompass all three branches of research. SA centers around cognition and working memory and does not focus on an individual's particular action or response. SA is considered a process and a state - the state being the product of SA and being separate from the process of maintaining SA. Additionally, while good SA may support good decision-making, it does not include the individual's final choice (Wickens, SA, 2002). This means the focus of SA is on dynamic, evolving situations, where the information gained from the SA process is used to construct mental models, creating links between the individual's goals and expectations (Wickens, SA, 2002; Endsley, 2015). SA also highlights the potential impact of task and system effects, such as workload and interface design, on model selection and decision processes and the need for attention to and proper integration of data into an individual's mental model (Wickens, SA, 2002; Endsley, 2015). SA hence involves analysis that centers on the impact of expertise and the potential misrepresentation of data. Good SA can therefore be characterized as necessary but not sufficient for good performance (Wickens, SA, 2008).

The most widely used SA construct “drills down” the state and process of maintaining situational awareness into three levels (Endsley, 1995; Jeon, Walker, and Gable, 2014) (Table 1). The diagnostic distinction between these levels of SA offers insight into the potential failures that may occur at each level and how to address those breakdowns (Wickens, SA, 2008). In outlining the essential factors for maintaining SA, for example, attention to direct and acquired information is vital for Level 1 SA while long-term memory is involved in all levels of SA (Wickens, SA, 2008). Hence, in terms of engineering applications, a breakdown in Level 1 SA indicates the need for better alerts while a breakdown in Level 3 SA indicates the need for the incorporation of predictive displays (Wickens, SA, 2008).

Table 1. *Overview of the levels of SA*

Level SA	**Definition**	**Sub-Components**	**Typical Cognitive Processes**
***Level 1* SA**	*Perception* of relevant information - involves filtering all incoming information from the outside world	Processes of monitoring, cue detection, and simple recognition	Attention and perception
***Level 2* SA**	Integrating the perceived relevant information with the operator’s goals (*comprehension*) – adds meaning and understanding to the information	Processes of pattern recognition, interpretation, and evaluation	Interpretation and judgment
***Level 3* SA**	Using information gained from (1) and (2) to predict future events in the system (*projection*)	Comprehension of the situation, with information being extrapolated forward in time to determine how it will affect future states of the operational environment	Judgment and decision-making

2.2 Mental Workload (MWL)

Workload is the “work” that is “loaded” on an operator and generally refers to all aspects of interaction between an operator and a structurally defined task (Huey, Messick, & Wickens, 1993). In terms of physical workload, the dimensions of workload are “stress” and “strain”, with “stress” being the demand imposed by the load and “strain” being the impact the load has on the user (Young, Brookhuis, Wickens, & Hancock, 2015). The analogous representation in terms of mental workload (MWL) is that “stress” represents task demands (e.g., time pressure and task complexity) and “strain” represents the impact on the human (e.g., mental expressions of the operator which often depend on the resources available) (Young et al., 2015).

Workload is, however, often only used in cases where the components required for successful completion of the task cause task demands that exceed the capacity of the worker (Tsang & Vidulich, 2006). Mental workload (MWL) is often referred to in terms of the resources demanded from a situation and typically reflects the level of attentional demand a task set places on the operator (Vidulich & Tsang, 2015; Young et al., 2015). These delineations have led to the development of assessment tools for determining the amount of residual attention leftover when performing a task (Parasuraman, Sheridan, & Wickens, 2008).

2.3 Relationship of Situation Awareness (SA) to Mental Workload (MWL)

In contrast to MWL, which focuses on the attentional resources demanded by the task, SA is predominantly associated with the information contained within the operator’s memory during task performance (Vidulich & Tsang, 2015). Wickens (2001) further contrasts these two concepts by deeming MWL as “fundamentally energetic” and having predominantly quantitative properties and by deeming SA as “fundamentally cognitive” and having predominantly qualitative properties. Despite these differences, however, both SA and MWL are shaped by a similar set of exogenous and endogenous factors. Exogenous factors reflect the operational demands of the task and constraints on the system while endogenous factors reflect the operator’s inherent skills and abilities (Vidulich & Tsang, 2015).

Figure 1 further explains the relationship between MWL and SA. The constructs are related through two major loops: (1) The attention and MWL loop, and (2) The SA loop (Tsang & Vidulich, 2006). The components of each of these loops reflect the previously discussed elements of MWL and SA. In the figure, MWL and SA are more directly connected through “strategic management” (i.e., executive control), which is used for maintaining adequate SA while coordinating multiple tasks and avoiding excessive workload (Tsang & Vidulich, 2006).

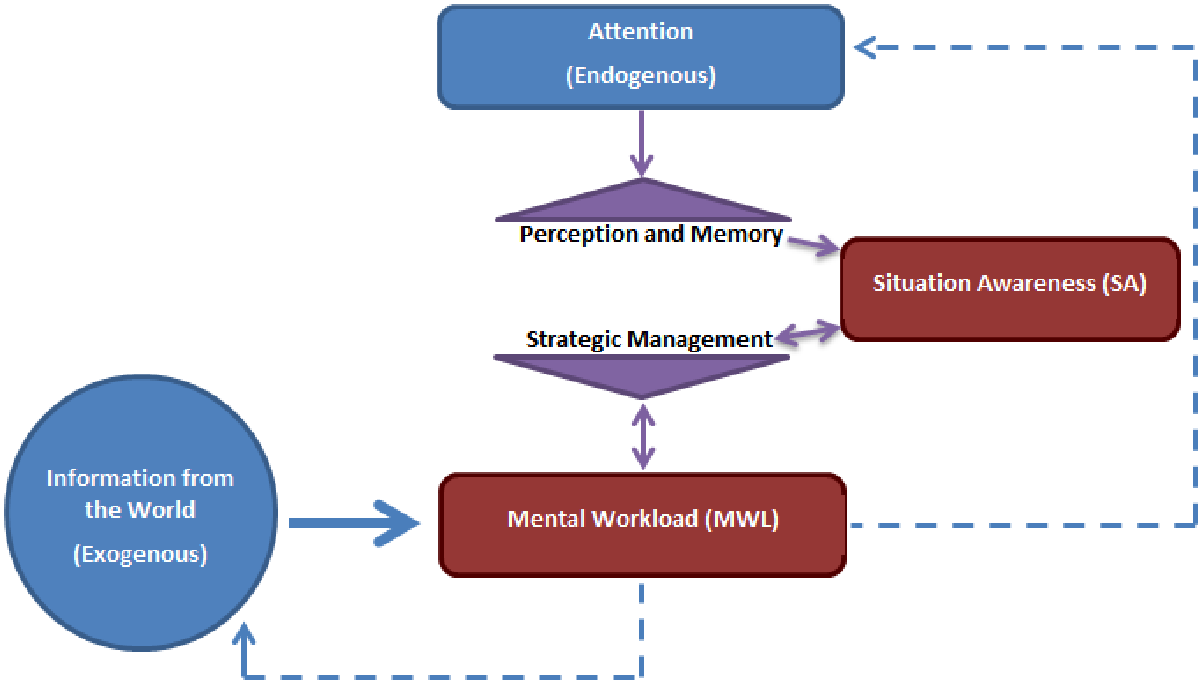

Figure 1. Illustration of the theoretical framework underlying the relationship between SA and MWL (adapted from Tsang & Vidulich, 2006). (1) The attention and MWL loop is represented by nearly the entire graphic and shows how attention and perception and memory relate MWL to SA. (2) The SA loop reflects the relationship of MWL and SA through strategic management. Both loops highlight how exogenous information is endogenously managed by the individual.

The complex connection between MWL and SA means that both positive and negative associations may result from this relationship (Table 2). Both workload and SA are competing for the same limited attentional resources during task performance (Tsang & Vidulich, 2006); however, higher workload levels, such as when the operator has an increased sampling rate, may contribute to higher levels of SA, particularly when the operator uses effective strategic management skills. In the optimal scenario, the operator would be able to efficiently obtain a high level of SA which would in turn promote a lower level of MWL (Tsang & Vidulich, 2006).

Table 2. *Potential cases of Workload and SA (Endsley, 1995)*

Mental Workload (MWL)	**Situation Awareness (SA)**	
	Low	*High*
Low	Information from the environment is either not present or not being used effectively (e.g., inattention, vigilance problems, or low motivation) -> The operator both has little idea of and is not actively working to determine the state of the situation	**Ideal State:** Information required to maintain SA is presented in an easily processable manner -> The operator has a good internal representation of the system and system state
High	Either too much information is in the environment, the information is not effectively displayed, or the number of tasks is too great -> The operator can only attend to a subset of the information available	A large amount of information is present in the environment -> The operator is working very hard to successfully achieve an accurate and complete picture of the situation

2.4 Multitasking

Mental workload (MWL) and SA compete for the same pool of limited attentional resources within a single task. The concept of multitasking involves performing several tasks concurrently (Wickens, MRT, 2008); the types of tasks being performed concurrently impact the operator's time-sharing abilities (Wickens, MRT, 2008). In this dissertation, one task is defined as the primary, ongoing task, and the other task is defined as the secondary task. The secondary task may be discrete or continuous but must be performed in addition to the primary task. Each secondary task will involve adding a technological device to the environment.

For purposes of this dissertation, multitasking is divided into what will be operationally referred to as “embedded” versus “non-embedded” multitasking. “Embedded” multitasking is the performance of a secondary task that contributes to the SA of the primary task. “Non-embedded” multitasking is the performance of a secondary task that is completely independent of the primary task.

The primary theoretical underpinnings for managing multitasking are reflected in Multiple Resource Theory (MRT) (Wickens 1980; 2002) and interference theory (Ivry, Diesrichson, Spencer, Hazeltine, & Semjem, 2004), both of which involve the distribution of a limited set of physical and mental resources within a task set. These are not distinct constructs as interference is also an essential component of MRT.

“Physical resources” refer to the resources required to physically perform the task. “Mental resources” refer to the perceptual, cognitive, and response resources required to mentally perform the task (Wickens, 1980; 2002). In terms of MRT, “mental resources” are divided into dimensions (Figure 2), which characterize the influences of resource demand, resource structure, and resource allocation strategies between concurrently performed tasks. MRT also predicts the level of interference between or among tasks and which task will likely suffer due to competition for resources (Grier, Wickens, Kaber, Strayer, Boehm-Davis, Trafton, & John, 2008).

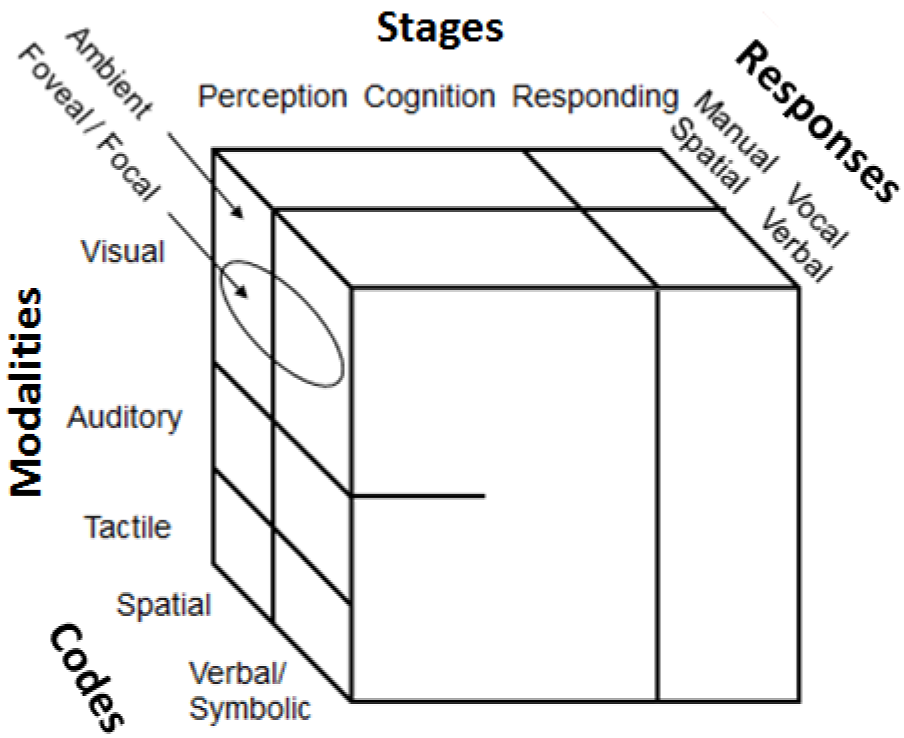

Figure 2. Representation of how resources may potentially be distributed in Multiple Resource Theory (adapted from Wickens, MRT, 2002). Note that, for example, while one modality may be overloaded, resources may still be available within another modality. Additionally, the resource "blocks" (created from codes, modalities, stages, and responses) are not independent.

In multitasking situations performance decrements between primary and secondary task activities may result from either structural (e.g., eyes can only support one field of view at a time, individual appendages can only perform one motor activity at a time) or cognitive interferences (e.g., among engaged working memory resources) (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Hurts et al., 2011; Sawyer, Finomore, Calvo, & Hancock, 2014). Performance decrements may also result from exceeding the capacity for physical or mental resources (i.e., the "cognitive redline", e.g., Grier et al., 2008; Rodriguez, Yang, Tippey, & Ferris, 2015). Moreover, if the multitasking load becomes high enough, the operator will switch to a strategy where they perform the tasks sequentially, with the secondary discrete task then acting as an interruption (Grier et al., 2008; Wickens, Hollands, Parasuraman, & Banbury, 2012).

The dimensions of MRT also have neurophysiological plausibility (i.e., neurophysiological research suggests that resource allocation in the brain may actually function in this manner) (Tsang & Vidulich, 2006). A growing body of research explores the use of Functional Near Infrared Spectroscopy (fNIRS) for discriminating the level of MWL imposed by a task (e.g., Power, Kushki, & Chau, 2011; Ciftci, Sankur, Kahya, & Akin 2008; Hoshi et al., 2002; Hirshfield et al., 2009) and for determining the relationship between workload and different levels of vigilance (e.g., Bogler, Mehnert, Steinberk, & Haynes, 2014; Helton, Warm, Tripp, Matthews, Parasuramen, & Hancock, 2010; Brunce, Izzetoglu, Ayaz, Shewokis, Izzetoglu, Pourrezaei, & Onaral, 2011). Several studies also evaluate the relationship between subjective, physiological, and performance metrics (e.g., Gupta, Laghari, Arndt, Schleicher, Moller, & O'Shaughnessy, 2013; Hirshfield, Girouard, Solovey, Jacob, Sassaroli, Tong, & Fantini, 2007; Peck, Yuksel, Ottley, Jacob, & Chang, 2013) and suggest that physiological data, such as that obtained using fNIRS, may explain apparent differences between subjective and performance outcomes.

The three primary advantages of observing multitasking in the construct of MRT for this dissertation are

(1) The model incorporates and aligns with interference theory (and calculations of interference, which is beneficial in understanding how and what or which resources are competing with each other at any given time (Wickens, MRT, 2008);
(2) The model addresses potential single-channel bottlenecking (Pashler, 1998); and
(3) The four dimensions of the model coincide with potential design decisions that may be addressed in the applied engineering context to better support multitasking activities (Wickens, MRT, 2008).

MRT is limited in that it does not account for phenomena that may distort attention (i.e., unwanted operator attention to interruptions, cognitive tunneling, and auditory preemption) and phenomena that may result in perceptual abnormalities (e.g., the moon illusion, perceptual masking).

2.5 Supporting Mental Workload (MWL) and Situation Awareness (SA) in Multitasking Using Interface Design

"Interface knowledge" significantly influences the ability to achieve a sufficient state of SA and may be manipulated within the context of MRT to better support the user (Figure 3). Interface design determines the amount of information that can be acquired, how accurately that information can be acquired, and to what degree that information is compatible with operator's SA needs (Endsley, 1995). In designing an interface to better support SA, developers must evaluate the amount of information processed at each SA level and determine how that information contributes to the operator's goals, must take into account the potential for attentional distortions and phenomena, and must support the operator in projecting future states of the system (Endsley, 1995; Tsang & Vidulich, 2006).

Developers should account for potential information overloading issues and determine how the system may support attentional-sharing between mutiple tasks (Endsley, 1995), both of which can be accounted for in research and design through the use of MRT.

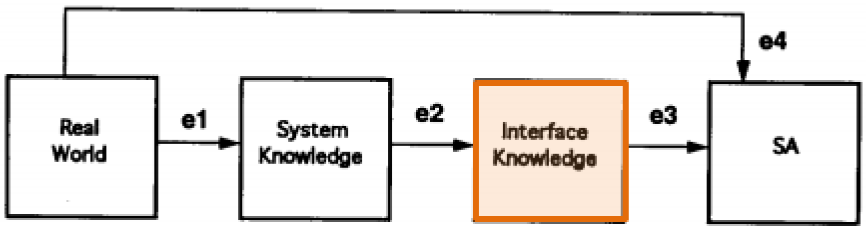

Figure 3. Task and system factor inputs to SA (adapted from Endsley, 1995)

Alerts predominantly contribute to Level 1 SA. Alerts can be particularly useful tasks of change detection (Nikolic, Orr, & Sarter, 2004). Grier, Wickens, Kaber, Strayer, Boehm-Davis, Trafton, & John (2008), in their discussion of the red-line of cognitive workload, make several recommendations about how interface design can support SA through aiding in change detection:

- The interface should automatically detect and notify operators in a relatively unobtrusive manner in order to minimize the amount of distraction from the display and to not overtax working memory (Tsang & Vidulich, 2006);
- If the display is cluttered, then the change information should only be available on-demand by the operator; and

- The interface should provide a summary of each significant change that occurs to allow the user to scan and prioritize the order in which changes are reviewed.

2.6 Multitasking Environments Studied

This dissertation involved performing experiments to evaluate novel interface characteristics for multitasking using task sets constructed in two real-world multitasking environments: driving and flying. The acts of driving and flying are dynamic, continuous, on-going process control tasks that require the operator to have a high level of SA to maintain safety. Hence, performing secondary tasks in these environments that use the same limited mental resources, such as the driving and texting case mentioned previously, is quite dangerous.

Driving and flying are composed of three major categories of tasks: strategic tasks, tactical tasks, and operational/control tasks (Matthews, Bryant, Webb, & Harbluck, 2001). Strategic tasks require both long-term planning, such as planning a route, and setting immediate goals, such as the execution of navigation plans and monitoring the environment for proximity cues (Matthews et al., 2001). Tactical tasks involve primarily short-term objectives; for example, local maneuvering of a vehicle or plane through a traffic stream or airspace, respectively (Matthews et al., 2001). Operational/Control tasks primarily involve the execution of routine actions, such as controlling the vehicle or airplane by doing things like steering and braking.

Each task category requires different SA levels (further represented in Figure 4). Strategic tasks chiefly involve planning and hence require a large amount of prediction, requiring increased Level 3 SA. However, strategic tasks also involve carrying out immediate goals, requiring increased Level 2 SA, and monitoring the environment for salient cues, requiring increased Level 1 SA. As tactical tasks possess less need to project events, these tasks primarily involve Levels 1 and 2 SA (Matthews et al., 2001). Routine operational/control tasks are composed of mostly automatic processes that are periodically monitored; hence, the tasks predominantly require Level 1 SA to ensure the automatic processes are behaving appropriately and will only involve higher levels of SA (primarily Level 2 SA) if an error occurs in a lower process (Matthews et al., 2001). Figure 4 represents the approximate amount of each Level of SA required to perform each category of task, and Table 3 contains additional examples of tasks in each category and the levels of SA used in those tasks.

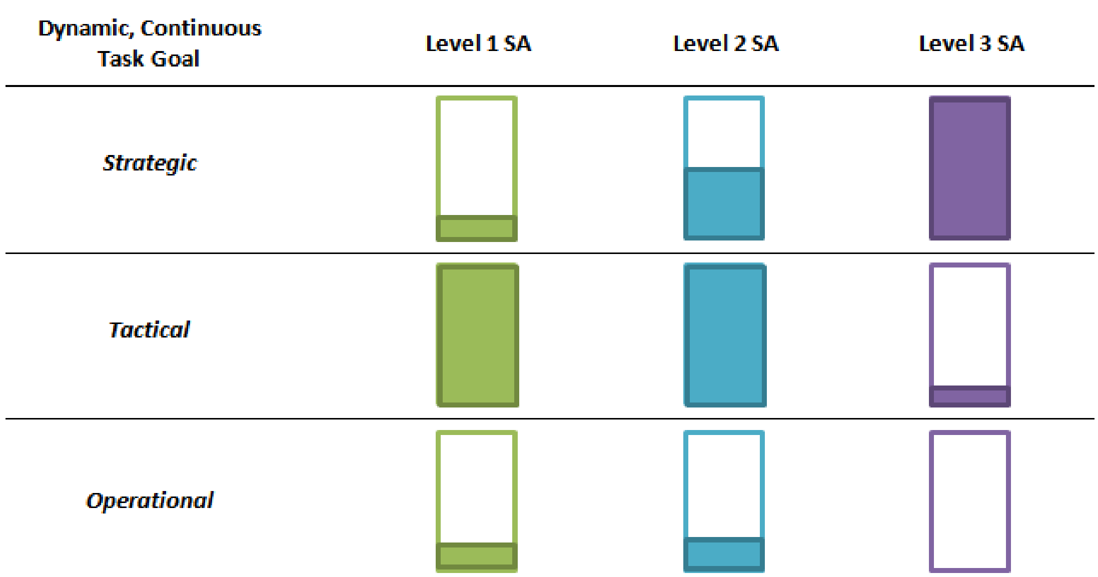

Figure 4. Representation of the relationship between the different driving and flying task categories and the different levels of SA (adapted from Matthews et al., 2001)

Table 3. *Examples of the different levels of SA for each task category within the driving and flying domains (Jeon, Walker, & Gable, 2014; Matthews et al., 2001; Endsley, 2015)*

			Domain	
Task Level	Task Example	***SA Level***	**Driving**	**Flying**
Strategic	*Route planning*	***1***	Perception of relevant landmarks	Perception of terrain location and height
		2	Comprehension of status in rout and updating of route plan	Comprehension of status and ability to reach destination
		3	Anticipation (projection) of delays and alternative routes	Projection of deviations in schedule
Tactical	*Hazard avoidance*	***1***	Perception of traffic in the environment	Perception of weather and altitudes effected
		2	Evaluation (comprehension) of safety margin	Comprehension of the validity of the indications
		3	Anticipation (projection) of hazard and prediction of future path	Projection of areas of severe weather may encounter
Control/ Operational	*Wind gust*	***1***	Detection (perception) of gust	Perception of wind gust
		2	Evaluation (comprehension) of relevance as a hazard	Evaluation (comprehension) of relevance as a hazard
		3	Projection of vehicle displacement due to the gust	Projection of airplane displacement due to the gust

Within the Levels of SA, both driving and flying require three different types of SA: spatial awareness, system awareness, and task awareness (Matthews et al., 2001; Wickens, SA, 2002). Spatial awareness involves maintaining knowledge of the locations of important features in the environment; this includes keeping track of both exterior (e.g., out-the-window situation) and interior (e.g., instrument displays) information (Matthews et al., 2001; Wickens, SA, 2002). System awareness involves knowing relevant system information and its relationship to the environment, which includes keeping the operator informed of actions taken by automated systems (Matthews et al., 2001; Wickens, SA, 2002). Task awareness directly involves how the operator maintains knowledge of where they are within the goals and sub-goals of the overall task, with task management contributing to task awareness (Matthews et al., 2001; Wickens, SA, 2002). System and task awareness also involve the temporal management of information; however, spatial awareness only involves the temporal information management to the extent that the operator is predicting the next locations and potential important features in the environment (Matthews et al., 2001; Wickens, SA, 2002). Additionally, salient items within any of these three contexts may either enhance the operator's knowledge of the situation or distract the operator from more valuable changes that must be monitored within the environment (Matthews et al., 2001; Wickens, SA, 2002).

2.6.1 Primary Task Environments

The following two sections explain the mental resources required when performing just the task of driving in the driving environment and just the task of flying in the flying environment.

2.6.1.1 Driving

Prior research indicates the dangers of performing both embedded and non-embedded secondary tasks while driving (e.g., Drews et al., 2009; Filtness et al., 2013; Tsimhoni & Green, 2001; Lyngsie, Pedersen, Stage, & Vestergaard, 2013; Yager, 2013). For example, take the representative task of driving and texting, which is predominantly a non-embedded task. The theoretical structures of MRT and interference provide insights into why the secondary task of texting on a mobile device– an activity that requires visual, spatial, and manual resources – is so problematic to attempt while driving (e.g., Drews et al., 2009; Fitch et al., 2013; Horrey & Wickens, 2007) (Figure 5). The detrimental effects of texting-and-driving on multitask performance and driving safety have been demonstrated in both controlled experimental contexts (e.g., Drews et al., 2009; Fitch, Hanowski, & Guo, 2015; Horrey and Wickens, 2007; Lyngsie, Pedersen, Stage, & Vestergaard, 2013; Tsimhoni & Green, 2001; Yager, 2013) and in naturalistic studies, which have shown texting to be the secondary activity associated with the largest increase in crash risk (23-fold) compared to non-distracted driving (e.g., Fitch et al., 2013; Olson, Hanowski, Hickman, & Bocanegra, 2009; Horrey & Wickens, 2007).

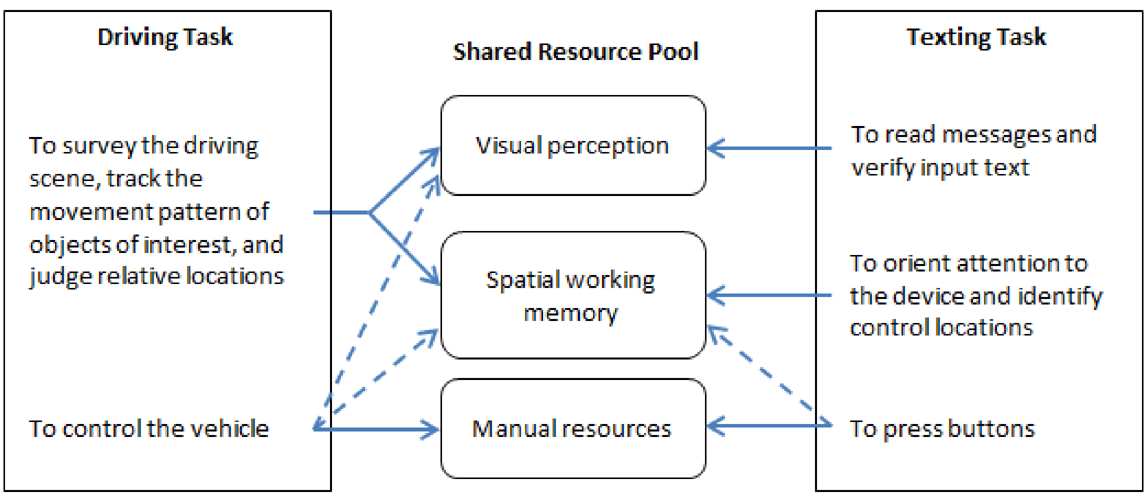

Figure 5. Representation of how the tasks of texting and driving share similar limited perceptual, cognitive, and manual resources. The dashed arrows represent that manual activities require the engagement of spatial working memory and that vehicle control additionally requires the engagement of visual perception, although the automatic nature of these motor activities suggests relatively little demand is imposed on memory or visual resources.

This same logic can be expanded to the use of Mobile Command Terminals (MCTs) (a.k.a., Mobile Data Terminals (MDTs)) by police officers who work alone in their vehicles and are hence responsible for both driving and operation their MCT (Yager, Dinakar, Sanagaram, & Ferris, 2015). However, while texting-and-driving can wait until the driver encounters a lower workload setting, MCT use may be more urgent and is by definition a required task for the successful completion of job duties. Emergency personnel spend approximately 13 percent of their shift time each day interacting with an MCT (Girouard, Rae, Croll, Callaghan, McKinnon, & Albert, 2013).

According to a usability analysis using the "Safety Checklist for the Assessment of In-Vehicle Information Systems", most current MCT systems are incompatible with the task of driving a vehicle (Yager et al., 2015). As with texting, MCTs require both manual and visual resources, including the reorientation of attention away from the roadway, and increase MWL, potentially causing officers to approach the red-line of cognitive workload (Grier et al., 2008; Wickens, 2002). Table 4 (Yeager et al., 2015) provides a breakdown of the potentially distracting tasks that are commonly performed on an MCT.

Table 4. *Summary table highlighting potentially distracting tasks performed by emergency personnel while operating their vehicles (copied from Yeager et al., 2015*)*

🔴High Demand; 🟡Moderate Demand; 🔵 Low Demand.

Task	Problems	Visual	Spatial	Manual	Solutions	References
Reading Call Notes	• Information in the form of textual phrases • Need for search through timeline • Increased glance time • Look for any updates on the current scenario	🔴	🟡	🔵	• Use of coded language as compared to naturalistic language • Use of audio-voice interface to interact with MCT	• Garrison, Williams, & Carruth, 2012 • Filtness, Mitsopoulos-Rubens, & Lenné, 2013
Using Map Function	• No turn-by-turn voice guided navigation system on MDT • Basic maps may not show current relative position with respect to destination • Map may not show traffic movement direction	🔴	🔴	🔵	• Use of voice based, turn-by-turn navigation system • Information on landmarks, current relative position displayed	• Srinivasan and Jovanis, 1997
Touch and Mouse-pad based interaction	• Requires continuous visual attention to identify current cursor location • Touch screen makes it difficult to access smaller icons/buttons • Requires arm to be rested on a support surface for stability	🔴	🟡	🟡	• Use of audio-voice interface to interact with MCT • Use of larger buttons to improve ease of use through touch-screen (40x40 pixels)	• Filtness, Mitsopoulos-Rubens, & Lenné, 2013 • Sun, Plocher, & Qu, 2007
License plate search (patrol vehicle)	• Involves reading a license plate number and entering it on the MCT • 7 character license plate number to be entered on a keyboard • Reading textual data about vehicle status and background, insurance information, vehicle owner background	🔴	🔴	🟡	• Use of coded language as compared to naturalistic language • Use of audio-voice interface to interact with MCT	• Garrison, Williams, & Carruth, 2012 • Filtness, Mitsopoulos-Rubens, & Lenné, 2013
Hospital Reporting (EMS vehicle)	• Locating closest hospital notifying them on the MDT • Input information on patient status and reason of emergency • Check for updates from hospital or dispatch	🟡	🔵	🔵	• Operation to be performed over the radio to dispatch	
Placement and location of in-vehicle devices	• Multiple equipment can obstruct the view of the driver by creating blind-spots • Require movement of the head away from the road and need to perform uncomfortable manual actions to interact with devices • Prolonged working in non-ergonomic environment can lead to musculoskeletal fatigue and injuries	🔴	🟡	🟡	• More ergonomically designed driver cockpit • Better location of in-vehicle devices	

***** Reprinted with permission from “Emergency Vehicle Operator On-Board Device Distractions” by Yager, C., Dinakar, S., Sanagram, M., & Ferris, T. K., 2015. Prepared by Texas A&M Transportation Institute for the Intelligent Transportation Society of America.

2.6.1.2 General Aviation (GA) Flying

General aviation (GA) flying builds on the potential resource competition found in driving. Flying, however, requires an additional tracking dimension (up-down) and monitoring of complex instrument displays found in general aviation flight panels. These displays indicate the plane's status and position and are strewn with a plethora of auditory alerts (Valasek, Ferris, Brown, Rantz, & Whitehurst, 2015) (Figure 6).

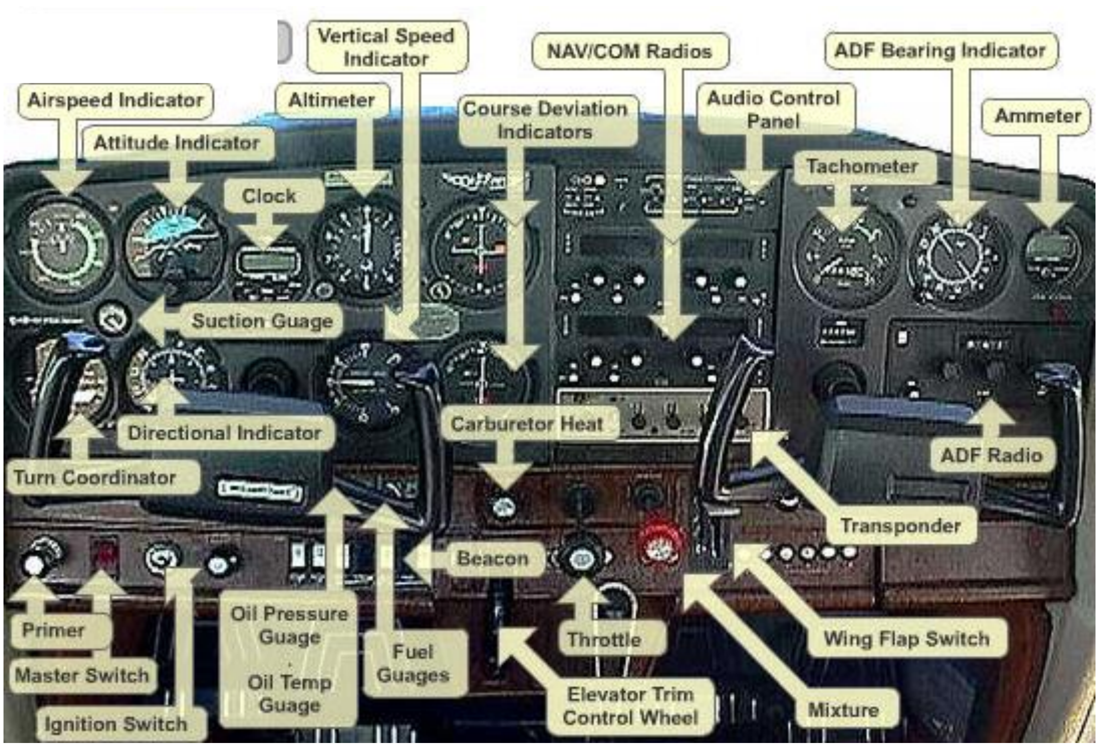

Figure 6. The instrument panel in a general aviation cockpit (adapted from Robert, 2008)

In flight six variables must be monitored simultaneously: three orientation variables (pitch, roll, and yaw) and three position variables (position on the flight path, altitude, and lateral deviation (from the flight path)) (Wickens, SA, 2002). These six variables are not independent: for example, pitch determines the plane’s future altitude and roll determines the plane’s future heading, lateral deviation, and pitch (Wickens, SA, 2002). Hence, the pilot must maintain an awareness of these six variables along with a temporal awareness of changes that occur as the pilot continues along the flight path, creating cognitive challenges relating to how the pilot internally represents tracking and lags. Additionally, pilots must coordinate potentially conflicting goals, such as maintaining the proper orientation to preserve lift (Wickens, SA, 2002). Maintaining system awareness then results in the potential for problems with

- Mode awareness - When automated systems on the flight panel change mode without the pilot realizing it;
- Change blindness - When a change in the environment occurs without the operator noticing it (this is more common when additional events or distractions occur at the same time as the environmental change); and
- Inattentional blindness - When an operator fails to recognize objects in plain sight (Haines, 1991).

While research suggests the optimal sound characteristics for auditory alerts during flight, cockpits are notorious for overloading pilots’ auditory channel (Valasek et al., 2015; Wiener & Nagel, 1988).

As with driving, pilots must manage flight activities in an environment that contains dynamic hazards. Weather is one of the most prominent examples of an embedded secondary task that poses significant dangers to general aviation pilots, with 50 weather-related general aviation accidents occurring in 2012 (38 fatal) (Nall Report, 2015). The majority of accidents result from pilots that are only certified to fly using Visual Flight Rules (VFR) in Visual Meteorological Conditions (VMC) unknowingly flying into Instrument Meteorological Conditions (IMC), with 80 percent of those accidents attributed to penetrating a thunderstorm or a deficiency in instrument techniques during an Instrument Flight Rules (IFR) flight (Nall Report, 2015). In fact, two-thirds of all accidents occurred in IMC, and two-thirds of accidents were made by private pilots (Nall Report, 2015). Maintaining a strictly visual awareness of weather is made even more difficult in the three-dimensional setting of flying due to phenomena such as the moon illusion, which describes how pilot's perception of their distance from the storm may be altered because of the absence of depth cues in the sky, making a storm that is relatively near appear further away (National Aeronautics and Space Administration (NASA), 2008). Pilots may then be caught off guard by the rapid deterioration in weather conditions, causing a strictly VFR pilot to end up in IMC.

While providing pilots with additional technology to aid in weather awareness is essential to mitigating these dangers, interacting with these new technologies themselves poses alternate hazards to pilot while flying. The following are examples of instances in which pilots were engaged in secondary tasks involving portable electronic devices (PEDs) that resulted in loss of life likely due to structural or cognitive interferences (NTSB, 2013):

- In August 2011, a helicopter pilot impacted the terrain due to engine failure because they did not confirm the helicopter had sufficient fuel prior to taking off. The accident investigation determined the pilot had engaged in frequent texting while preparing the helicopter for takeoff as well as during flight.

- In December 2007, a small plane impacted the terrain while doing a low-altitude fly-by of a friend's residence. The accident investigation determined that the pilot, who had been speaking with the friend on their cell phone during the fly-by, hit a turbulent wind and initiated a rapid climb. The airplane then stalled, resulting in the pilot's loss of control of the plane.
- In February 2006, a small plane hit a power line near the ground. The accident investigation determined that the pilot was speaking on their cell phone with a friend in a nearby tractor trailer that was driving the same stretch of highway as they were flying when the accident occurred.

Incidents such as these resulted in the NTSB issuing a safety alert urging pilots to recognize the potential distractions from the nonoperational use of PEDs while flying and to hence turn off PEDs before getting in the cockpit (NTSB, 2013). These incidents also accent the need for research into the safety issues of multitasking while flying and how to better support pilots when they encounter scenarios in which multitasking with technological devices brought into the cockpit is necessary, such as when approaching hazardous weather situations.

2.7 Supporting Multitasking in These Primary Task Environments

Altering interfaces in the driving and flying environments may result in two fundamental changes in the driving or flying task: (1) Impact on SA due to task automation (i.e., changes in information gained through and maintenance of a mental representation of the system and system state), and (2) Impact on behavioral adaption due to changes in perceived safety (i.e., if operators view the system as increasing in safety, then they may be more likely to rely on the system and reduce the amount of information they are processing from the environment) (Ward, 2000). Hence, while adding Intelligent Transportation Systems (ITSs) to the driving or flying environment may enhance the operator's experience and increase operator safety, such systems also have the potential to overload the operator and may additionally pose problems if they become the primary task (Matthews et al., 2001).

The focus of this dissertation is on the category of ITSs that aid the operator by supporting Level 1 SA. Such systems may expand the operator's sensory base, provide the operator with new kinds of sensory information, or expand the capabilities of current technologies that act as sensory aids. While this may involve using different types of signals that enhance the operator's perception or attention by engaging alternate modalities and processing channels than typical driving or flying alerts, developers must still be careful that these novel displays do not impair the operator's SA (Matthews et al., 2001). Theoretical inferences from multitasking literature help form the base for the development of potential design candidates that may better support SA and MWL in multitasking environments.

2.7.1 Supporting the Primary Task: Reducing Conflict with the Secondary Task

First consider how the secondary task may be supported to minimize interference with the primary task (i.e., driving or flying). This may result in interface changes affecting both the commands the operator inputs into the device and the output from the secondary task device to the user.

2.7.1.1 Input Characteristics

Many complex PEDs have replaced physical keyboards with touchscreens. This input change makes interacting with PEDS potentially even more detrimental to driving or flying performance (Lyngsie et al., 2013), because confirming touchscreen button activation puts a higher demand on visual resources. This is because touchscreen keyboards do not produce the haptic feedback found when using physical keys. An alternative to manual input is to employ voice-to-text input, which frees the user's hands thus reducing the user's manual resource demand and the resulting in lower levels of structural interference. Some evidence suggests voice-to-text input may be less detrimental to driving performance (He et al., 2014), promoting more eyes-on-road time and reducing subjective MWL compared to manual input methods (Tsimhoni & Green, 2001). However, voice-to-text methods still require eyes-off-road time to read incoming messages and visual attention to verify the correctness of text translations of the spoken input (Yager, 2013). Additionally, while MWL may be reduced with verbal entry compared to manual entry, the load imposed by verbal annunciation is similar to that required for phone conversations while driving and can result in similar performance decrements (Filtness, Mitsopoulos-Rubens, & Lenne, 2013). Moreover, device input and output characteristics can affect the level of demand imposed on perceptual (e.g., vision, audition), cognitive (e.g., spatial and verbal working memory), and response (e.g., hands, voice) resources in complex ways, thus the effects of device characteristics on operator safety and performance must be considered as an emergent property that depends on the individual and interacting demands these characteristics impose on human information processing resources.

2.7.1.2 Output Display Characteristics

Reading or manually interacting with a PED requires reorienting focal visual attention away from the horizon towards the device, with both the magnitude of this reorientation and duration of eyes-off-horizon time impacting safety (Horrey, Wickens, & Consalus, 2006; Hosking, Young, & Regan, 2009; Wittmann, Kiss, Gugg, Steffen, Finka, Poppela, & Kamiya, 2006). Since smartphones and other PED technologies are often positioned outside the field-of-view used for driving or flying, they require more disruptive reorientations than displays that at least partially share the same field-of-view as the horizon. This suggests that performance and safety may be improved with technologies that allow secondary tasks to be conducted within the same field-of-view as the horizon, thus making visual resources more "sharable" with the concurrent driving or flying task.

Screen size is the main dimension within traditional devices that may impact that "sharability" of visual resources. Comparatively larger screens (e.g., a standard smartphone versus a standard tablet PC) may require shorter glances away from the horizon to view the screen. Technologies of varying sizes are being brought into both automobiles and airplanes. For example, drivers began bringing Global Position System (GPS) devices into vehicles over 20 years ago. More recently, in aviation Garmin recently developed a smartwatch designed for GA pilots to wear in the cockpit, which includes distance from waypoints and airports as well as bearings and glide ratio (Garmin, 2013). Tablets, such as the iPad and Windows Surface, are also becoming more commonly used in the cockpit through apps such as ForeFlight, which, among other functions, aids pilots in flight planning and provides them with weather information (Joslin, 2013; ForeFlight, 2007). Studies also suggests that how users interact with cursors and button size may contribute to the amount of attentional resources required to complete a task, with research suggesting that performance is better with pointing over dragging tasks (MacKenzie, Sellen, & Buxton, 1991), which are best performed using a stylus over a mouse or trackball, and that key sizes should be no smaller than 20 mm (Colle & Hiszem, 2004).

Recent years have seen a growth in development of advanced technologies that offer promise for better supporting device interactions in multitasking environments. Of note, Bluetooth-connected smartphone extensions such as Google Glass, Sony's SmartEyeglass, and Samsung's Galaxy Glass are becoming more commonplace, and the usage of these devices while driving or flying is an increasingly urgent issue for policymakers, transportation engineers, and hardware and software designers. These devices combine head-up display (HUD) functionality – which provides output within the same field-of-view as the horizon, improving the operator's ability to perceive events in the forward scene (e.g., Kiefer, 1991; Kiefer & Gellatly, 1996; Flannagan & Harrison; 1994; Okabayashi, Sakata, Furukawa, & Hatada, 1990; Sojourner & Antin, 1990) – with alternative input methods (e.g., voice input) that together can reduce visual and manual interference.

Recent research conducted with Google Glass illustrates its potential to support some in-vehicle and cockpit tasks along with concurrent driving or flying performance, respectively. He, Ellis, Choi, & Wang (2015) compared reading performance on a smartphone versus Glass and found that while medium and long text messages both impaired driving performing, using Glass resulted in smaller driving performance decrements than using a smartphone. Building on this study, He, Choi, McCarley, & Chapparo (2015) sought to compare vocal text entry with both a smartphone and Glass using a short answer texting task on a simulated three-lane freeway. They found that while all texting conditions negatively impacted driving performance, Glass did so the least, including when compared with using smartphone voice-to-text. Similarly, Sawyer et al. (2014) compared driving and secondary arithmetic task performances with either Glass voice-to-text entry or smartphone manual entry and found that Glass better supported performance for recovery from brake events. Finally, Beckers et al. (2014) found similar results with an in-vehicle secondary task that is arguably more beneficial than texting: GPS destination entry. The study showed how using the voice input functionality with either Glass or a smartphone to enter an address resulted in significantly smaller driving performance decrements than did manual input. Burke (2015) addressed how Glass may help to better support pilots as they approach for landing by presenting information from approach plates on the device, either in conjunction with a tablet or when using Glass as the sole display. Pilots performed the best when using the combination of tablet and Glass, displaying a reduced amount of heads-down time and a quicker reaction time when they made navigational errors.

The use of alternate sensory channels may also help offload the visual channel and reduce the amount of time operators spend with their eyes off the horizon. The tactile channel is another potential avenue for presenting information to drivers and pilots, particularly since this sensory channel is not already overloaded (e.g., Ardoin & Ferris, 2014; Fitch et al., 2013; Sklar & Sarter, 1999). The visual channel is responsible for multiple tracking tasks while the auditory channel already presents a multitude of alerts, particularly to pilots. Other potential benefits of using the tactile channel include the following (Gallace, 2007):

- Tactile alerts degrade less in environments with high G-loads than information from other modalities;

- Tactile alerts are not adversely affected by the high level of auditory information already present in the environment (and tactile acuity is better than auditory acuity (Williams, Shenasa, & Chapman, 1998)); and
- The effectiveness of tactile alerts is not dependent on the current direction of the operator's attention.

The tactile channel, however, does already contain some load while driving and flying. For example, many backroads and highways have rumble strips to alert drivers when they veering out of their lane. Similarly, planes are equipped with a stick shaker for stall warnings.

Independent information presented via multiple sensory channels, with each information stream on its own channel, may present processing limitations, particularly when both auditory and tactile channels are involved (Gallace, 2007). Prior studies suggest that the information presented via the visual channel is independent of both the auditory and tactile channels and hence does not result in conflicting processing limitations (Duncan, Humphreys, & Ward, 1997). However, in the case of auditory and visual response performance, operators respond to auditory cues faster than visual cues, particularly when the complexity of the ongoing task was high (Lu, Wickens, Prinet, Hutchins, Sarter, & Sebok, 2013). In contrast to the absence of conflict with visual sensory representations, as both the auditory and tactile channels present information temporally, users can only process a maximum of two auditory or tactile streams at the same time (i.e., they are not independent) (Duncan et al., 1997). Moreover, both auditory and tactile cues have advantages in different circumstances (Lu, Wickens, Prinet, Hutchins, Sarter, & Sebok, 2013):

- In general, operator's respond faster to tactile than auditory cues;

- As task difficulty increases, performance with auditory cues becomes better than performance with tactile cues; and
- In terms of processing codes, for spatial cues, auditory responses are faster than tactile, and for categorical cues, tactile responses are faster than auditory.

In addition to using alternative sensory modalities to reduce the likelihood of interference between primary and secondary tasks, graded alerts may also be used to draw the attention of operators, particularly in high workload or high stress environments when urgent situations occur. Graded alerts are multi-stage displays that can present an alarm signal proportional to danger posed by the situation (Lee, Hoffman, & Hayes, 2004; Sorkin, Kantowitz, & Kantowitz, 1988; Woods, 1995). These types of alerts have been shown to improve attention allocation techniques among tasks (Sorkin et al., 1988). For example, a graded alert that tells a pilot about a developing weather situation gives the pilot additional information about the urgency of a situation over a binary alert. As graded alerts give the pilot a preview of the weather situation, they allow operators to more effectively determine whether or not addressing the alert demands an immediate attentional shift. If the alert warns the pilot that the weather is rapidly deteriorating and they urgently need to engage in an attentional shift, a more pronounced signal would be presented to encourage the operator to immediately gain new information about the weather and to ready them to make a decision about potentially altering their flight path.

2.7.2 Supporting the Primary Task: Adding to the Primary Task Feedback Loop

Alternatively, consider how the primary task may be better supported by using multiple modalities or processing channels, hence reducing interference with the secondary task.

As previously explained, both driving and flying involve multiple divided attention tasks. Moreover, driving and flying can be divided into multiple tracking tasks: both drivers and pilots must monitor speed and position. In these cases both the driver and pilot divide attention within the primary task, creating a "multi-tracking" environment. An additional feedback loop could hence be used to support one of these multi-tracking tasks.

Concurrent feedback is the process of providing the user with feedback in real-time. The benefits to concurrent feedback are that it has the potential to immediately alter driving or flying behaviors and that it can help the operator learn safer maneuvers (e.g., a safe following distance from a lead vehicle) (Donmez, Boyle, & Lee, 2009). Prior studies suggest concurrent tracking feedback can help drivers to adjust their engagement in distracted driving activities in real-time (Donmez, Boyle, & Lee, 2009). To more effectively aid in supporting a multi-tracking task, this dissertation looks at the use of concurrent feedback in the context of a continuously informing tactile display presented in a multitasking environment.

A continuously informing tactile display is the tactile analogue of a sonification. Sonifications are continuous auditory displays that transform data (or data relations) into a sound display (Watson & Sanderson, 2004). The most widely known example of a sonification is the pulse oximetry display used by hospitals for patient monitoring. Within the tactile modality, the use of continuously informing displays (over strictly continuous displays) avoids perceptual phenomena that result in user difficulties in distinguishing changes that occur in the signal (Ferris, Sarter, 2008; Ferris, Sarter, 2011).

The idea behind having a display that provides continuous information is that it will function similarly to the way peripheral vision works in the visual channel. As opposed to focal vision, peripheral vision does not require significant attentional resources and allows ongoing tasks to continue undisrupted unless some form of partial information signals the need for a shift in attention (Woods, 1995). Hence, a continuously informing display should "preattentively" tell the operator the state of the component of the system that it is representing, making it so that the operator can interpret the display's information in parallel with other ongoing tasks or activities (Woods, 1995; Watson & Sanderson 2004).

2.8 Situation Awareness (SA) and Mental Workload (MWL) Metrics

Metrics for task assessment of MWL and SA can be divided into three main categories: operator performance, subjective ratings, and psychophysiological measures (Figure 7, Table 5) (Tsang & Vidulich, 2006). Each of these categories contributes to the understanding of the impact of display characteristics on the human system (i.e., no category alone can provide a comprehensive evaluation of the impact of the interface).

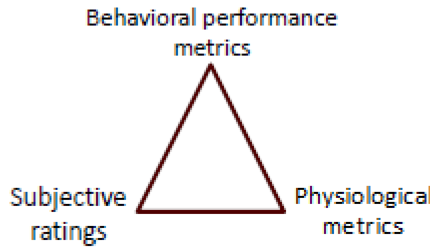

Figure 7. Relationship between types of metrics

Table 5. *Examples of SA and MWL metrics within each task category*

Category	**Situation Awareness (SA)**	**Mental Workload (MWL)**
Operator performance	• Situation Awareness Global Assessment Technique (SAGAT) • Situation Present Assessment Model (SPAM) • Global Implicit Measure (GIM) • Embedded task performance	• Primary and secondary task performance
Subjective ratings	• Situation Awareness Rating Technique (SART) • Subjective Workload Dominance (SWORD)	• National Aeronautics and Space Administration-Task Load Index (NASA-TLX) • Subjective Workload Assessment Technique (SWAT)
Physiological measures	• Potentially fNIRS	• Electroencephalography (EEG) • Heart Rate Variability (HRV) • Skin Conductance Level (SCL) • Functional Near Infrared Spectroscopy (fNIRS)

2.8.1 Operator Performance

Operator performance can be used to measure SA via two main types of methods: (1) Real-time SA memory probes and (2) Implanted SA tasks dependent upon direct performance measurements. Significant debate exists over the pros and cons of each methodology. Performance measures are often thought to be more sensitive than subjective ratings in measuring the severity of resource competition in multitasking (Tsang & Vidulich, 2006).

2.8.1.1 Performance Metrics of Situation Awareness (SA)

The use of real-time memory probes to measure SA has become a mainstream measurement method. The focus of real-time memory probes involves evaluating the response time and accuracy of the data obtained using the probes. The logic behind the use of these probes is that if the operator has sufficient knowledge of that task and environment, then they will be able to answer questions relating to those factors in a timely manner (Tsang & Vidulich, 2006). The primary arguments for the use of real-time probes to measure SA are that correct SA is supposedly more sensitive than a performance measurement and that faulty SA may contribute to performance problems even if this is not apparent from performance data (Durso, Dattel, Banbury, & Tremblay, 2004).

One of the first approaches developed to measure SA was the SA Global Assessment Technique (SAGAT) (Endsley, 1995). When using SAGAT, the primary task scenario is periodically frozen and all the information relating to the task is removed from the scene; the participant is then required to answer randomly selected questions relating to the task and environment (Endsley, 1995). The primary limitation of this technique is that is cannot be used outside of the simulation environment, which places limits on these studies' ecological validity (Endsley, 1995). Additionally, while Endsley (1995) performed multiple experiments validating the technique and showing that these interruptions to the task do not significantly impact overall task performance, a large amount of skepticism relating to the impact of SAGAT on the overall task scenario still exists.

An alternative method to measuring SA that was developed out of the limitations and skepticism surrounding SAGAT is the Situation Present Assessment Model (SPAM), which requires experimenters to ask participants questions relating to the task while they are performing the task (i.e., SPAM is an on-line query technique) (Durso, et al., 2004). This method relies more heavily on response time than accuracy as participants are nearly 100 percent accurate; in fact, response times for inaccurate responses are typically discarded, with a focus only on correct values (Durso, et al., 2004). Proponents of this method argue that SPAM is a viable metric to use in real-world data collection and that SPAM's use of response times is more sensitive than SAGAT's use of accuracy, noting that the use of only correct response time allows researchers to evaluate SA when it succeeds over when it fails (Durso, et al., 2004). Similarly to SAGAT, studies have also been performed validating that the technique does not appear to impact operator performance (Durso, et al., 2004). The logic behind this methodology is that if a participant has the query response stored in active memory, then the response time should be faster than if the participant does not have the response stored in active memory (Durso, et al., 2004).

Implanted tasks that do not require the operator to respond to verbal queries are also used as implicit SA metrics. Though often considered less sensitive than real-time memory probes, implanted tasks are sometimes more appropriate to use for different kinds of research than queries. Researchers predominantly argue that using query methods will disrupt primary task performance, will distract the participant, and may prompt participants to memorize contextual information, all of which can influence the data collected (Jeon, Walker, and Gable, 2014). For example, Jeon, Walker, and Gable (2014) measured SA using implicit performance metrics during each task scenario that they defined as the participants coping strategies with hazardous events, effectively arguing that hazard perception is a viable option to measure SA for dangerous situations. In noting the limitations of this study, however, the authors suggested further research into the use of hazards with a mediated model as they could not effectively distinguish between SA and MWL (Jeon, Walker, and Gable, 2014).

A purely observational alternative to the use of any task protocol is the Global Implicit Measure (GIM). The logic behind this metric is that the operator is attempting to accomplish a set of goals, each with a varying priority level (Tsang & Vidulich, 2006). Hence, experimenters may consider the step-wise progress toward accomplishing specific goals as a performance-based SA measure (Tsang & Vidulich, 2006). To perform this method, task analysis is used to link measureable behaviors with the accomplishment of specific goals; successful accomplishment of properties of these goals are then run through the GIM algorithm to indicate how well the participant performed (Tsang & Vidulich, 2006).

2.8.1.2 Performance Metrics of Mental Workload (MWL)

Measures of workload based on operator performance typically involve the evaluation of primary and secondary task performance.

Direct measures of primary task performance have a limited ability to indicate MWL as they do not reflect variations in allocation of resources due to changes in difficulty level (Tsang & Vidulich, 2006). The information inferred from direct measures of primary task performance must account for the relationship between performance and workload (Figure

8). If the task is not of sufficient difficulty to be on the downward side of the slope (box indicated in Figure 8), then results indicating changes in performance may lead to faulty conclusions about the relationship between experimental conditions, objective performance, and MWL.

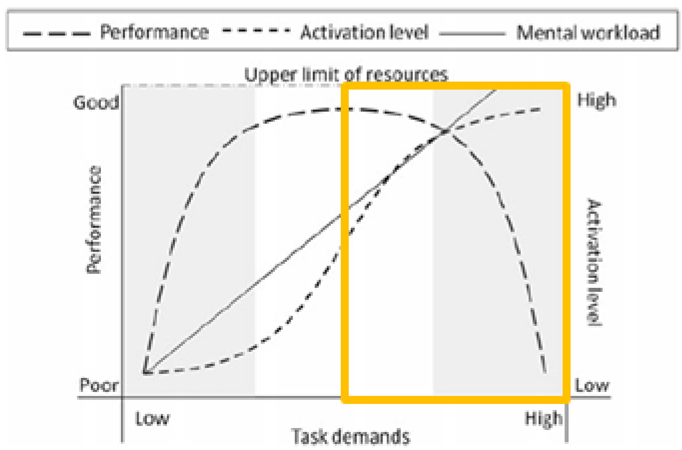

Figure 8. Graph of the relationship between performance, activation level, and MWL (adapted from Young et al., 2015)

The secondary task method attempts to discriminate variations in resource allocation and whether or not the operator has exceeded their information processing capacity (Tsang & Vidulich, 2006). More specifically, to use this method the operator performs a concurrent secondary task, which the experiment explicitly states is of a lower priority than the primary task. Theories on multiple resources and interference suggest that the greater the demand for use of the same resources (i.e., requirement for time-sharing of those resources), the higher the degree of interference between the two tasks (Tsang & Vidulich, 2006). Hence, performance on the primary and secondary task may be analyzed to determine the impact of altering the tasks on MWL (Tsang & Vidulich, 2006).

2.8.2 Subjective Ratings

Subjective ratings offer insight into the operator's perception of SA and MWL when performing a task. Three variables are most important in categorizing subjective metrics:

- Whether the ratings occur along a single or multiple dimensions;
- Whether the ratings are relative/comparative or absolute; and
- Whether the metric must be administered directly after the to-be-rated experience or can be administered at the end of the experiment.

Results from subjective ratings may or may not align with operational performance metrics. Subjective ratings tend to be more sensitive to the number of tasks that must be time-shared in the task set and to the conscious, central processing demand of the task set. However, subjective metrics may be ineffective in capturing workload under low workload conditions that create optimal performance regardless of changes that occur during the experiment or when subjects are performing data-limited tasks where the participant's performance is more heavily reliant on the quality of information than the availability of resources (Tsang & Vidulich, 2006). A higher subjective workload may also legitimately result in better performance as the ratings also indicate the level of effort the participant gives to the task (Tsang & Vidulich, 2006). Subjective data are also often representative of qualitative information gained through post-experiment surveys.

2.8.2.1 Subjective Ratings of Situation Awareness (SA)

The Situation Awareness Rating Technique (SART) is a rating system that measures both perceived workload and perceived understanding of the system components (Endsley, 1995). Hence, while SART is correlated with performance measures, this metric does not effectively delineate between what part of the scale is attributable to workload and what part is attributable to the actual understanding of the system components (i.e., SA) (Endsley, 1995).

The Subjective Workload Dominance (SWORD) metric is an alternative to SART that requires participants to make pairwise comparisons of competing design concepts (Endsley, 1995). The ratings are along a continuum and allow experimenters to assess how much of a difference in workload exists; these preferences are then combined using a hierarchical processing technique and transformed into linear ordering rankings of the design concepts (Endsley, 1995). However, as with SART, some difficulty still exists in determining SA using this subjective preference scaling technique (Endsley, 1995).

2.8.2.2 Subjective Ratings of Mental Workload (MWL)

The National Aeronautics and Space Administration-Task Load Index (NASA-TLX) is composed to six dimensions designed to represent independent "clusters" of variables: mental demands, physical demands, temporal demands, frustration, effort, and performance (Hart, 2006; Tsang & Vidulich, 2006). Each dimension contains a scale in which subjects rate themselves for the scenario, with the assumption being that some combination of these clusters is representative of the subject's perceived workload (Hart, 2006; Tsang & Vidulich, 2006).

An alternative to NASA-TLX is the Subjective Workload Assessment Tool (SWAT). In contrast to NASA-TLX, SWAT is based on three ratings scales: time load, mental effort load, and physiological stress load (Tsang & Vidulich, 2006). However, most studies have found that both NASA-TLX and SWAT have concurrent validity, though NASA-TLX may be more sensitive to detecting lower levels of workload (Tsang & Vidulich, 2006).

2.8.3 Physiological Measures

Physiological measures are typically considered a poor candidate for capturing SA good candidate for capturing MWL (Tsang & Vidulich, 2006). Research is ongoing to determine the best method of using the available data to approximate SA using physiological measures. Hence, the current focus of this section is on the use of physiological measures to measure MWL. Physiological metrics are thought to be the potential "missing link" between discrepancies found in objective performance and subjective ratings.

2.8.3.1 Physiological Measures of Eye Movement

Eye trackers, pupil trackers, and video cameras can be used to obtain eye movement data. While eye movement metrics may be computed using a variety of techniques, all of these techniques center around fixations and saccades (Poole & Ball, 2003). Fixations are periods when the eye is relatively stationary, which allows for the encoding of information; common metrics include the frequency of sampling of a target (with greater frequencies indicating increased interest in the target) and duration of fixation (with longer durations suggesting longer processing times for the object) (Poole & Ball, 2003). In contrast, saccades are the quick eye movements that occur between fixations; common metrics include the number of saccades (with higher counts indicating that more searching is occurring in the environment) and saccade amplitude (with higher amplitudes (i.e., larger saccades) indicating that more salient cues are drawing the user's attention from a distance) (Poole & Ball, 2003).

Two metrics are typically used in driving research: the total amount of time spent looking away from the road and a count of the number of glances away from the road exceeding 1.6 seconds, a critical duration linked with impaired vehicle control and increased crash risk (Horrey & Wickens, 2007). A greater amount of time or number of glances signals that the operator is spending more time with their eyes off the road, which typically results in performance decrements and raises safety concerns. Alternatively, eye tracking devices have also been used in driving simulators, and metrics from these devices have a high correspondence to cognitive load estimations (Palinko, Kun, Shyrokov, & Heeman, 2010).

2.8.3.2 Physiological Measures of Functional Near Infrared Spectroscopy (fNIRS)

Functional Near Infrared Spectroscopy (fNIRS) is an optical imaging technique. The technique provides measurements of ceberal oxy- and deoxygenated hemoglobin levels; these levels reflect changes in inputs to processing during cognitive activities, with higher levels of blood flow to the Preftronal Cortex (PFC) indicating higher MWL.

While recent research suggests that fNIRS has the potential to discriminate workload levels among tasks, the use of fNIRS to test MWL in dynamic task environments is relatively new and poses several difficulties due to a lag between the task time and the data collection points. The primary categories of human factors studies where fNIRS has been used thus far are task discrimination and interference (e.g., Peck et al., 2013; Power et al., 2011), adaptive interfaces and learning (e.g., Solovey et al., 2011; McKendrick, Ayaz, Olmstead, &

Parasuraman, 2014), and working memory and workload (e.g., Brunce et al., 2011; Herff, Heger, Fortmann, Hennrich, Putze, & Schultz, 2013).

2.8.3.3 Other Physiological Measures

Additional metrics for measuring MWL are summarized in Table 6. This summary provides information about the definition of the metrics, location of the sensors, how the measurements are done, and what the measurements indicate about MWL.

Table 6. *List of physiological measures used in required mental resources quantifications (adapted from Yang & Ferris, submitted 2015; Skin Conductance Explained, 2015; Tests and Procedures: EEG (electroencephalogram), 2015; Kawachi, 1997)*

Physiological Measure (Device)	Definition	Sensor Location	Measurement Index	Relationship to MWL
Skin Conductance Level (SCL) (Iom® Wild Divine biofeedback sensor system)	Level of sweat the body excretes	Tips of index and ring fingers of the non-dominant hand	The average skin conductance level over display-processing interval	Higher SCL associates with higher MWL
Electroencephalography (EEG) (NeuroSky® MindWave hardware)	Indication of the electrical activity in a person's brain	Single electrode on the Fp1 position on ventrolateral prefrontal cortex (VLPFC) according to 20-10 international system	The average desynchronization percentage (ERD%) of lower alpha band of the EEG	Larger ERD% of lower alpha band indicates more overall MWL
Heart Rate Variability (HRV) (Zephyr Bioharness 3)	Beat-to-beat alterations in a person's heart rate	Under participants' clothing around the torso	pNN20 and pNN50	Lower pNN20 or pNN50 associated with higher MWL

3. RESEARCH QUESTIONS

Driving uses visual perceptual and spatial working memory for surveying the driving scene, for tracking movements of objects in the environment, and for judging relative locations; driving also uses manual resources, as the hands are on the steering wheel to control the vehicle. Flying adds to this workload by requiring pilots to survey a complex set of instruments in the cockpit; however, airspace typically contains fewer dynamic hazards than the roadway, such as if a deer ran into the road while driving.

Newer technologies brought into vehicles, such as the smartphone in the texting example discussed in the *Motivation* section, commonly require a similar set of mental resources as driving, resulting in interference between the tasks. For example, texting requires visual perception to read incoming messages and verify the text of responses; spatial working memory for the orientation of attention towards the device and for the identification of control locations; and manual resources to press the buttons on the device.

Table 7 lists each novel interface characteristics (i.e., mediating measures) addressed in this dissertation. Each characteristic has the potential to reduce the level of interference between the primary and secondary task. The results from this dissertation are relevant to all cases when the primary task is a dynamic, continuous, on-going process control task; the secondary task is framed as a discrete task. The numbers in the cells are the numbers of the experiments that address the proposed novel interface characteristic.

Table 7. *Dissertation framework*

	Problem States (Types of Interference)					
	Aiding Primary Task		*Aiding Secondary Task*			
			Not Embedded		Embedded	
Solution States (Mediating Measures)	Structural interference	Cognitive interference	Structural interference	Cognitive interference	Structural interference	Cognitive interference
Voice Input			**1**		**3**	
Head-up display			**1**			
Size of Display					**2, 3**	**2**
Discrete tactile signals					**2**	**2**
Continuously informing tactile signals	**3**					

Prior research in the driving and aviation domains has demonstrated both the potential detrimental and beneficial impact of novel interface design components to impact performance and safety. The goal of this dissertation is to address better supporting the operator in multitasking environments by answering three questions.

3.1 Research Question 1: How Does Manipulating the Modalities Used in Presenting the Secondary Task Increase or Decrease the Mental Workload (MWL) of the Primary Task? And of the Task Set?

As noted, when two tasks compete for attention from the same pool of limited mental resources, decrements in performance of the primary or secondary task may occur. Altering the modalities used in processing the information from the secondary task has the potential to reduce the interference between the primary and secondary tasks and allow for improved performance throughout the task set.

Prior research suggests that improved performance can be attained in both the driving and aviation domains through altering interface characteristics (e.g., Drews et al., 2009; Fitch, Hanowski, & Guo, 2015; Valasek et al., 2015). As expressed in Table 7, the research presented here tests the impact of the following display characteristics on MWL of the secondary task by manipulating how users perform the secondary task: voice input, head-up display, size of display, and use of discrete tactile signals.

3.2 Research Question 2: How Does Increasing or Decreasing the MWL Impact the Situation Awareness (SA) of the Primary Task?

As MWL competes for the same pool of limited mental resources as SA, altering the distribution of MWL should alter the operator's SA. Correspondingly to Question 1, as expressed in Table 7, the research presented here tests the impact of the following display characteristics on SA of the primary task manipulating how users perform the secondary task: voice input, head-up display, size of display, and use of discrete tactile signals.

3.3 Research Question 3: How Does Giving the User Additional Information for the Primary Task (in Alternate Modalities) Influence MWL of the Primary Task and of the Task Set as well as SA of the Primary Task?

As opposed to altering the modalities used in the secondary task, this question discusses the potential of providing additional support for the primary task. This refers to evaluating the same MWL and SA explained in Questions 1 and 2, except for the primary as opposed to the secondary task. Question 3 also relies on the same theoretical underpinnings used in Questions 1 and 2, implying that altering the physical and mental resource distribution somewhere else in the task set should also produce reduced or redistributed MWL and higher SA.

As expressed in Table 7, the research presented here tests the impact of using a continuously informing tactile display in the primary task. As discussed in the *Literature Review*, the use of a continuously informing display should "preattentively" give the operator additional information without drawing substantial resources away from the task set.

4. APPROACH

This dissertation is constructed of three experiments. At the center of each experiment is a different multitasking scenario designed to test the components outlined in Table 7.

4.1 Experiment 1

The multitasking setup for the first experiment involved the primary task of driving and the "non-embedded" secondary task of texting. The experiment tested the impact of using voice input and head-up display output in the secondary task to determine the effect of those components on structural interference. The study evaluated four texting scenarios (within-subjects): (1) baseline controlled driving task, and a controlled driving task plus a secondary texting task using (2) a smartphone and its manual touchscreen input, (2) a smartphone and its voice-to-text input, and (3) a head-up display and its voice-to-text input.

Across these scenarios, the time to complete inputting the text message and video-based glance data were used to determine differences in the workload imposed by the secondary task (Question 1). Performance on the driving task was used to determine potential differences in SA in the driving task when the secondary task involved using different modalities (Question 2). Additionally, differences in the reaction times to anticipate and unanticipated driving events also aided in the evaluation of the impact of display type on SA (Question 2).

4.2 Experiment 2

The multitasking setup for the second experiment involved the primary task of flying and the "embedded" secondary task of responding to weather alerts. The experiment tested the impact of size of display output and use of discrete tactile signals the secondary task to determine the effect of those components on structural interference. The study evaluated use of two graphical displays that contained the text of the alert (within-subjects variable): (1) a tablet map interface and (2) a smartwatch interface. The study also evaluated three different types of vibrations that were delivered at the same type as the appearance of the text alert (between-subjects factors): (1) baseline no vibration, (2) single level vibration (i.e., all alerts received the same time of vibration), and (3) graded vibration (i.e., alerts received different types of vibration based on their urgency).

Across these scenarios, the time the pilot took to start and through finishing their response to the weather alerts was used to determine differences in the MWL imposed by the two sizes of display (Question 1). Between the subjects, these same metrics were used to determine the impact of the three types of discrete tactile alerts (Question 1). For both of the within- and between-subjects variables, fNIRS and NASA-TLX data for each scenario were used to evaluate each characteristic's impact on MWL (Question 1). Situation Awareness Probes (SAPs) were used to evaluate the impact of within- and between-subjects variables on SA (Question 2).

4.3 Experiment 3

The multitasking setup for the third experiment involved the primary task of driving and the "embedded" secondary tasks of three types of interaction with a mockup of a police

Mobile Command Terminal (MCT). The experiment tested the impact of using manual versus voice input and size of output display the secondary task to determine the effect of those components on structural interference. Additionally, the experiment also evaluated the potential of a continuously informing tactile display for better supporting the primary driving task.

During the study, each participant performed a set of three tasks on the following devices while driving (within-subjects): (1) a smartphone using manual input, (2) a smartphone using voice input, (3) a touchscreen laptop using manual input, or (4) a touchscreen laptop using voice input. Half the participants (between-subjects variable) were also presented with a continuously informing tactile alert designed to provide the participant with additional information about their current speed, which was matched against a target speed.

Across these scenarios, the time to complete input text was used to determine differences in the workload imposed by the secondary task (Question 1). Performance on the driving task was used to determine potential differences in SA in the driving task when the secondary task involved using different modalities (Question 2). Between the subjects, both these sets of metrics were used to determine the impact of using a continuously informing display to support the primary task on overall MWL and SA (Question 3).

5. EXPERIMENT 1: TEXTING AND DRIVING WITH GOOGLE GLASS

This study was designed to compare the individual and interacting effects of voice-to-text input (vs. manual input) and head-up display (vs. head-down display) on performance in a dual-task. The dual-task included navigating in a driving simulation and reading and responding to short, semi-open-ended text messages in a manner reflecting participant's natural response tendencies. Participants completed a controlled driving scenario while performing the "non-embedded" secondary texting task via three methods: (1) using a smartphone with manual input, (2) using a smartphone with voice-to-text input, and (3) using Google Glass with voice-to-text input. Participants also completed a baseline (no-texting) condition. Google Glass is a small head-mounted transparent prism screen that sits in front of the right eye, and the frame of the device includes multi-axis accelerometers for head-based gesture controls and voice-command and read-aloud functionalities. In addition to texting task measures, which measured MWL, driving performance was assessed according to common metrics associated with driving safety and SA, including the mean of RMS absolute steering rate and standard deviation of lane position (SDLP). The impact of texting on SA was additionally inferred via the mean following distance and differences in brake response time to a lead vehicle during contextually predictable and unpredictable braking events ("pacecar") (Hurts, Angell, & Perez, 2011; Horrey et al., 2006), and video-based analysis of eyes-off-road glance durations were used to further analyze MWL as they indicated the orientation of visual attention.

By comparing driving and texting performance in the baseline and three texting cases, this study distinguishes the benefits of voice-to-text input from those of voice input+HUD displays on MWL and SA. This study builds on others' recent work by including eye orientation metrics and evaluating performance in a diverse set driving environments. In addition to contributing to the knowledge base on human information processing and multitasking performance, current findings can be used to inform policymakers, app developers, and the general public about the implications of interacting with Glass and similar technologies while driving, providing insights into ways to make secondary tasks that require similar resources as texting *less unsafe* (e.g., Liu & Wen, 2004).

5.1 Hypotheses

For this study, as with previous studies, the addition of a secondary texting task was expected to negatively impact driving performance, and hence negatively impact SA, in all cases. Texting with Glass was, however, expected to negatively impact performance the least as its voice input+HUD functionalities make visual resources easier to share and reduce the driver's structural interference with manual resources. Similarly, compared to texting manually on a smartphone, the conditions involving voice input with a smartphone and with Glass were expected to support relatively better driving performance, due to reduced need for visual and manual resources, thus resulting in lower mean of RMS absolute steering rates, lower SDLP, larger differences between contextually predictable and unpredictable even response times, and shorter following distances, all of which indicate increased SA. Glass' HUD functionality was expected to provide added driving performance benefits, further increasing the participant's SA, over using a smartphone (head-down display (HDD)) with voice-to-text entry due to its potential to reduce the driver's eyes-off-road time. The manual texting condition was expected to result in a higher MWL, presenting with longer texting response times and a larger number of safety-critical eyes-off-road glances compared to both voice input conditions as manual texting requires a more attention be given to typing and visual verification. The Glass condition was expected to support the fastest texting times and fewest eyes-off-road glances, imposing the lowest MWL, again due to its combined voice input and HUD functionalities.

5.2 Method

Data collection and analysis activities were completed for 24 participants (15 men and 9 women) aged 20 to 32 years (men: *M*=24.5, *SD*=3.11; women: *M*=23.8, *SD*=1.92) from Texas A&M University. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Texas A&M University. All participants reported normal or corrected-to-normal vision, familiarity with smartphone texting, and had a valid driver's license.

Participants completed a primary driving task in all four experimental conditions and a secondary texting task in three of those conditions. The driving scenarios were constructed in STISIM Drive^{TM}, a medium-fidelity, stationary desktop driving simulator displayed on a 30-inch screen. Drivers used a Logitech G27 force-feedback steering wheel and floor-mounted pedals to control the vehicle (Figure 9).

Figure 9. Experimental setup in the Glass texting condition. In conditions involving smartphone interaction, the device was placed on the table near the mouse. The picture in the right-hand corner is a still shot from the video recordings collected and used for coding the eye movement data.

After signing an informed consent form and completing a background questionnaire, participants received a short training session with the driving simulator. Simulator training involved completing a short scenario with a pacecar that was repeated until it was both satisfactorily completed (i.e., without observing collisions or other unsafe behaviors) and each participant stated that they were comfortable driving in the simulation environment. All participants were able to demonstrate proficiency in the driving task. Participants then completed four test conditions, the order of which was completely counterbalanced: (1) baseline (driving-only); and driving plus (2) reading texts on a smartphone and responding via the smartphone's touchscreen keyboard, (3) reading texts on a smartphone and responding via the smartphone's voice-to-text input (no manual input was permitted), and (4) reading and listening to texts with Google Glass and responding via Glass's voice-to-text input. Prior to each texting condition, participants were trained on how to use and tested for proficiency in use of the respective texting device. As nearly all participants had never used Glass, experimenters aided participants in physically adjusting the device so that the prism was properly positioned within a "sharable" field of view with the roadway. Then each participant completed a short tutorial, which included learning how to navigate the Glass interface and practicing texting. Participants then repeated the simulator training scenario while receiving and sending practice text messages with Glass; the scenario was repeated until participants were able to correctly send two successive text messages while driving. All participants were able to demonstrate proficiency in using all the required texting methods. Prior to every scenario, participants were instructed that their first priority was to drive safely and that their second priority was to answer the texts in a timely manner but understood that their response times for texting were being recorded. After driving all four scenarios, participants completed a post-experiment questionnaire, which included questions about their experiences using each texting device. The experiment lasted approximately one hour.

Information collected in the background questionnaire was used to categorize participants into "experience" levels in driving, texting, and multitasking contexts. The levels ranged from 1 to 5, with 1 being the least experienced and 5 being the most experienced, and each "experience" rating was used as a covariate.

5.2.1 Primary Driving Task

The driving task involved driving a scenario that was 6.3 miles in length, took approximately seven minutes to complete, and spanned both urban and rural driving environments. Participants were instructed to drive safely and near posted speed limits. Each of the four counterbalanced scenarios (one for each test condition) included three stages that were presented in a randomized order: (1) a winding mountain road (Mountain-Road), (2) a city highway with interchanges (Ramp-Highway), and (3) a town square with pedestrians and a sharp left-hand turn (Town-Square). Each stage involved varying densities of vehicle and pedestrian traffic and standard traffic control devices, such as speed limit signs, stop signs, and traffic lights. A "pacecar" led the driver's vehicle throughout each scenario, and participants were instructed to follow this vehicle at a comfortable distance. The pacecar periodically braked, with roughly half of brake "events" occurring at relatively unpredictable times (i.e., when no other roadway events would have suggested a braking response) and the other half occurring at contextually predictable locations (e.g., when approaching steep curves or a sharp turn). The pacecar acted as an implanted task resulting in measurements of MWL. All of the brake event data was considered in a single dataset.

Prior to each scenario, drivers were instructed to drive safely and to obey traffic rules as their highest-priority task. Data were sampled from the driving simulator every foot. Dependent measures are described in Table 8. All dependent measures discussed here are intended to measure SA.

Table 8. *Dependent measures of driving performance*

Measurement	Description	Inferences	Sources
RMS absolute steering rate (deg/sec)	How fast the participant turns the steering wheel (measure of steering activity)	Reflects increases in the number of large and potentially abrupt steering movements to correct heading errors, which indicate higher levels of MWL	Menhour, Lechner, & Charara, 2009; Rosenthal, 1999; Young, Lee, & Regan, 2008
Standard deviation of lane position (SDLP) (ft)	Standard deviation in the location of the participant's vehicle with respect to the roadway's dividing line	Decrements in the amount of lateral position control (i.e., higher levels of SDLP) indicate higher levels of workload	Angell et al., 2006; Rosenthal, 1999; Young, Lee, & Regan, 2008
Mean following distance from the pacecar (ft) (i.e., headway)	Distance between the pacecar and the participant's vehicle	Reflects potential driver control strategies, such as maintaining longer distances for a safety compensation, when drivers are under higher workload	Boer, Ward, Manser, Yamamura, & Kuge, 2005; Sawyer et al., 2014; Young, Lee, & Regan, 2008
Brake response time (sec) (computed the same way for both contextually predictable events and unpredictable events)	Time interval between the activation of the pacecar's brake lights and when the participant's foot let up from the accelerator	Traditionally, longer times reflect higher MWL for resources used to track hazards in the environment In this case, drivers are able to respond to contextually predictable events better when they have a higher level of SA, making a larger difference between the contextually predictable and unpredictable times (i.e., controls for driver's natural response time)	Green, 2000

5.2.2 Secondary Texting Task

The secondary texting task required participants to read and respond to incoming text messages. The order of test conditions (Baseline, Touchscreen_Keyboard, Voice-to-Text, and Glass) was completely counterbalanced among participants. Participants used their own smartphones for the Touchscreen_Keyboard and Voice-to-Text conditions, except for four participants whose phones did not contain the voice-to-text feature. These participants instead used an experimenter-provided phone with the same operating system as their respective phones. In total 18 participants used Android phones and 6 participants used Apple phones.

Participants were trained with each device prior to starting the corresponding scenario and demonstrated proficiency using a baseline texting task. Modeled after Drews, et al. (2009), this baseline texting task involved starting on the smartphone home screen, navigating to the texting interface, and entering and sending the message, “The quick brown fox jumps over the lazy dog”. Prior to the Glass condition, participants were trained on how to access received text messages (i.e., by tilting one’s head up or tapping the side of the glasses frame), how to read messages visually or with Glass’ read-aloud functionality, and how to compose and send responses using Glass’ voice-to-text input. Participants were allowed to use Glass’ read-aloud function to listen to incoming messages but were encouraged and tended to use Glass’ visual display either to quickly read or to visually verify displayed content. This visual verification behavior was confirmed by reviewing video recordings of participant’s eye movements after receiving a text message. All but 4 of the 24 participants showed evidence of visually sampling incoming messages and entered text. The four exceptions presented difficulties in video-coding glances when using Glass due to eye characteristics or to Glass obstructing the eyes, and similar visual verification behavior was assumed because none were associated with outlier data for any dependent measure.

The texting task consisted of reading and responding to six text messages sent by the experimenters, two during each of the three stages of the scenario, using the condition’s assigned texting method. Messages were delivered at predetermined locations that were designed to impose higher workload (e.g., merging onto a highway, approaching an intersection, completing a turn) and at intervals that allowed for at least 45 seconds for participants to respond between messages. Across the three texting conditions, participants received 18 messages that were selected and randomly ordered from a set of 20 prewritten questions. Each question was designed to be of roughly equivalent difficulty for the participant population (refined through pilot testing), involved reading at least three lines of text (~50 characters), and required responses of several words. See examples of these text message questions in Table 9.

Table 9. *Example text messages used in this study and common expected responses*

Text Message	**Expected Response (understandable variants were equally acceptable)**
What are the names of two kinds of fruits and what colors are they?	*bananas are yellow and apples are red*
What are Texas A&M's school colors and what is the mascot?	*maroon and white and a collie*
What is a major sport and two professional teams that play that sport?	*football and the Dallas Cowboys and the Miami Dolphins*
What are the names of two of the major roads in College Station?	*University Drive and Texas Avenue*

Participants were told to respond as they naturally would in a texting conversation with a familiar party, to maintain a consistent response style throughout the experiment, and to address each message completely in their response (i.e., participants could not send multiple texts to address a single question). In order to determine if requiring clarity and accuracy in messages impacted voice input versus manual texting methods, half of the participants (*N*=12) were instructed to correct typing and transcription errors in entered text until they felt the response was satisfactory while the other half were to send responses without correction, regardless of whether the text was entered as intended. Dependent measures of texting performance included texting response times, defined as the time from when the device announced the arrival of a message to when the participant submitted a response, and accuracy of response content. Since response content accuracy rates were all near 100 percent, ultimately this measure was not analyzed.

Video recordings of participant eye movement during all texting scenarios were collected, and glances away from the roadway towards the texting device were manually coded by counting frames in QuickTime. The frame count was then used to tally the number of glances away from the road of 1.6 seconds or greater during each stage. This is considered a critical eyes-off-road glance duration linked with impaired vehicle control and increased crash risk (Horrey & Wickens, 2007). Eye movement metrics were intended to measure MWL.

5.3 Results

All data were analyzed using SAS 9.3 with a significance level of α=0.05 (Tippey, Ritchey, & Ferris, 2015). The four experimental conditions (Baseline, Touch_Keyboard, Voice-to-Text, and Glass) each contained three measurements per participant (i.e., the stages of the scenario). Carryover was tested for as each participant performed all four experimental conditions; as each scenario included three measurements (from the stages), this was the repeated measures variable.

For driving performance analyses, repeated-measures ANOVAs were performed to determine the effects of the four texting conditions using the Proc Mixed function (REML estimates) with an unstructured covariance matrix. Given that all the data approximately met the ANOVA normality and equal variances assumptions, this procedure was chosen because it compensates for missing data values and alternatively accommodates for sphericity (i.e., inconsistencies were observed between the likelihood ratio and univariate tests when using Proc GLM for this analysis). Correspondingly, for the texting response and eye movement analyses, repeated-measures ANOVAs were performed to assess the impact of the three texting methods using the Proc GLM function (i.e., no inconsistencies were observed between the likelihood ratio and univariate tests). The Huynh-Feldt correction was used for all violations of sphericity when using Proc GLM. All means reported, used in confidence intervals, and used in graphs are the adjusted least squares means. Effect size (η_p^2) values for the Proc Mixed procedure were estimated using analogous GLM estimation procedures due to limitations of the SAS software; no effect size was estimated for the Friedman test. Tukey-Kramer post-hoc tests were used to determine differences among means using an α=0.05. None of the covariates, which were computed as 1 to 5 "experience" ratings (with separate ratings for each driving, texting, and multitasking experience) using information from the background survey, were significant and hence all were excluded from the analysis. For analysis of the subjective rankings from the post-experiment survey, a nonparametric Friedman test was performed.

5.3.1 Texting Correction Factor

The effect of requiring text response errors to be corrected or not prior to sending a text was insignificant across participants with respect to all dependent measures. Few instances were observed when participants corrected texts, regardless of input method, perhaps because of the simplicity of the responses required. Since this between-subjects factor did not reach statistical significance, it was eliminated from further analyses.

5.3.2 Situation Awareness (SA) Metrics: Driving Performance

The following are the results of the driving performance metrics.

5.3.2.1 Driver Control Metrics

Mean of RMS absolute steering rate, which represents the speed at which drivers make steering inputs, was significantly affected by texting condition ($F(3,253)=18.03$, $p<.001$, Mountain-Road $\eta_p^2=.183$, Ramp-Highway $\eta_p^2=.026$, Town-Square $\eta_p^2=.250$), which significantly interacted with the repeated-measures variable stage ($F(6,253)=3.62$, $p=.002$). With this significant interaction, post-hoc comparisons were performed for all texting conditions under each stage (Figure 10). During the Mountain-Road and Ramp-Highway stages, performance in the Baseline (no-texting) condition (5.19 deg/sec; 1.55 deg/sec) was significantly lower than in the Touch_Keyboard (8.29 deg/sec, $p=<.001$; 2.99 deg/sec, $p=<.001$) and Voice-to-Text (7.10 deg/sec, $p=.007$; 2.81 deg/sec, $p=<.001$) conditions, which did not significantly differ. Steering rate in the Glass condition (5.92 deg/sec; 2.02 deg/sec) was also significantly lower than in the Touch_Keyboard ($p<.001$; $p=.007$) condition but not significantly different from the Baseline or Voice-to-Text conditions. During the Town-Square stage, only performance in the Baseline condition (4.63 deg/sec) was significantly lower than in the Touch_Keyboard condition (6.14 deg/sec, $p=.004$). No other comparisons reached significance. Across all stages non-significant comparisons had p-values ranging from .117 to .987.

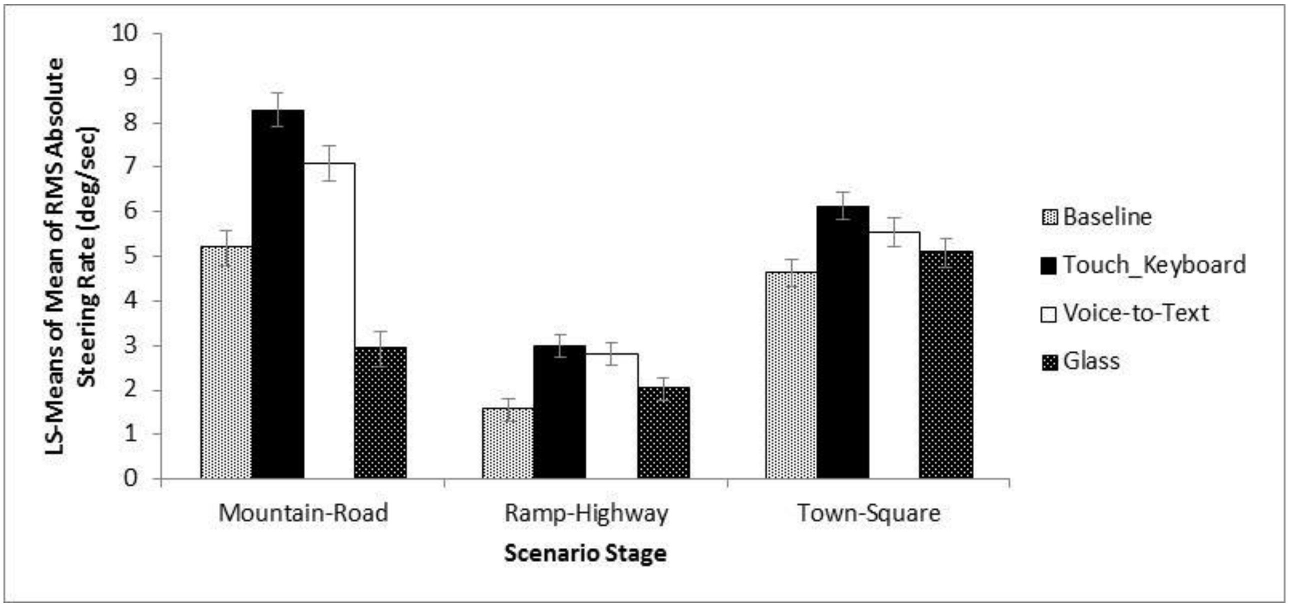

Figure 10. Least squares-means of mean of RMS absolute steering rate (deg/sec) for each texting condition by stage. Error bars represent standard error values.

Standard deviation of lane position (SDLP) did not significantly differ due to for texting condition ((F(3,253)=2.36, p=.072, highest η_p^2 for any stage was .075).

5.3.2.2 Pacecar Metrics

Both the response time and mean following distance pacecar metrics displayed evidence of carryover effect, suggesting that participants improved in these metrics as the experiment progressed (i.e., did not fully plateau in driving performance during the training). Hence only the first texting condition performed by each participant was analyzed.

Mean following distance behind the pacecar (with longer distances associated with driving behavior under higher workload) showed significant main effects for texting condition (F(3,40)=3.95, p=.015, η_p^2=.076) across all stages. Post-hoc comparisons (Figure

11) indicated that mean following distance in the Baseline condition (134.63 ft) was significantly shorter than in the Touch_Keyboard condition (286.74 ft, *p*=.011). No other comparisons reached significance, with *p*-values ranging from .185 to .968.

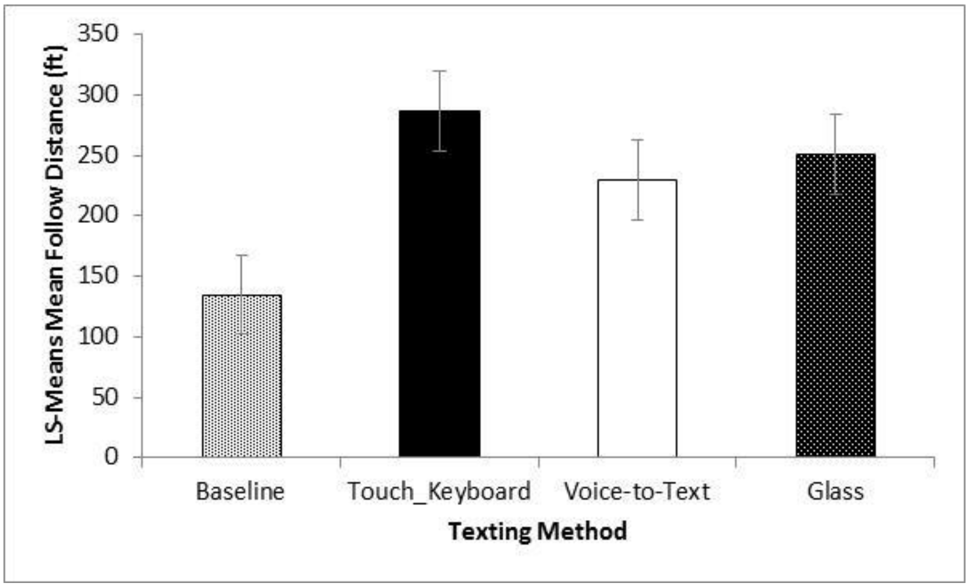

Figure 11. Least-squares means of mean following distance (ft) for each texting condition. Error bars represent standard error values.

Response time to pacecar events was calculated as the time participants took to release the accelerator after the onset of the pacecar's brake lights (with shorter times indicating better roadway vigilance). Data were missing when: (1) the participant did not brake because they were far enough behind the braking pacecar that it was unnecessary; and (2) the participant was already actively applying the brake at the moment the pacecar braked.

The latter case primarily occurred when roadway elements, such as curves, led some participants to initiate an early brake response.

The difference in response times for the events that contained context clues versus the more unpredictable events that did not contain context clues are currently being analyzed. The current analysis that evaluates all response times as a single group is listed here. For response time texting condition (F(3, 49)=2.86, p=.0462, η_p^2=.081) showed main effects across all stages. Post-hoc comparisons indicated that performance in the Baseline condition (0.55 sec) was significantly better than all of the texting conditions: Touch_Keyboard (0.77 sec), Voice-to-Text (1.29 sec), and Glass (1.07 sec). These differences were inferred from the F-test, as the Baseline condition's mean brake response times were non-estimable, but those for the three device conditions were all estimable and did not significantly differ, with p-values ranging from .218 to.872.

5.3.3 Mental Workload (MWL): Texting Response Time

The texting response times, which represent time spent with partial attention devoted to the secondary task (with longer times being worse), were significantly different among texting methods (F(6,226)=7.70, p=<.001, Wilk's Λ=.69, Mountain-Road η_p^2=.163, Ramp-Highway η_p^2=.056, Town-Square η_p^2=.128). Proc GLM analysis suggested that testing for the interaction between texting method and stage violated sphericity ($\chi^2_{2,0.05}$=13.90, p=.001); therefore degrees of freedom was corrected using Huynh-Feldt estimates of sphericity (ε=0.91). Texting method was significant and significantly interacted with the repeated-measures variable stage (F(4,230)=5.77, p<.001) to affect response time. Therefore, post-hoc comparisons were performed within each stage of the scenario (Figure 12).

During the Mountain-Road stage, mean response times in the Touch_Keyboard condition (25.88 sec) were significantly longer than in the Voice-to-Text (20.08 sec, p=<.001) and Glass (21.52 sec, p=.0032) conditions. During the Ramp-Highway stage, mean response times in the Touch_Keyboard condition (21.25 sec) were significantly shorter than in the Glass condition (25.78 sec, p=.035), but no other differences were found. During the Town-Square stage, response times in the Voice-to-Text condition (18.71 sec) were significantly shorter than in the Glass (24.25 sec, p=.031) and Touch_Keyboard (27.25 sec, p=.003) conditions, which did not significantly differ. Across all stages non-significant comparisons had p-values ranging from .151 to .785.

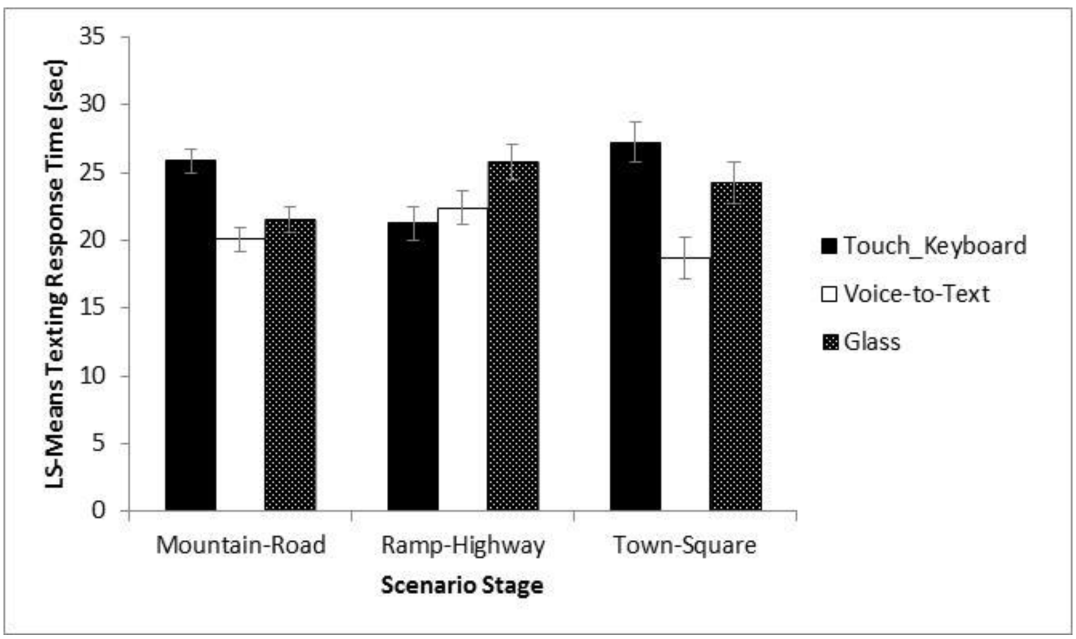

Figure 12. Least squares-means of texting response time (sec) for each texting method by stage. Error bars represent standard error values.

5.3.4 Mental Workload (MWL): Eyes-Off-Road Time

Glances away from the road that were greater than 1.6 seconds were tallied and compared. This threshold is one that was defined by Horrey & Wickens (2007) as indicating increased crash risk (with more glances over 1.6 seconds indicating heightened risk). This threshold, however, differs from the criteria for long glances used in the NHTSA Phase 1 Voluntary Guidelines and is one of three glance metrics suggested by those guidelines. The number of such glances was significantly different among texting methods (F(6,208)=20.47, p=<.001, Wilk's Λ=.40, Mountain-Road η_p^2=0.282, Ramp-Highway η_p^2=0.190, Town-Square η_p^2=0.372), which significantly interacted with the repeated-measures variable stage (F(4, 212)=7.42, p=<.001). Therefore, post-hoc comparisons were performed within each stage of the scenario (Figure 13).

During the Mountain-Road stage, the number of glances in the Touch_Keyboard condition (2.37 glances) was significantly greater than in the Voice-to-Text (0.63 glances, p<.001) and Glass (1.23 glances, p=.007) conditions. In the Town-Square stage, the Voice-to-Text (1.13 glances) and Glass (0.92 glances) conditions similarly involved fewer glances than did the Touch_Keyboard condition (3.22 glances, p<.001; p<.001). During the Ramp-Highway stage, significantly fewer glances occurred in the Glass condition (0.49 glances) than in both the Voice-to-Text (1.75 glances, p<.001) and Touch_Keyboard (2.10 glances, p=<.001) conditions, which did not significantly differ. Across all stages non-significant comparisons had p-values ranging from .099 to .807.

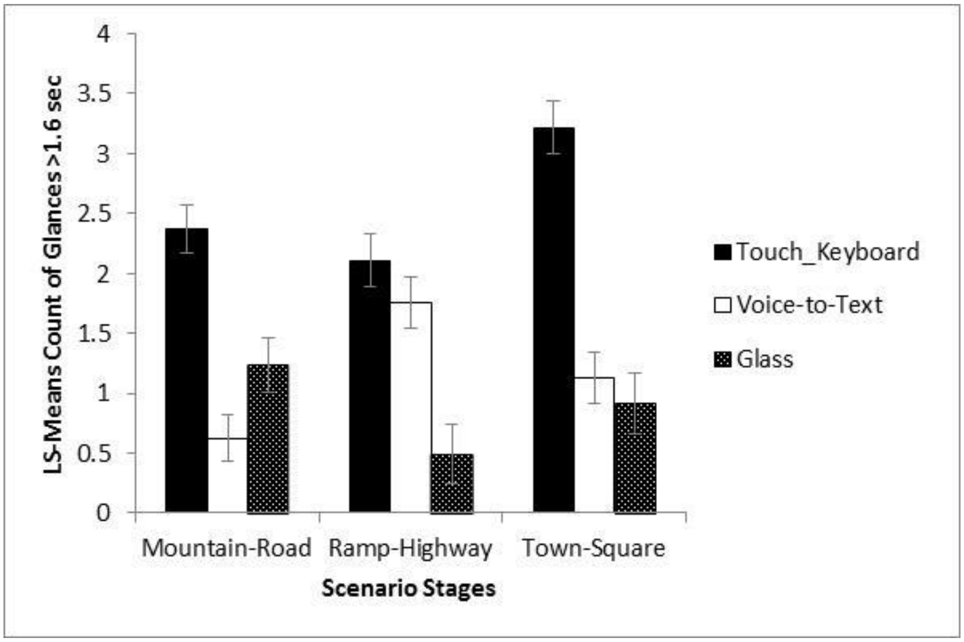

Figure 13. Least squares-means of count of glances >1.6s for each texting method by stage. Error bars represent standard error values.

5.3.5 Subjective Measures

All of the subjective measures differed significantly among texting methods (Table 10). The Touch_Keyboard condition was rated significantly more difficult and involving more dual-task interference than the Voice-to-Text and Glass conditions. The Voice-to-Text condition was also rated significantly worse than the Glass condition for both metrics. Glass' overall rank was significantly higher than the Touch_Keyboard and Voice-to-Text devices, which did not significantly differ.

Table 10. *Summary of analyses of subjective ratings and rankings*

Metric	**Test-statistic**	**P-value**	$\boldsymbol{\eta_p^2}$	**Tukey Groupings (LS-Means)**
Difficulty Rating (1=not difficult, 10=difficult)	F(2, 44)=21.27	<.001	0.363	{Touch_Keyboard (7.26)}, {Voice-to-Text (5.57)}, {Glass (3.52)}
Interference Rating (1=no interference, 10=complete interference)	F(2, 44)=45.47	<.001	0.460	{Touch_Keyboard (8.4348)}, {Voice-to-Text (5.96)}, {Glass (3.57)}
Preferred Device Ranking (1=most preferred, 3=least preferred)	$\chi^2_{2,0.05}$=28.72	<.001	-	{Touch_Keyboard (2.6087), Voice-to-Text (2.22)}, {Glass (1.09)}

5.3.6 Multitask Performance Summary

Table 11 lists the Tukey HSD (α=0.05) groupings for each driving and texting performance metric. Texting conditions listed in a grouping did not significantly differ from each other for the given performance measure and stage.

Table 11. *Letters directly adjacent to each other (i.e., Tukey groupings) indicate device conditions that are not significantly different from each other (B: Baseline, G: Glass, V: Voice-to-Text, and T: Touch Keyboard).*

Driving Performance	Stage	Baseline	Glass	Voice-to-Text	Touch Keyboard
Average absolute steering rate	*Mountain-Road*	BG	BG GV	GV VT	VT
	Ramp-Highway	BG	BG GV	GV VT	VT
	Town-Square	BG	BG GVT	GVT	GVT
Mean following distance	*Same for all 3 stages*	BGV	BGV GVT	BGV GVT	GVT
Reaction time to pace car events (inferred)	*Same for all 3 stages*	B	GVT	GVT	GVT

Texting Performance	Stage	Glass	Voice-to-Text	Touch Keyboard
Response time during scenarios	*Mountain-Road*	GV	GV	T
	Ramp-Highway	GV	GV VT	VT
	Town-Square	GT	V	GT

Eyes-Off-Road time	Stage	Glass	Voice-to-Text	Touch Keyboard
Number glances >1.6 seconds	*Mountain-Road*	GV	GV	T
	Ramp-Highway	G	VT	VT
	Town-Square	GV	GV	T

5.4 Discussion

With regard to input characteristics, this study's results are in accord with prior studies involving general in-vehicle tasks (Tsimhoni & Green, 2001; Horrey & Wickens, 2007) and studies involving interactions with Glass (Sawyer et al., 2014; Beckers et al., 2014; He et al., 2015), showing benefits to voice input (Voice-to-Text and Glass conditions) over manual input (Touch_Keyboard condition) methods in both SA and MWL. Compared to the Baseline (no-texting, driving-only) condition, the Touch_Keyboard condition was associated with the worst driving performance for all driving metrics. Voice input conditions generally showed smaller performance decrements than did the Touch_Keyboard condition, illustrating a relative improvement with the addition of the voice input functionality. The Touch_Keyboard condition was subjectively rated as inducing the highest workload and was the only condition to show significantly greater pacecar mean following distances than the Baseline condition. As longer following distances can indicate higher MWL because drivers facing increased resource demands tend to compensate by maintaining larger safety buffers from lead vehicles (Boer et al., 2005), this suggests the at using a touch keyboard imposes the highest workload, and hence may reduce SA.

When comparing across all dependent measures (i.e., driving, texting, and eyes-off-road metrics (Table 11)), the Glass and Voice-to-Text conditions were part of the same Tukey groupings (i.e., were not statistically different) for nearly all measurements, with two exceptions found in two single stages. For the texting and eyes-off-road metrics, the Touch_Keyboard condition was often significantly different from the other two texting conditions, with exceptions exhibiting no clear grouping pattern. These findings strongly suggest a multitask performance benefit for voice input over manual input, resulting in both higher SA and lower MWL. However, the added benefit from using a head-up display, such as Glass, may not be as consistently impactful.

The impact of display characteristics, though not as pronounced as the effects of input characteristics, significantly influenced multitask performance. The results indicate similar advantages for HUDs as the results found by others investigating Glass in the driving environment (e.g., Becker et al., 2014; He et al., 2015; Sawyer et al., 2014). The benefits of using Glass are clearest when examining the mean RMS absolute steering rate data, indicating its potential benefit to increase SA. Mean RMS absolute steering rate indicates increases in workload through the presence of a greater number of "erratic" steering inputs and corrections. This study's results agree with previous research (e.g., He et al., 2015; Sawyer et al., 2014), showing that across all texting conditions Glass supported significantly fewer abrupt or dramatic steering revisions compared to the other texting methods.

The results of this study suggest that when using Glass, captured visual attention can quickly be reoriented back to the road with relatively minimal consequences to driving, resulting in lower MWL and higher SA. The eye glance analyses confirmed that when a text message arrived on Glass, the driver reoriented their attention to the screen; however, the number of safety-critical glances (i.e., those that exceed the 1.6 second threshold for increased safety risk (Horrey & Wickens, 2007)) while using Glass for messaging was significantly reduced compared to when using the Touch_Keyboard in all stages. Using the smartphone Voice-to-Text method also showed a significant safety improvement over using the Touch_Keyboard in all stages except the Ramp-Highway stage, suggesting that at least some of the benefit Glass provides could be attributable to its voice input functionality. The lack of significant difference between using the smartphone Voice-to-Text versus Touch_Keyboard method in the Ramp-Highway stage may reflect that the longer stretches of straightaway driving in that stage resulted in a lower cognitive workload, potentially allowing participants to more readily glance at the smartphone for longer durations while reading incoming messages and verifying input text.

Despite Glass' reduction in the driving performance decrement due to adding a secondary texting task, using Glass still impaired performance when compared to the Baseline (no-texting) condition. This finding is consistent with other studies whose results indicate that the best overall driving performance occurs in driving-only, no-texting conditions (e.g., Drews et al., 2009; Filtness et al., 2013; Tsimhoni & Green, 2001; Lyngsie, Pedersen, Stage, & Vestergaard, 2013; Yager, 2013).

The subjective metrics indicate that participants perceived texting with Glass as easier (i.e., lower MWL) and as interfering less with driving (i.e., higher SA) than with the other devices. Glass received the highest overall preference rating for supporting the two concurrent tasks among the devices tested. These positive rankings occurred despite participants lack of familiarity with the Glass; however, some of the positive reviews for Glass were likely due to the novelty and perceived "coolness" of the device. Glass' positive assessments were also despite operational problems, such as the device's tendency to overheat while in use, which slowed its processing time and sometimes required experimenters to stop and restart a scenario after allowing it to cool.

Qualitative observations noted while reviewing the video when coding the eye glance data suggested that participants exhibited several differences in behaviors due to their level of "trust" in the Glass technology. Some participants did not appear to trust the functional or navigational features of Glass and thus repeatedly visually sampled the display, re-reading incoming texts and verifying their location within the interface. Other participants exhibited higher trust in the system and visually verified input and output infrequently, most often glancing to confirm the content of an outgoing message. With increasing familiarity using advanced interfaces such as Glass, general user trust levels will likely increase (e.g., Riley, 1996). As display format and quality are two factors known to affect trust in technologies (Lee & See, 2004), future technological developments should consider how both task-related and trust-related factors may impact the frequency and duration of visual reorientations away from the roadway and hence driver safety.

The experimental nature of the Glass technology was a limitation in the current study. As Glass was not commercially available at the time of this study, nearly all participants were unfamiliar with the technology, and some design issues, such as overheating problems, will likely be resolved in future head-mounted wearable technologies. Participants were also primarily from a younger demographic, which affects both driving behavior and familiarity with texting tasks and devices; hence, future work should evaluate performance over broader age range and vary technological experience.

The fidelity of the driving simulator and design of the experimental scenarios were also study limitations. The driving simulator was presented on one (large) monitor and did not include side or over-the-shoulder views; the simulator was also unable to provide the vestibular feedback that results from real-world vehicle movement. The scenarios were designed to emphasize realism, which resulted in a reduction in experimental control of the pacecar's behavior; this led to some larger than expected variances and missing data, which was most problematic in computing the response times to pacecar braking events. Additionally, texts were sent during particularly high-workload contexts during the scenarios, which is not necessarily representative of real-world texting, and participants may have responded to texts more quickly than in the real-world because their safety was not truly at risk.

6. EXPERIMENT 2: WEATHER TECHNOLOGY IN GENERAL AVIATION COCKPITS

This study was designed to compare the individual and interacting effects of screen size and discrete tactile alerts on performance in a dual-task that included navigating in a flight simulation and reading and responding to weather alerts. Participants completed a controlled flying scenario while performing an “embedded” secondary weather alert response task using two graphical weather displays devices: (1) a Windows Surface PC and (2) a Pebble Smartwatch. Participants also received one of three types of vibration from the Smartwatch that was associated with the graphical alert displayed on either device throughout the entire experiment: (1) no vibration, (2) a single vibration, (3) a graded vibration (based on the urgency of the weather alert). In addition to Alert Decision-Action Point (ADAP) metrics, Functional Near Infrared Spectroscopy (fNIRS) was also used to determine MWL. The impact of different display types on SA was evaluated using Situation Awareness Probes (SAPs), which were developed using the Situation Present Awareness Method (SPAM) technique.

By comparing the displays within and across participants, this study distinguishes the benefits of screen size and discrete tactile alerts on MWL and SA, as well as the interacting effects of those characteristics. This study builds on work by the Federal Aviation Administration (FAA) Partnership to Enhance General Aviation Safety, Accessibility, and Sustainability (PEGASAS) Weather Technology in the Cockpit (WTIC) project groups, whose focus is on evaluating weather incidents and developing requirements for the potential certification of new weather displays that general aviation pilots are bringing into the cockpit on external devices such as iPads.

6.1 Hypotheses

Both SA and MWL response metrics were expected to be better in scenarios involving larger screen size and graded vibrations. The larger screen size was expected to reduce the level and number of attention shift that pilots must engage in to gain information from the display, thereby allowing them more time with their eyes on the horizon. The use of graded vibrations was expected to give the pilots more information about the urgency of the scenario and thus allow them to determine how much attention to give to the graphical weather display without having to first evaluate the display, allowing them to more effectively prioritize between flying and weather monitoring activities.

6.2 Methods

Data collection and analysis activities were completed for 32 participants (3 women, 29 men; average age 53 years, ranging from 20 to 79) who were all recruited through a private contractor commonly used at the FAA WJHTC. All participants reported being GA pilots, being at least 18 years old, and holding at least a private pilot certificate. Using a threshold age of 45 years, 11 of participants were categorized as “Younger” and 21 were categorized as “Older”. The mean reported flight experience was roughly 4,900 hours (median 2,200 hours), ranging from 100 to 35,000 flight hours.

Participants were given a primary flying task and a secondary weather alert response task. The flying scenarios were constructed on two high-fidelity Flight Training Devices

(FTDs) at the William J Hughes Technical Center (WJHTC) in Atlantic City, New Jersey (Figure 14).

Figure 14. Pictures from the inside of the FTDs at the WJHTC.

After signing a consent form and completing a background questionnaire, which included information such as age and flight experience (which were used as covariates), participants received a PowerPoint briefing by Certified Flight Instructors (CFIs) that included an introduction to the experimental equipment (e.g., instrument panels, simulated cockpit, alert displays) and information about each of two flight scenarios: the first set in Alaska and the second in New Mexico. Participants were then taken to the simulator room, equipped with and trained on using the smartwatch, and then connected to the fNIRS equipment. Pilots then entered the simulator and received a hands-on introduction to the flight equipment, which was followed by completion of the Atlantic City training scenario. Participants then completed two test conditions (one set in Alaska and one set in New Mexico), the order of which was counterbalanced, that included flying plus the secondary weather alert task, which was performed on (1) the graphical Windows *Tablet* display or (2) the text Pebble

Smartwatch display. Participants received one of three categories of tactile display: (1) no vibration (i.e., *NoVibe*), (2) a single vibration (i.e., *SingleVibe*), or (3) a graded vibration (i.e., *GradedVibe*). Participants were instructed that their first objective was to fly safely and their second objective was to engage in think-aloud verbal protocols which gave experimenters information about participant's response to weather alerts. After each experimental condition, participants completed a NASA-TLX workload index survey. After both conditions, participants completed a post-experiment survey. The experiment lasted approximately two and a half hours.

6.2.1 Primary Flying Task

Two VFR scenarios were developed, both of which were based on NTSB reports of weather-related accidents: (1) a flight from Juneau (PAJN) to Skagway, AK (PAGY) and (2) a flight from Santa Fe (KSAF) to Albuquerque, NM (KABQ). Timelines containing all the events that occurred within each scenario were developed in Excel and used by experimenters to ensure all information was consistently presented across scenarios. During each scenario, the experimenters collected observational data based on the information from the think-aloud verbal protocols using an Excel macro. Participants were also presented with Situation Awareness Probes (SAPs) at key times and locations that were delivered via radio communications from air traffic controllers or other pilots and asked participants to give a status report relaying some aspect of their current location, altitude, airspeed, or other flight-relevant data. Video and audio data were also recorded. By the end of both the scenarios conditions had turned from VMC to IMC.

During the Alaska scenario, pilots flew through a narrowing pass in the Alaska canyon area. Figure 15 shows where deteriorations in weather conditions along the flight path occurred. These deteriorations were manually performed by simulator technicians during the scenario. Participants were given two Situation Awareness Probes (SAPs) during the Alaska scenario (Table 12).

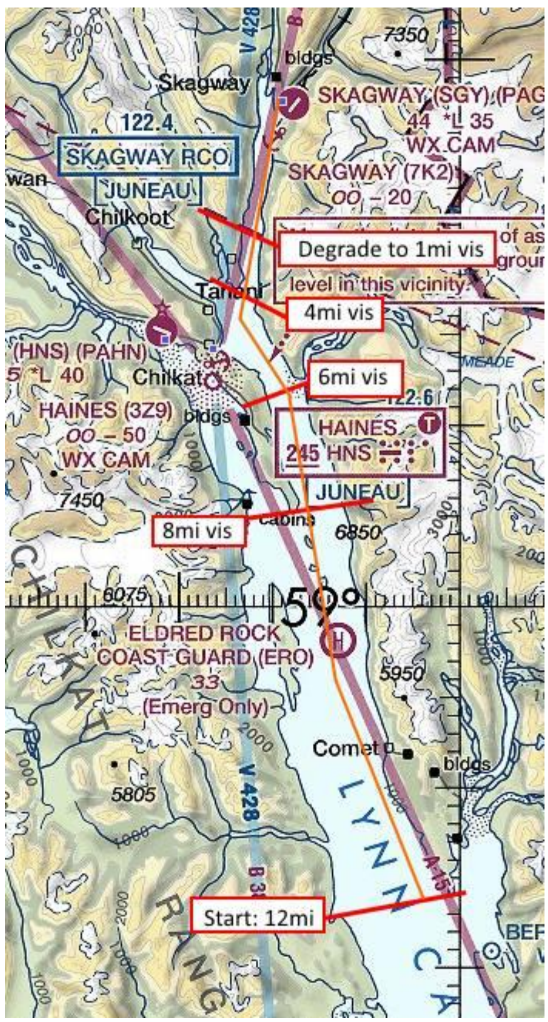

Figure 15. Alaska scenario flight map with markers where changes in weather were rendered. Low visibility and icing develop at destination of Skagway, AK (PAGY).

Table 12. *SAPs used in the Alaska scenario*.

Time During Scenario	ATC Request	Anticipated Participant Response
0:01:00	Traffic in Lynn Canal on CTAF: "Any aircraft within 10 nautical miles of Point St. Mary-say altitude and if experiencing turbulence." **[SAP-1 EVENT OPEN]** [**if** subject asks "state your position"...]: "just departed Juneau enroute to Skagway"	"This is 6JW-currently 10 miles south of Point St. Mary-northbound at 2,500-negative turbulence" (or similar response) **[SAP-1 EVENT CLOSED]**
0:09:30	Traffic in Lynn Canal on CTAF: "Float plane 12Mike is 10 nm north of Chilkat Inlet landing Chilkat… Any aircraft within Chilkat inlet-say position and current flight conditions" **[SAP-2 EVENT OPEN]**	"This is 6WJM-currently abeam Seduction point -enroute to Skagway northbound at [altitude-dependent on action from turbulence encounter]" **[SAP-2 EVENT CLOSED]**

During the New Mexico scenario, pilots flew over gradually rising terrain for the first two-thirds of the flight. This was followed by a dramatic increase in elevation with lowering ceilings for the remainder of the flight. Figure 16 shows where deteriorations in conditions along the flight path occurred. These deteriorations were manually performed by simulator technicians during the scenario. Participants were given two Situation Awareness Probes (SAPs) during the New Mexico scenario (Table 13).

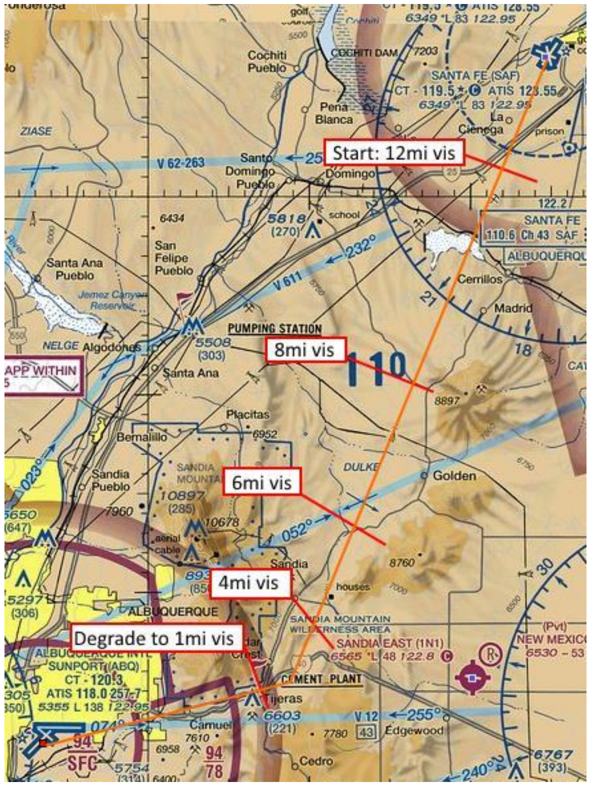

Figure 16. New Mexico scenario flight map with markers where changes in weather were rendered. The map demonstrates the likely diversion route around the mountains when the severe weather conditions are encountered.

Table 13. *SAPs used in the New Mexico scenario*.

Time During Scenario	**ATC Request**	**Anticipated Participant Response**
0:02:00	SAF TWR: "6JW traffic is a bonanza inbound from the south-continue on course---state your altitude." **[SAP-1 EVENT OPEN]**	Subject should relay his/her altitude **[SAP-1 EVENT CLOSED]**
0:10:30	"6JW- state your position and flight conditions." **[SAP-2 EVENT OPEN]**	[Subject should verify squawk code and relay his/her position] **[SAP-2 EVENT CLOSED]**

Dependent measures of primary task performance based on the SAPs included the pilot's Reaction Time (RT) (sec) and the pilot's Completed Response Time (CR) (sec) to the SAP. Using the SPAM assessment method, faster responses were the primary measure used to indicate higher levels of SA as accuracy is expected to be near 100 percent, which was the case for this experiment (Durso et al., 2004).

6.2.2 Secondary Weather Alerts Task

The weather alert response task was performed on either (1) the graphical Windows *Tablet* display or (2) the Pebble *Smartwatch* display. Each participant performed once scenario on each device. Across all participants, pilots were divided into groups and received one of three categories of vibrotactile cues: (1) no vibration alert (i.e., NoVibe alert), (2) single vibration alert (i.e., SingleVibe alert), or (3) graded vibration alert (i.e., GradedVibe alert). The vibrations were delivered via the smartwatch in both graphical display conditions.

All coded text for alerts was modeled after the Lockheed Martin Flight Service (LMFS) Adverse Conditions Alerting System (ACAS) format and included pilot reports and advisories (i.e. PIREPs, AIRMETs, and SIGMETs). The alert text was displayed visually either in a popup textbox embedded in the graphical Tablet display (Figure 17) or on the face of the Smartwatch (Figure 18). The Tablet display included complex graphical content, including a VFR map with an "own ship" indicator that updated its position with global positioning system (GPS) data from the FTDs, and highlighted areas containing weather information that were designed to be of potential interest to the pilot. When incoming weather information was classified as an "alert", the summaries for those alerts would appear in a small text box adjacent to the highlighted area on the map. Participants could touch the highlighted area or alert box to reveal detail about the alert. The alert could be accessed as long as it remained active, and could be hidden or re-revealed by touching the alert text box.

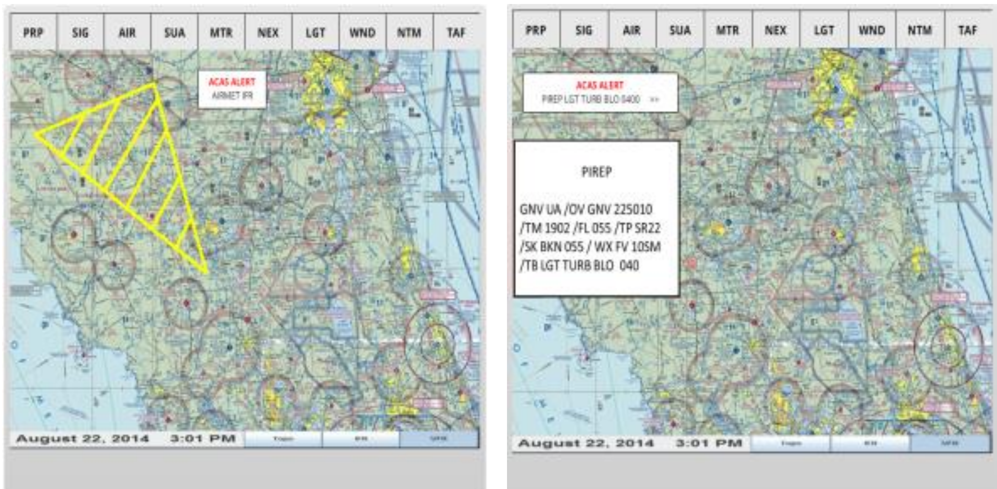

Figure 17. Example screens illustrating alert functions on the Windows tablet display. (Left) An alert summary popup and applicable highlighted area. (Right) The alert detail that is displayed when either the highlighted area or alert summary was selected by pilots by touching the screen at the relevant location.

The Smartwatches were connected via Bluetooth to smartphones controlled by the experimenters. Experimenters manually triggered watch alerts at pre-scripted times. The alert summary was visible immediately on the watch face; up and down buttons on the right side of the watch could be used to scroll through screens to read the whole text of the alert (Figure 18).

Figure 18. Pebble smartwatch display. (Left) The smartphone software controller. (Right) An example text of the ACAS alert summary.

Depending on the vibration conditions associated with the participant, the Smartwatch presented SingleVibe, GradedVibe, or NoVibe with each alert. As the watch was worn throughout both scenarios, in the Tablet display conditions, the vibrations were issued to coincide with Tablet-displayed alerts (popup textboxes), but no text was displayed on the watch in these conditions.

Both the Alaska and New Mexico scenarios each contained three Alert Decision-Action Points (ADAPs) where alerts of low, medium, and high urgency, respectfully, were issued (Tables 14 and 15). Responses to ADAPs were used as indicators of MWL.

Table 14. *Alert Decision-Action Points (ADAPs) used in the Alaska scenario*

Time During Scenario	**Weather/Position**
0:01:57	**Alert 1** PIREP Low urgency JNU UA/ OV 25SSE HNS /TM 1702/ FL025/ TP PA32 /SKC SCT 55/ WX FV 10SM/LGT TURB BLO- 040/RM PT SHERMAN VCNTY
0:07:55	**ALERT 2** Med Urgency AIRMET WA7O JNUS WA 111700 AIRMET SIERRA FOR MT OBSC VALID UNTIL 1122000 LYNN CANAL AND GLACIER BAY MTS OCNL OBSC IN CLDS/PCPN. OTLK VALID 111700-112000.
0:13:57	**Alert 3** High Urgency JNU UUA/ OV AGY /TM 1710/ FL010/ TP C206 /SKC BKN 08/ WX FV 01SM BR/RM FRZA TAIYA INLET.

Table 15. *Alert Decision-Action Points (ADAPs) used in the New Mexico scenario*

Time During Scenario	**Weather/Position**
0:06:58	**ALERT 1** Low Urgency PIREP ABQ UA/ OV 1N1 /TM 1705/ FL085/ TP C402 /SKC SCT 550/ WX FV 6SM/LGT TURB BLO- 090/RM NEAR SANDIA PARK
0:10:57	**ALERT 2 Medium urgency** **AIRMET** **ABQS WA 201710Z AIRMET SIERRA FOR MT OBSC VALID UNTIL 201900** **MTN OBSCN VALID UNTIL 1900Z** **MTS OCNL OBSC IN CLDS/PCPN.** **OTLK VALID 201710 -201900.**
0:11:59	**ALERT 3 High urgency** **AB CWA 201710** **ZAB CWA101 VALID UNTIL 1900** **ISOLD SVR TSTM OVER ABQ MOVG SSE** **10 KTS MVFR-IFR CIG** **New Mexico is 5/20/ at 1700Z**

Dependent measures of secondary task performance centered around the measurement of MWL, which was done by analyzing responses to the ADAPs and filtering the fNIRS data. For ADAPs, the same measurements of Reaction Time (RT) (sec) and Completed Response Time (CR) (sec) were used as with SAPs, with faster times indicating lower levels of MWL.

The fNIRS system used in this study was the fNIR100B stand-alone functional brain imaging system originally designed by Drexel University (Merzagora & Izzetoglu, 2016; Ayaz et al., 2013) (Figure 19). The device includes 16 diodes (i.e., light detectors) that measure changes in the concentration of oxy- (HbO_2) and deoxygenation (Hb) with respect to a control (Merzagora & Izzetoglu, 2016). HbO_2 was analyzed in this study. COBI control device software was used to calibrate the device and collect data. Initial data was filtered using fNIRSOFT analysis software, which uses a Beer-Lambert filter. Scenario means and maxes for each diode were then computed in using R scripts.

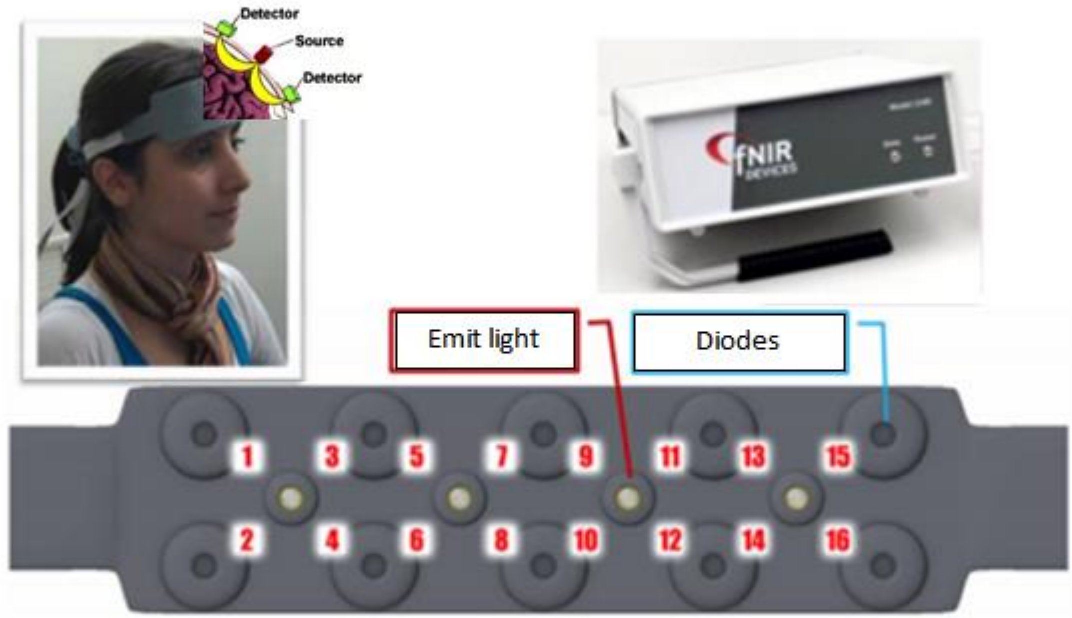

Figure 19. fNIR100B system overview (adapted from Merzagora & Izzetoglu, 2016; Ayaz et al., 2013)

6.3 Results and Discussion

All data were analyzed using SAS 9.3 with α=0.05. Repeated-measures ANOVAs and ANCOVAs were performed using the Proc Mixed function (REML estimates) with *Graphical* (within-subjects) and *Tactile* (between-subjects) as the main fixed variables. Participant and Scenario were both modeled as random variables, each serving as blocking terms. The multiple ADAP and SAP measurements within each scenario served as the repeated measurements variables within each respective model. fNIRS measurements were modeled at the Scenario level using diode locations as repeated measurements. NASA-TLX measurements were modeled at the Scenario level and did not include repeated measurements. Participant Age and Experience level were both analyzed as covariates.

Tukey-Kramer post-hoc tests were used to determine differences between means (α=0.05). Means and standard errors reported in the text are LS-Means values; those depicted in graphical illustrations, however, are means calculated manually from the raw data. Data were lost in coding the ADAPs and SAPs due to issues with the video recordings.

6.3.1 Background Questionnaires

Review of the survey results suggests approximately 65 percent of participants never received any weather training beyond that given in basic pilot training. However, nearly 70 percent reported familiarity with various weather information systems that are accessible in-flight, such as weather radar displays in embedded GPS systems or tablet or smartphone apps such as ForeFlight.

6.3.2 Situation Awareness (SA): Situation Awareness Probes (SAPs)

Both the SAP RT and CR datasets indicated the presence of outliers. A standard method of computing quartiles was used to remove outliers.

6.3.2.1 SAP Reaction Time (RT)

RT was significantly affected by the interaction of *Tactile* and *Graphical* display factors ($F(2,27)=8.29$; $p=.005$). Post-hoc comparisons showed a significant difference under the NoVibe condition between the Tablet (M=27.57 sec, SD=10.78 sec) and Smartwatch (M=50.47 sec, SD=10.16 sec, p=.006) displays and under the GradedVibe condition between the Tablet (M=20.36 sec, SD=2.66 sec) and Smartwatch (M=14.99 sec, SD=2.46 sec, p=.011) displays. A significant difference was also present under the Tablet condition between the

NoVibe (*M*=12.06 sec, *SD*=3.24 sec) and GradedVibe conditions (*M*=20.36 sec, *SD*=2.66 sec, *p*=.044) and between the SingleVibe (*M*=11.72 sec, *SD*=2.75 sec) and GradedVibe conditions (*p*=.032). No other combinations were significantly different from each other (Figure 20).

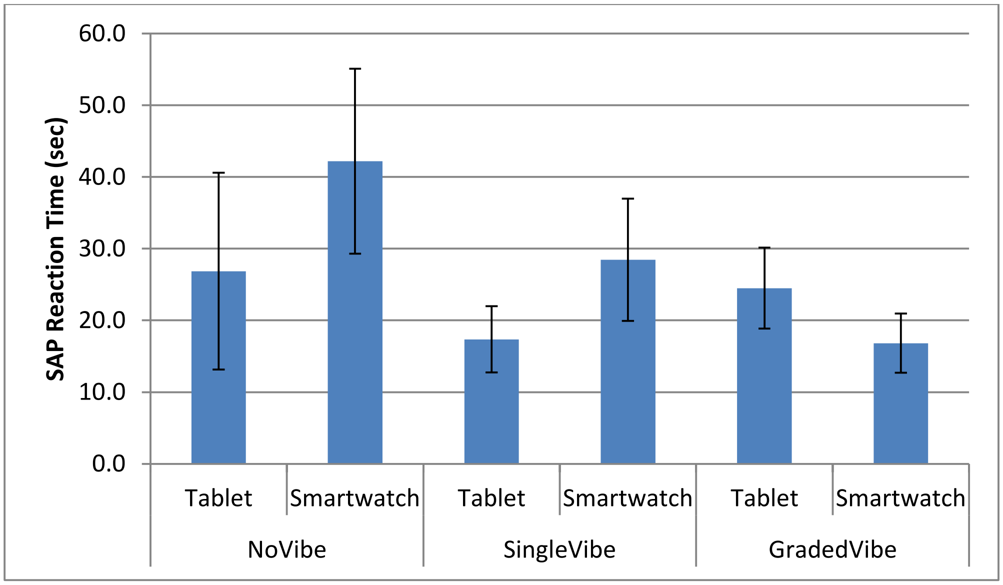

Figure 20. Mean SAP Reaction Time (RT) for each combination of Tactile and Graphical display factors. Error bars represent standard error.

6.3.2.2 SAP Completed Response (CR) Time

None of the independent variables (*Graphical* and *Tactile)* or covariates (Age and Experience) were significant, with *p*-values ranging from .240 to .855.

6.3.3 Mental Workload (MWL) Performance Metrics: Alert Decision-Action Points (ADAPs)

Of the 55 live-coded recordings of participants across the two scenarios, nine pilots correctly decided to divert when weather conditions degraded and approached IMC levels in one of the two scenarios. Note that no participant made the correct decision in both scenarios, and that more often decisions were made correctly in the New Mexico scenario than in Alaska, which was always presented second.

6.3.3.1 ADAP Reaction Time (RT)

Reaction Time (RT) to alerts was significantly affected by *Tactile* ($F(2,17)=3.82$; $p=.043$) but not *Graphical* or the interaction between *Graphical* and *Tactile* display components, with p-values ranging from .534 to .906. Post-hoc comparisons showed a significant difference between the NoVibe ($M=69.33$ sec, $SD=15.61$ sec) and SingleVibe conditions ($M=27.34$ sec, $SD=13.19$ sec, $p=.049$) and between the NoVibe and GradedVibe conditions ($M=20.44$ sec, $SD=10.67$ sec, $p=.016$). However, the SingleVibe and GradedVibe alert conditions were not significantly different from each other (Figure 21).

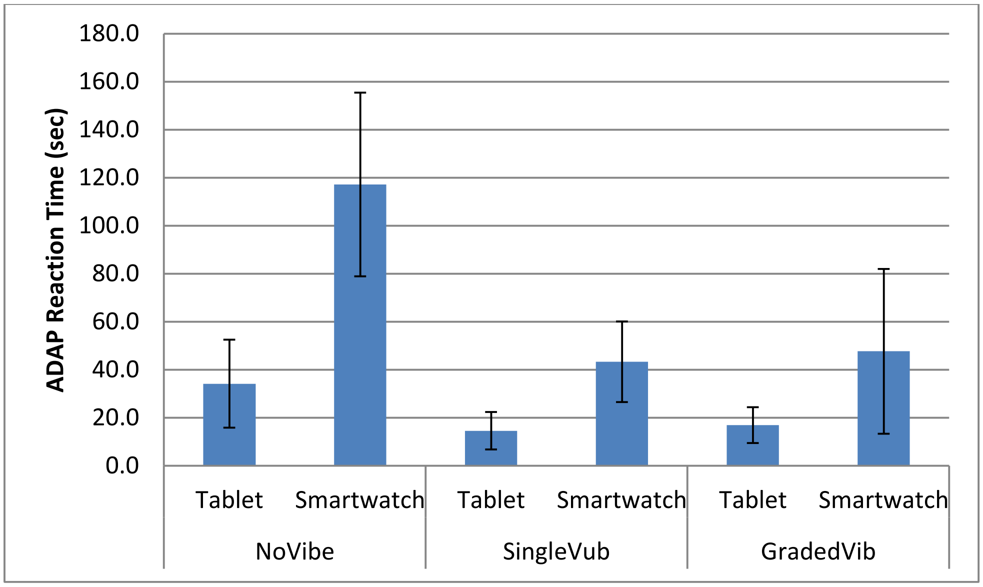

Figure 21. Mean ADAP Reaction Time (RT) for each combination of Tactile and Graphical display factors. Error bars represent standard error.

6.3.3.2 ADAP Completed Response (CR) Time

None of the independent variables (*Graphical* or *Tactile*) or the covariates (Age and Experience) were found to be significant, with *p*-values of the independent variables ranging from .179 to .214 (Figure 24).

6.3.4 Mental Workload (MWL) Physiological Metrics: fNIRS

Modeling the fNIRS means and maxes using all 16 diodes resulted in non-estimable values due to missing data, which was the result of poor diode connectivity. Upon further analysis of the data from each diode, the data set was reduced to only the even numbered diodes because those presented with better connectivity (i.e., a more complete dataset). Based on continued issues with repeated measurements analysis, *Diode* was then used as a random blocking factor as opposed to a repeated measurements factor in the model.

6.3.4.1 fNIRS Means

fNIRS means were significantly affected by the interaction of *Tactile* and *Graphical* components ($F(2,359)=3.73$; $p=.025$), with the *Graphical* components also displaying significant effects ($F(1,359)=8.91$; $p=.003$). Post-hoc comparisons showed that under the SingleVibe condition, using the Tablet ($M=-0.30$ sec, $SD=0.80$ sec) required significantly less HbO_2 than using the Smartwatch ($M=0.82$ sec, $SD=0.80$ sec, $p=.003$).

6.3.4.2 fNIRS Maxes

fNIRS maxes significantly affected the *Graphical* ($F(1,355)=14.08$; $p=.001$) but neither *Tactile* nor the interaction between *Graphical* and *Tactile* display components, with *p*-values ranging from .258 to .743. Post-hoc comparisons showed that using the Tablet ($M=2.94$, $SD=0.76$) required significantly less HbO_2 than using the Smartwatch ($M=3.34$, $SD=0.76$ sec, $p=.001$).

6.3.5 Mental Workload (MWL) Subjective Metrics: NASA-TLX

A Cronbach's alpha was performed to test for the internal consistency among the dimensions of the workload survey ($\alpha=.74$). The combined α in the low acceptability range along with scatterplots of the data indicated that each dimension should be evaluated separately. No independent variables were significant, with ranges of *p*-values listed in Table 16. This was likely because participants all viewed the scenarios as being very difficult prior to the changes in the *Graphical* and *Tactile* display factors. For Physical Demand, the covariate *Age* ($F(1,28)=7.31$, $p=.012$) was significant, with older pilots rating the scenarios to be more difficult than younger pilots (Figure 22). The dataset was hence divided into *Young* and *Old* and the model re-run. None of the independent variables were significant, with *p*-values ranging from .122 to .335 in the *Young* group and .131 to .636 in the *Old* group.

Table 16. *Range of p-values for dimensions of NASA-TLX. Italicized measures are thought to have factors that may be approaching significance.*

Dimension	**Range of *p*-values**
Mental Demand	*.073 to .957*
Physical Demand	.300 to .683
Temporal Demand	*.183 to .481*
Performance	*.087 to .623*
Effort	.322 to .961
Frustration	*.351 to .640*

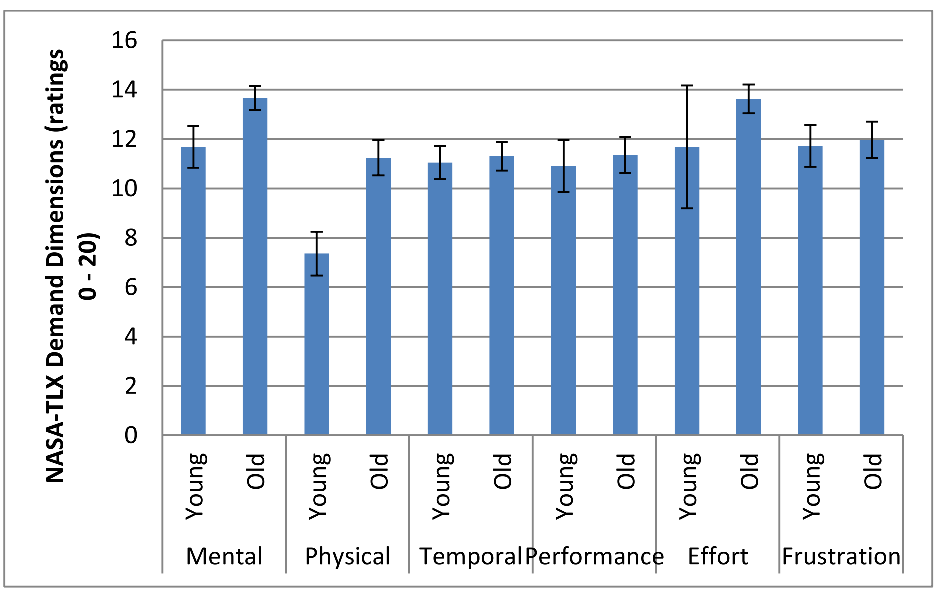

Figure 22. NASA-TLX subjective ratings for each dimension of demand. The ratings scale for each dimension ranges from 0 to 20.

6.3.6 Post-Flight and Post-Experiment Survey Results of Interest

Over 90 percent of the participants stated preferring the Tablet over the Smartwatch, noting issues with the smaller display on the watch. Additionally, nearly 70 percent of participants receiving either a SingleVibe or GradedVibe viewed the vibration as useful, with only 25 percent viewing the vibration as distracting. As a post-test check, the post-experiment questionnaire asked participants to identify which of the adverse weather events were presented in the scenarios. For the Alaska scenario, 8 of the 32 participants recalled having encountered all three of these adverse weather conditions, with 7 of those 8 participants having received vibrations for ACAS alerts. In the New Mexico scenario, 16 of the 32 participants recalled having seen at least the first two 2 of the 3 adverse weather conditions presented in the scenario (as more diverted before encountering IFR conditions in this scenario), with 10 of those 16 having received vibrations for ACAS alerts. Other notable observations from data collection included pilots' lack of familiarity with and inability to read the coded PIREP information on the Smartwatch and pilots' loss of SA when processing underspecified alert information.

6.4 Discussion

Monitoring incoming weather information poses significant challenges to GA pilots, especially as workload increases when they encounter potentially hazardous weather situations (Nall Report, 2015). This study suggests flight safety potentially benefits from the use of larger, graphical displays and the use of vibrotactile cues to call attention to new weather-related information.

Both the ADAP and SAP RT data indicated significant benefits for the Tablet over the Smartwatch display. The ADAP RT results suggested increased ease of processing information from the Tablet, resulting in lower MWL than when pilots used the Smartwatch. Trends in the ADAP CR data toward longer response times with the Smartwatch corroborated these findings, indicating the Smartwatch required higher levels of interaction with the display (i.e., scrolling to read the entire ACAS) and thus resulted in higher levels of MWL. The SAP RT results also indicated a clear benefit for the Tablet, suggesting that the larger screen and consolidation of information within the map on the screen kept users better informed of situational details. The fNIRS data further corroborated this finding, indicating that the overall MWL was less when pilots used the Tablet as opposed to the Smartwatch. This finding also aligned with responses to the post-experiment questionnaire, which indicated pilots had a significant preference for the Tablet over the Smartwatch.

Both the ADAP and SAP RT data also indicate significant benefits when pilots were provided with some form of vibrotactile notification, suggesting that these alerts attract the pilot's attention speeded alert identification, thus aiding in the redistribution of MWL. The lack of difference between the SingleVibe and GradedVibe alert cases reflects the use of dependent measures that correspond with alert acknowledgement and not information contained within the alert, suggesting that a SingleVibe is sufficient to register a "reaction" from the pilot. This is in contrast to previous findings indicating that graded alerts benefit interruption management (Hameed et al., 2009; Ferris & Sarter, 2011; Lee, Hoffman, & Hayes, 2004). Correspondingly, the SAP RT data suggested that the pilots that received tactile alerts underwent fewer attention shifts as they likely felt more informed about the situation and were hence less likely to seek additional data sources. The post-experiment questionnaire and the post-test check data further support the argument for the advantages of vibrotactile cuing.

The increase in average age of GA pilots has prompted concerns about the extent to which novel technologies are implemented in the cockpit (Air Safety Institute, 2012). The NASA-TLX workload survey showed a significant difference in perceived workload based on pilot age, with older pilots perceiving the physical demands of the scenarios to be more taxing. Additionally, in the post-flight questionnaire, older pilots cited having more difficulty with relying on the Smartwatch technology due to its small font size and insufficient lighting.

Experimental limitations of this study centered on the participants lack of familiarity with smartwatch technology and their inability to read coded PIREPS on the Smartwatch. Several cases also occurred when the Smartwatch was not adequately secured, resulting in the vibrations being masked by the engine and other environmental vibrations. Data limitations of this study centered on difficulties coding the ADAP and SAP times due to audio and video recording issues. Given this, experimenters were also not able to confirm cases when participants did not respond (i.e., omissions) and thus were unable to analyze that data; accounting for this would have refined the results. Additionally, at times when impoverished alerts were presented, data analysis did not correct for cases when the pilots were already looking at a display when the alert appeared.

Continued research in this area should accommodate for these limitations as well as approach the issue of best practices for training pilots on how to use wrist-based devices.

7. EXPERIMENT 3: SUPPORTING EMERGENCY VEHICLE MOBILE COMMAND TERMINAL USE WHILE DRIVING

This study was designed to compare the individual and interacting effects of voice-to-text input (vs. manual input) and display size on performance in a dual-task set that included navigating in a driving simulation and three "embedded" secondary tasks that represent tasks commonly encountered by police officers while on duty when interacting with their Mobile Command Terminals (MCTs). Participants completed a controlled driving scenario that involved maintaining a speed of exactly 50 mph while performing a "non-embedded" secondary texting task via four methods (within-subjects): (1) using a smartphone with manual input, (2) using a smartphone with voice input, (3) using a touchscreen laptop with manual input, and (4) using a touchscreen laptop with voice input. Additionally, half of the participants (between-subjects) were given a continuously informing tactile display to aid in speed maintenance. The three secondary tasks were designed to measure different processing capacities while driving and included a spatial (navigation) task, a search task, and a tracking or ranking task. In addition to secondary task measures, which measured performance-based MWL, subjective MWL was also measured after each scenario using the NASA-TLX. Driving performance was also assessed via common metrics of driving safety and SA, including the mean of RMS absolute steering rate and standard deviation of lane position (SDLP).

By comparing driving and secondary task performance in the four (within-subjects) experimental cases, this study is able to distinguish the benefits of display size when performing each of the three types of secondary tasks. SA and MWL performance across participants (between-subjects) (i.e., whether or not the participant used the continuously informing tactile display) indicated the potential of alternative displays in better supporting the primary task in dynamic environments.

This study builds on others' recent work concerning MCT use in emergency vehicles (Yeager et al., 2015) as well as in using concurrent tracking as a driving aid (Yang et al., 2015). This research contributes both to the knowledge base on human information processing and timesharing and can inform developers and emergency responders on the best way to mitigate the risks associated with using MCTs while driving.

7.1 Hypotheses

Based on the literature review conducted by Yeager et al. (2015), the addition of the secondary MCT-representative tasks was expected to negatively impact performance in all cases by directly increasing MWL and thus indirectly decreasing SA. In accordance with the findings from Experiment 1 (Google Glass study), verbal input on both the smartphone and mMCT were expected to result in better driving and secondary task performance than manual input as voice input results in less competition for visual and manual resources, indicating the redistribution of MWL and indirectly resulting in higher SA. Similarly, in accordance with the findings from Experiment 2, the larger touchscreen laptop display was expected to result in better driving and secondary task performance by directly reducing MWL and indirectly increasing SA. The addition of the continuously informing tactile display, which Yang et al. (2015) showed independently aids in driving performance, was expected to result in better performance across all device-entry combinations for both the driving and secondary tasks as additional information in this alternate modality frees up visual resources to aid in faster completion of secondary tasks, thereby reducing eyes-off-road time. This would directly lead to the redistribution of MWL and increase SA.

7.2 Methods

Data collection and analysis activities were completed for 28 participants from Texas A&M University. All participants reported normal or corrected-to-normal vision, familiarity with smartphone texting, and had a valid driver's license.

Participants were given a primary driving task that centered around maintaining a speed of 50 mph and a set of three secondary tasks commonly performed by a police officer while driving (Yang, You, & Ferris, 2013). The driving scenarios were constructed in STISIM Drive^{TM}, a medium-fidelity, stationary desktop driving simulator displayed on a 30-inch screen. Drivers used a Logitech G27 force-feedback steering wheel and floor-mounted pedals to control the vehicle.

After signing a consent form and completing a background questionnaire, participants received short training sessions with the mock Mobile Command Terminal (mMCT) and iPhone 4S smartphone, the driving simulator, and, if applicable, the tactors. For mMCT training, each participant practiced the set of three secondary tasks until they were able to complete all three without making any errors. For the smartphone training, participants were familiarized with the iPhone 4S interface and instructed to practice entering text through manual and voice input. After this, for simulator training, each participant completed a five minute drive through hilly roads while maintaining a target speed of 50 mph. Participants were trained to accelerate as quickly as possible to 50mph at the beginning of each scenario and maintain that speed throughout. Those participants using the tactors as additional sensory information for speed maintenance were guided through putting on the tactor belt prior to receiving the mMCT or simulator training. Participants using the tactors were trained on how to identify speed deviations from that target based on the location the tactile vibrations and then asked to speed up and slow down to feel the difference in the tactors during the simulator training. All participants demonstrated proficiency in their respective trainings. Participants then completed four test conditions, the order of which was counterbalanced, that included driving plus the secondary task set, which was performed on (1) a smartphone using manual input, (2) a smartphone using voice input, (3) a touchscreen laptop using manual input, or (4) a touchscreen laptop using voice input. The tasks performed on the cell phone were analogous to those developed for the mMCT. Half of the participants also received a continuously informing tactile display to aid in maintaining the target speed for the primary task in all four scenarios. Participants were instructed that their first objective was to drive safely and their second objective was complete the secondary tasks but to also be aware that their times to complete the tasks were being recorded. After finishing each scenario, participants completed the NASA-TLX. After finishing all scenarios, participants completed a post-experiment survey. The experiment lasted approximately one hour.

7.2.1 Primary Task

The primary driving task involved driving safely and maintaining a target speed of 50 mph on winding rural roads containing moderate traffic density. Each scenario lasted five minutes and involved three randomized stages, each containing three curves (Figure 23).

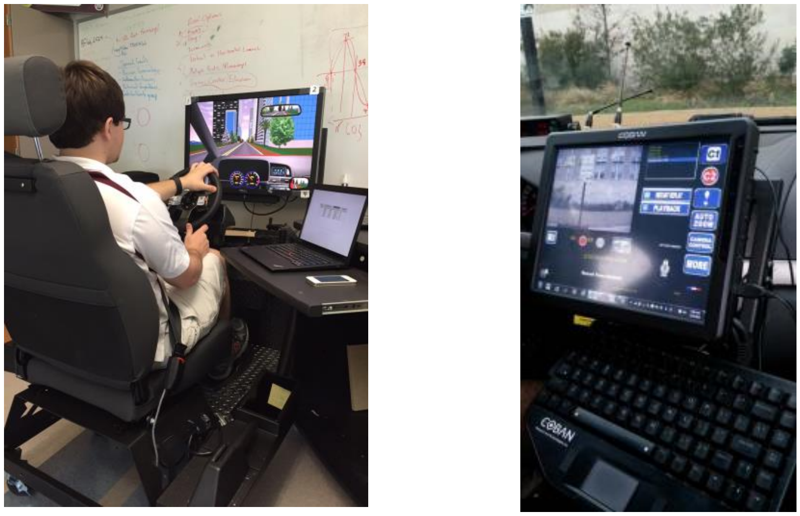

Figure 23. Comparison of setup of the driving simulator and mMCT (left) with that of a real patrol vehicle (right) (copied from Yeager et al., 2015).

To evaluate whether or not providing operators with additional information for the primary driving task may promote safer driving, half of participants were trained using a continuously informing display to aid in maintaining the 50 mph target speed (i.e., speed-tracking). The display used in this study was based on Yang et al. (2013) and used their tactile-spatial encoding method. The tactile display used set of eight C-2 tactors and an ATC 3 controller, which were developed by Engineering Acoustics, Inc.. The eight tactors used in the display were affixed with Velcro to a modified weight-lighting belt (designed to support the torso by providing a small amount of compression) and arranged horizontally and symmetrically across the participant’s lower back (Figure 24). The tactors were set to maximum gain and communicated the participant’s relative speed using sets of continuously informing vibrations (Ferris & Starter, 2011) (Table 17). In contrast to the experiments by

Yang et al. (2013), participants received no vibration when maintaining the target speed in this experiment.

Table 17. *The nine speed levels and their respective tactile presentations across the eight tactors (adapted from Yang, Nevins, & Ferris, 2015). For example, when the participant was driving at 52 mph, they received vibrations on tactors 5 and 6.*

	Speed (mph)	Tactile-Spatial (location; Hz)
	Greater than 54	8; 262
	53-54	7, 8; 258
	52-53	6, 7; 254
	51-52	5, 6; 251
Acceptable Speed	49-51	No vibration
	48-49	3, 4; 249
	47-48	2, 3; 246
	46-47	1, 2; 242
	Less than 46	1; 238

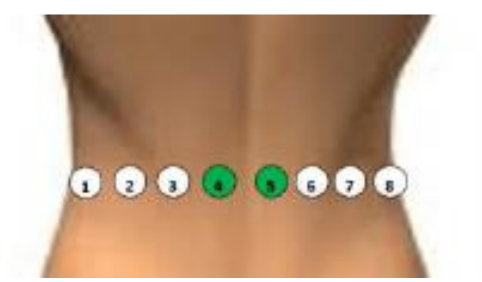

Figure 24. The eight tactor locations that were presented on the participant's lower back (adapted from Yang, Nevins, & Ferris, 2015).

Prior to each scenario, drivers were instructed that driving safely and obeying traffic rules was their highest-priority task. Data were sampled from the driving simulator at 1 cycle/foot-driven. Dependent measures are described in Table 18.

Table 18. *Dependent measures of driving performance*.

Measurement	Description	Inferences	Sources
RMS absolute steering rate (deg/sec)	How fast the participant turns the steering wheel (measure of steering activity)	Reflects increases in the number of large and potentially abrupt steering movements to correct heading errors, which indicate higher levels of mental workload	Menhour, Lechner, & Charara, 2009; Rosenthal, 1999; Young, Lee, & Regan, 2008
Standard deviation of lane position (SDLP) (ft)	Standard deviation in the location of the participant's vehicle with respect to the roadway's dividing line	Decrements in the amount of lateral position control (i.e., higher levels of SDLP) indicate higher levels of workload	Angell et al., 2006; Rosenthal, 1999; Young, Lee, & Regan, 2008
Average Speed Deviation from Target (50 mph)	The amount of deviation of the average speed from the target speed indicates the driver's ability to maintain the target speed	Reflects the mental workload required to maintain the target speed	
Standard Deviation Longitudinal Speed	The standard deviation in the speed maintenance task indicates the driver's ability to effectively maintain the target speed	Reflects the mental workload required to maintain the target speed	

7.2.2 Secondary Task

The secondary task set required participants to perform three tasks, each once and in a random order, during each of the four driving-device scenarios: (1) a spatial task (navigation), (2) a search task (license plate), and (3) a tracking or ranking task (notification).

The secondary tasks were performed on a touchscreen laptop using a custom Java Application (combination referred to as mock MCT (mMCT)) and on a smartphone through text messages. During the voice input conditions, no manual interaction with the devices was permitted. The order of the device conditions (smartphone with manual input, smartphone with voice input, mMCT with manual input, and mMCT with voice input) was counterbalanced among participants.

The mMCT was run on a Lenovo Carbon X1 laptop with a 14-inch touchscreen. During the manual-mMCT condition, participants were instructed to use the touchscreen to navigate through the menus and the keyboard to enter data. During the voice input-mMCT condition, participants were instructed to use the touchscreen to navigate the menus and, once clicking the entry field, to verbally enter their response by speaking aloud. For this condition, experimenters remotely manipulated the mMCT interface using a Logitech Bluetooth connected keyboard in response to the commands by the participants. The smartphone condition was run using an iPhone 4S. For both the manual and voice input-smartphone conditions, the participants used the phone’s touchscreen to navigate the menus. During the manual-smartphone condition, participants used the touchscreen keyboard to enter text; during the voice input-smartphone condition, participants used the phone’s built-in voice-to-text functionality to enter text.

The spatial and search tasks were the same for the mMCT and cell phone. For the spatial task, participants were signaled to view the map, either by a beep they received from the simulator during the mMCT conditions or by the text-received sound on the smartphone, and instructed to provide directions from the start point to the end point listed on the map (Figure 20). The map presented was randomly selected from a set of three maps, where one map was repeated across the scenarios. For the search task, a license plate appeared on the top left of the simulator of the simulator screen. Participants then entered the license plate number into either the mMCT or smartphone and received information about the status of the car (i.e., insurance number, make, model, number of violation) (Figure 20). Participants then searched the status information for the total number of active violations for that vehicle and made a determination on whether or not to pull the car over – if the number of violations was greater than zero then the participant pressed the "pull over" button on the left side of the front of the steering wheel and if the number of violations was zero then the participant pressed the "clear" button on the right side of the front of the steering wheel. Each license plate presentation was randomly selected from a set of three license plates, where one license plate was repeated across the scenarios.

The tracking or ranking task was different for the mMCT and cell phone but designed to call upon the same mental resources (Figure 25). On the mMCT, participants were required to track the "Notifications" button throughout the scenario. When the "Notifications" button turned red, participants were required to enter that task from the main menu and verbally announce the police code in progress for all lines that were yellow or red. Participants touched the screen on the yellow or red code as they announced each code. On the cell phone, participants received a text message containing a list of items they had to rank. This required participants to read and mentally interpret the list and to respond to the text, thereby approximating the same visual, manual, and cognitive resources as they had to use for the analogous mMCT task.

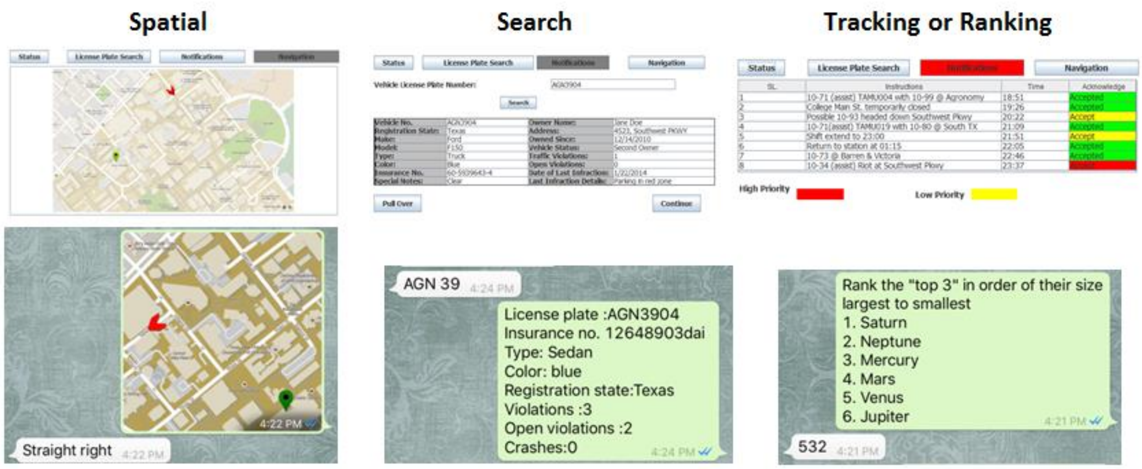

Figure 25. The top row are images of the three tasks (spatial, search, tracking, respectively) from the mMCT and the bottom row are the images of the analogous task that were presented on the iPhone 4S (spatial, search, ranking, respectively).

Dependent measures of texting performance included manual coding of response times in Excel, which was defined as the time from when the device announced the arrival of a message to when the participant submitted a response, and the evaluation of whether or not participants correctly pulled over the vehicle in the license plate task. Participants were told the accuracy of their voice-to-text entries and manual keyboard entries was not being measured.

7.3 Results

All data were analyzed using SAS 9.4 with a significance level of α=0.05. Values approaching the α-level are discussed to provide a more holistic perspective of the analysis. The Tactor and No-Tactor groups were analyzed separately, with Repeated-measures

ANOVAs and ANCOVAs used to determine differences among measures within each group, forming two $2 \times 2 \times 3$ (2 Devices, 2 Input Methods, 3 Types of Tasks (nested within Device × Input Method)) fractional factorial designs. Three covariates were used that were obtained from the background survey: Gender, Driving Experience, and Texting Experience. The models were also tested for order, sequence, and carryover effects across the treatment conditions. Models for each dependent measure (driving metrics, secondary task data, and driving performance metrics) were all estimated using the Proc Mixed function (REML estimates) with an unstructured covariance matrix. Each model was reduced to its simplest form when covariates and extraneous effects were insignificant, giving more degrees of freedom to the independent variables. All reduced datasets were run without nesting Type of Task in order to provide more appropriate degrees of freedom for calculations. Tukey-Kramer post-hoc tests were used to determine differences between means using an $\alpha=0.05$. For practical relevance, post-hoc means and standard deviations presented in the text are the real means as calculated in Excel, and those represented in the graphs are the fitted values (i.e., least-squares means) so that the data presented and images align with the model results, thus appropriately representing the statistics.

7.3.1 Tactors

As the use of tactors was between-subjects dichotomous variable, these data were split into two groups (Tactor and No-Tactor). Each group was analyzed separately to determine if the same trends emerged between both groups, indicating that the datasets could be merged. The same trend did not occur in each group, therefore the datasets were analyzed separately and a comparison of the differences in within group results is made between the two groups.

To analytically support the split of the dataset, T-tests were performed to determine if differences exist between means and variances. When variances differed, the appropriate Pooled or Satterwaite adjustments were made in computing the T-test. Variances between datasets were significantly different for the Mean RMS Absolute Steering Rate (*Folded F*(167,167)=3.91, $p<.001$), Standard Deviation of Lane Position (*Folded F*(167,167)=2.77, $p<.001$), Secondary Task Response Time (*Folded F*(158,160)=1.38, $p<.044$), and NASA-TLX (*Folded F*(51,51)=1.97, $p<.015$). Results suggest that the means significantly differed for Mean RMS Absolute Steering Rate ($t(247.24)=-2.46$, $p=.015$), with the mean for the Tactors condition ($M=11.53$ deg/sec, $SD=8.18$ deg/sec) being greater than the No-Tactors condition ($M=9.79$ deg/sec, $SD=4.14$ deg/sec), and that means are nearly significantly different for Standard Deviation of Lane Position ($t(273.64)=-1.82$, $p=.071$), Average Speed Deviation from Target ($t(334)=1.55$, $p=.121$), and Standard Deviation Longitudinal Speed ($t(334)=1.77$, $p=.077$), with the means for Tactors condition being greater than the No-Tactors condition for Standard Deviation of Lane Position (Tactors: $M=2.88$ ft, $SD=3.12$ ft, No-Tactors: $M=2.37$ ft, $SD=1.88$ ft) but less than No-Tactors condition for Average Speed Deviation from Target (Tactors: $M=1.26$ mph, $SD=1.49$ mph, No-Tactors: $M=1.52$ mph, $SD=1.56$) and Standard Deviation Longitudinal Speed (Tactors: $M=2.73$ mph, $SD=2.05$ mph, No-Tactors: $M=3.12$ mph, $SD=2.03$ mph). Significances between both means and variances among the metrics support the decision to divide the dataset into Tactor and No-Tactor groups.

7.3.2 Situation Awareness (SA): Driving Performance

The following are the results of the driving performance metrics.

7.3.2.1 Mean RMS Absolute Steering Rate

Mean of RMS absolute steering rate represents the speed at which drivers make steering inputs. For the Tactors condition, no within-subjects variables significantly differed due to the test conditions, with *p*-values of the main effects of the independent variables ranging between .469 and .839.

For the No-Tactors condition, no within-subjects variables significantly differed due to the test conditions, with *p*-values of the main effects of the independent variables ranging between .149 and .207. Both the Device ($F(1,142)=1.79$, $p=.183$) and Input Methods ($F(1,142)=2.10$, $p=.149$) conditions appeared to be approaching significance. Post-hoc tests indicated that using a Smartphone (M=10.24 deg/sec, SD=4.41 deg/sec) resulted in a greater mean RMS absolute steering rate than using the mMCT (M=9.34 deg/sec, SD=3.81 deg/sec) and that using Manual entry (M=10.01 deg/sec, SD=3.64 deg/sec) resulted in a greater mean RMS absolute steering rate than using Voice entry (M=9.58 deg/sec, SD=4.58 deg/sec).

7.3.2.2 Standard Deviation of Lane Position (SDLP)

SDLP represents the driver's amount of lateral vehicle control, with lower values indicating better control. For the Tactors condition, no within-subjects variables significantly differed due to the test conditions, with *p*-values of the main effects of the independent variables ranging between .469 and .839.

For the No-Tactors condition, the covariate Gender was significant ($F(1,142)=11.30$, $p=.001$). The data were hence split into a Male group and a Female group. Within the Female group, SDLP did not significantly differ due to any test condition, with p-values ranging from .127 to .630. Within the Male group, Input Method ($F(1,94)=5.20$, $p=.025$) significantly impacted SDLP (Figure 26), with Manual entry ($M=2.04$ ft, $SD=1.05$ ft) resulting in a larger SDLP than Voice entry ($M=1.70$ ft, $SD=0.70$ ft).

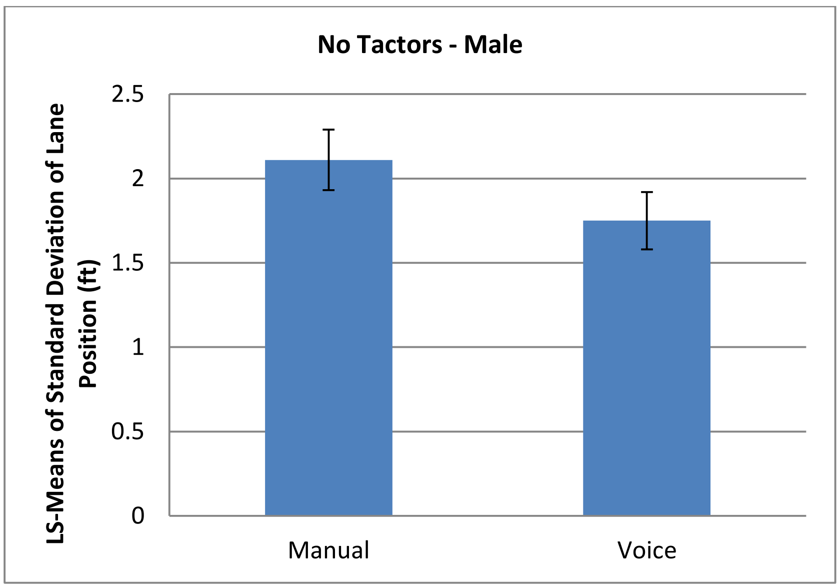

Figure 26. Least squares-means of the standard deviation of lane position (ft) for male participants under the No-Tactors condition. Error bars represent standard error values.

7.3.2.3 Average Speed Deviation from Target (50 mph)

The amount of deviation of the average speed from the target speed indicates the driver's ability to maintain the target speed, with smaller deviations indicating better performance. For the Tactors condition, the interaction of Device and Input Method (*F*(1,142)=5.66, *p*=.019) significantly impacted Average Speed Deviation from Target (Figure 27), with Input Method being significant (*F*(1,142)=5.66, *p*=.043), and all other effects for independent variable combinations with *p*-values ranging between .318 and .469. Post-hoc tests indicate that when using the mMCT device, participants had a significantly smaller deviation from the target speed when using the Manual entry (*M*=1.00 mph, *SD*=1.69 mph) versus the Voice entry (*M*=1.66 mph, *SD*=1.88 mph) method. When using the Voice input method, the difference between using a Smartphone (*M*=1.18 mph, *SD*=1.14 mph) to perform the tasks versus the mMCT (*M*=1.66 mph, *SD*=1.88 mph) was approaching significance. No other post-hoc interaction tests were significant.

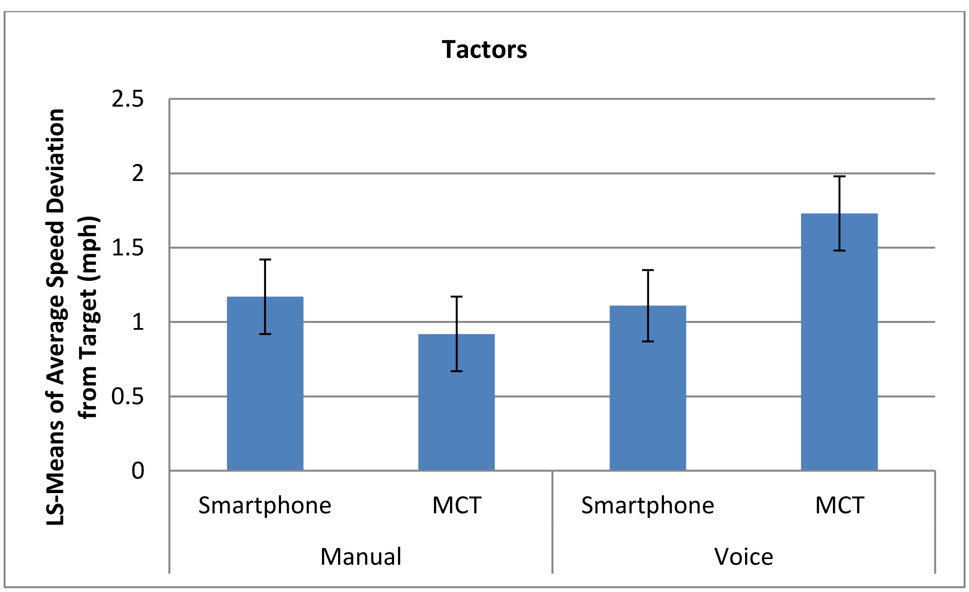

Figure 27. Least squares-means of average speed deviation from the target speed of 50 mph under the Tactors condition. Error bars represent standard error values.

For the No-Tactors condition, the carryover effect was significant ($F(5,137)=3.37$, $p=.007$). Hence, a reduced dataset was run. In that reduced dataset, none of the factors were significant, with *p*-values ranging from .224 to .932.

7.3.2.4 Standard Deviation Longitudinal Speed (SDLS)

SDLS indicates the deviation in speed that occurred while trying to maintain the target speed, with smaller deviations indicating better performance. For the Tactors condition, no within-subjects variables significantly differed due to the test conditions, with *p*-values of the variables ranging between .276 and .823.

For the No-Tactors condition, both the covariates for Gender ($F(1,142)=32.85$, $p<.001$) and Driving Experience ($F(1,142)=8.49$, $p=.004$) were significant. As the Gender variable had a lower *p*-value, the dataset was divided into groups first based on this variable. Within female participants, the Driving Experience covariate was no longer significant. The Type of Task ($F(2,49)=18.05$, $p<.001$) significantly impacted the SDLS, with the impact of the interaction between Device and Input Method showing strong indications of approaching significance ($F(1,49)=2.93$, $p=.093$) (Figure 28). The License Plate task ($M=5.19$ mph, $SD=2.72$ mph) resulted in a significantly greater SDLS than both the Navigation ($M=3.55$ mph, $SD=1.74$ mph) and Notification ($M=3.05$ mph, $SD=1.48$ mph) tasks. When using the Smartphone, the Manual entry method ($M=3.93$ mph, $SD=2.01$ mph) resulted in near significantly greater SDLS than the Voice entry method ($M=3.97$ mph, $SD=2.46$ mph).

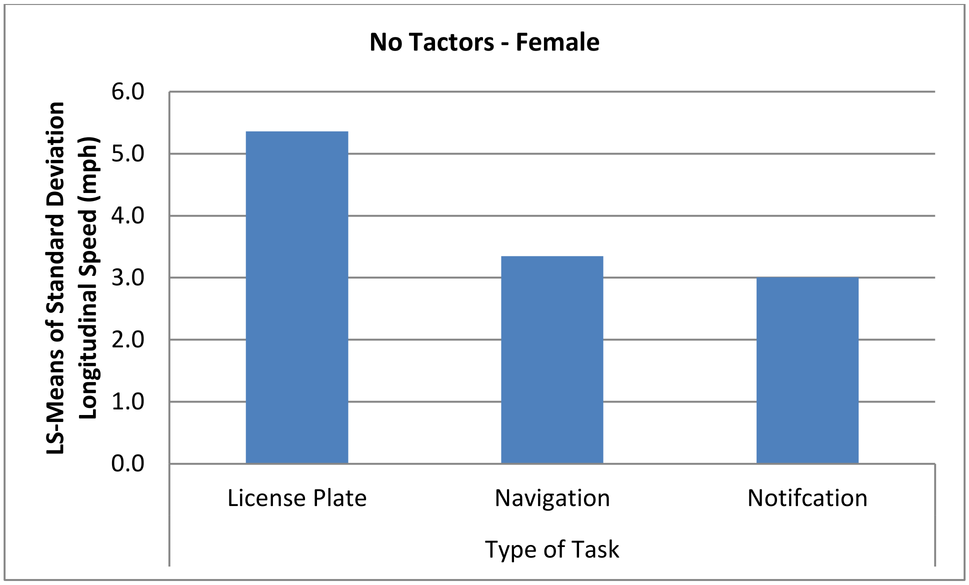

Figure 28. Least squares-means of standard deviation in longitudinal speed (mph) for female participants under the No-Tactors condition. Error bars represent standard error values.

Within male participants, the covariate Driving Experience (1-4 scale) was still significant (F(1,94)=7.16, p=.009), and hence the dataset was further broken into four groups based on driver experience (Figure 29). Those drivers with the least experience (level 1) exhibited a significant difference in SDLS based on Type of Task (F(2,6)=8.29, p=.019), with the Navigation task (M=5.08 sec, SD=2.22 mph) resulting in significantly higher SDLS than the Notifications (M=1.80 mph, SD=0.91 mph) task. Those drivers with slightly more experience (level 2) exhibited a significant difference in SDLS based on Input Method (F(1,28)=8.08, p=.008), with Manual entry (M=2.80 mph, SD=1.65 mph) resulting in significantly higher SDLS than Voice entry (M=2.67 mph, SD=0.93 mph). Those drivers with added experience (level 3) exhibited only an approaching significant difference in SDLS based on Input Method (F(1,28)=3.178, p=.086), with Manual entry (M=2.49 mph, SD=2.08

mph) resulting in higher SDLS than Voice entry (*M*=2.79 mph, *SD*=2.39 mph). Those drivers with the most experience (level 4) exhibited a significant difference in SDLS based on the interaction of Device and Input Method ($F(1,17)=4.54$, $p=.048$), with Device also exhibiting significant main effects ($F(1,17)=24.32$, $p=.001$). Post-hoc tests indicate that under the Manual entry condition, SDLS is significantly higher when using the mMCT (*M*=1.91, *SD*=0.66) versus the Smartphone (*M*=1.82 mph, *SD*=0.71 mph). Under the Voice entry condition, SDLS is also significantly higher when using the mMCT (*M*=2.64 mph, *SD*=0.76 mph) versus the Smartphone (*M*=1.74 mph, *SD*=0.25 mph).

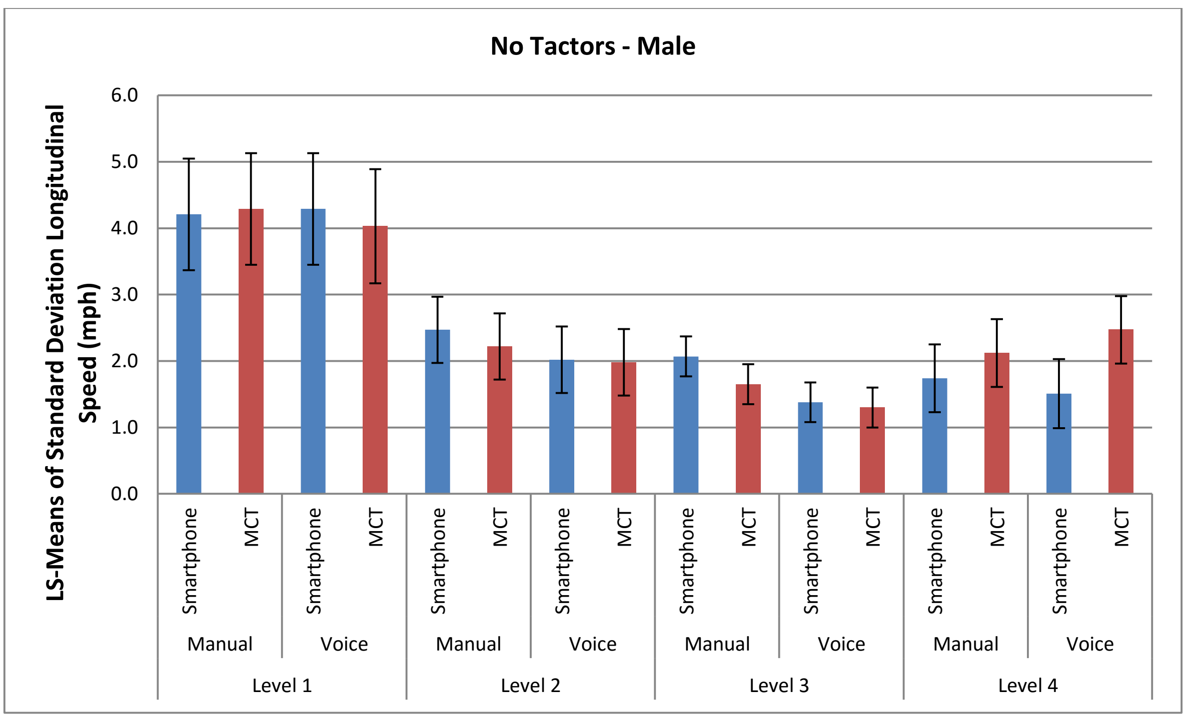

Figure 29. Least squares-means of standard deviation in longitudinal speed (mph) for male participants under the No-Tactors condition. Error bars represent standard error values.

7.3.3 Mental Workload (MWL) Performance Metrics: Secondary Task Response Time

Secondary task response times represent the time spent with partial attention devoted to the secondary task, with slower times associated with worse task performance. For the Tactors condition, the carryover effect was significant ($F(4,131)=2.98$, $p=.022$). Hence, a reduced dataset was run (Figure 30). Within the reduced dataset, Type of Task ($F(1,22)=7.13$, $p=.004$) significantly impacted response time for the task, with the License Plate task ($M=32.31$ s, $SD=11.79$ s) taking significantly less time to complete than the Navigation ($M=52.38$ s, $SD=20.38$ s) task and nearly significantly less time to complete than the Notifications ($M=45.23$ s, $SD=15.87$ s) task.

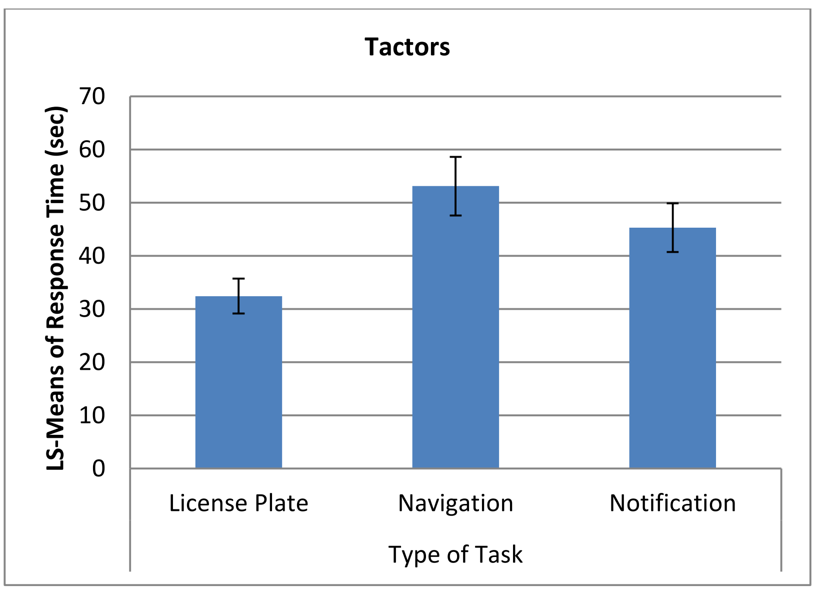

Figure 30. Least squares-means of response time to the secondary task (sec) under the Tactors condition. Error bars represent standard error values.

The carryover effect was also significant for the No-Tactors condition ($F(4,129)=2.67$, $p=.035$). Hence, a reduced dataset was run. Within the reduced dataset, the covariate Gender was significant ($F(1,18)=8.66$, $p=.004$). However, as the dataset was already reduced due to carryover effect, further splitting the dataset into Male and Female groups would have resulted in a reduction in the degrees of freedom to the point where many values were non-estimable. Hence, the reduced carryover dataset was left intact and the Gender variable was converted from a covariate into an independent variable to try to accommodate for differences due to Gender in the model.

The revised model suggests that Gender significantly impacted secondary task response time ($F(1,18)=8.66$, $p=.009$), with Males ($M=42.00$ s, $SD=17.06$ s) taking longer to respond than Females ($M=39.60$ s, $SD=18.45$ s). The interaction between Device and Input Method also significantly impacted response time ($F(1,18)=10.92$, $p=.004$), with both Device ($F(1,18)=19.39$, $p=.001$) and Input Method ($F(1,18)=15.82$, $p=.001$) having main effects as well (Figure 31). Under the Voice entry condition, performance using the Smartphone ($M=40.375$ s, $SD=16.34$ s) was significantly faster than when using the mMCT ($M=59.25$ s, $SD=11.79$ s). Without consideration for the interacting effects, evidence suggest that participants performed the task significantly faster using the Smartphone ($M=40.27$ s, $SD=17.31$ s) than using the mMCT ($M=43.08$ s, $SD=17.62$ s) and that participants performed the task significantly faster using Manual entry ($M=38.52$ s, $SD=16.96$ s) over Voice entry ($M=46.67$ s, $SD=17.15$ s). Additionally, Type of Task trends towards significance in the same fashion in the No-Tactors condition as it did in the Tactors condition ($F(2,18)=3.27$, $p=.061$), with the License Plate task ($M=35.17$ s, $SD=14.60$ s) taking nearly significantly less time to complete than the Navigation ($M=51.00$ s, $SD=20.77$ s) task.

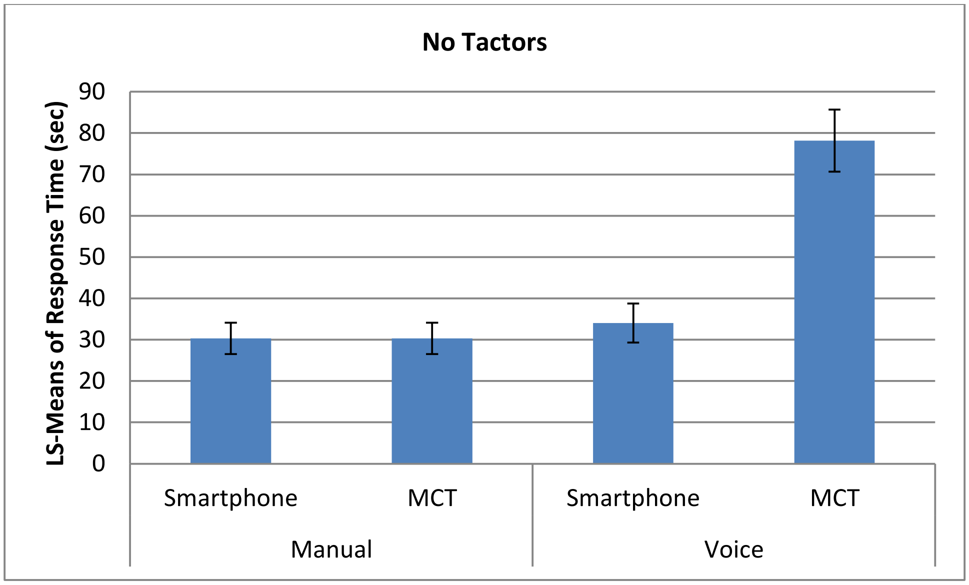

Figure 31. Least squares-means of response time to the secondary task (sec) under the No-Tactors condition. Error bars represent standard error values.

7.3.4 Mental Workload (MWL) Subjective Metrics: NASA-TLX

NASA-TLX indicates subjective perspective of MWL, with higher values indicating a higher perceived MWL. For the Tactors condition, no within-subjects variables significantly differed due to the test conditions, with *p*-values of the variables ranging between .149 and .792. For the No-Tactors condition, NASA-TLX scores were significantly impacted by Input Method ($F(1,40)=11.47$, $p=.002$) (Figure 32), with post-hoc tests indicating that Manual entry ($M=58.23$, $SD=24.46$) resulted in a higher subjective mental workload than Voice entry ($M=43.69$, $SD=21.28$).

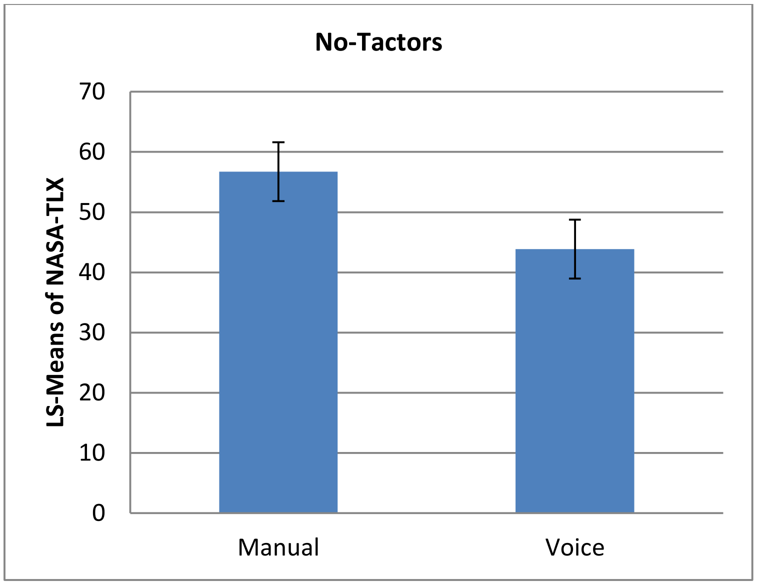

Figure 32. Least squares-means of NASA-TLX Subjective Workload Survey under the No-Tactors condition. Error bars represent standard error values.

7.3.5 Summary of Results

Table 19 summarizes the findings from this study by comparing significant differences found in the Tactors versus No-Tactors conditions. Those colors in the metrics column correspond with the cell colors of the variable condition where participants performed comparatively better. Areas shaded dark gray represent those cases where significant differences were found among levels of a variable while areas shaded light gray represent those cases where differences were approaching significance. The cells containing other letters represent those where analysis indicated a need to further breakdown the dataset and significance was then found.

Table 19: *Representation of significant findings under the Tactors and No-Tactors conditions*.

Variables / **Metrics**	Tactors			No Tactors		
	Device	*Input Method*	*Task*	*Device*	*Input Method*	*Task*
Mean of RMS Absolute Steering Rate						
Standard Deviation Lane Position					**M** F	
Average Speed Deviation from Target 50 mph						
Standard Deviation Longitudinal Speed					**M: D2** F	**M: D1,D3** F
				M: D4		
Secondary Task Response Time						
NASA-TLX						

7.4 Discussion

The results suggest that the use of Tactors supported the primary driving (i.e., speed maintenance) task but may have detracted from other driving control activities compared to the No-Tactors condition. The smaller amount of deviation from the target speed and of standard deviation in longitudinal speed when using the Tactors indicates the potential of continuously informing displays to support the primary task along at least one dimension (i.e., speed maintenance). Contrastingly, the larger Mean RMS Absolute Steering Rate and SDLP when using the Tactors likely reflects the observed desire to over rely on information from the tactors, particularly while performing the secondary task, resulting in drivers veering from the roadway and having to make larger steering adjustments. This poses the question of whether or not continuously informing vibrations can present driving information across more than one dimension; in this case, a continuously informing display that also presented information on lane deviation may have led to better driving performance via higher SA and a more effective redistribution of MWL.

Analysis suggests the continued benefit of voice over manual input, as found in Experiment 1 through the reduction in structural inference and redistribution of MWL. This was found across multiple driving performance metrics as well as the secondary task response time and subjective workload metrics. Furthermore, the potential benefits of voice over manual input are more pronounced in the No-Tactors case as drivers must allocate more visual resources to tracking the speedometer, thus using manual entry results in greater structural interference and consequently lower SA and higher MWL.

The results on the impact of screen size contrasted to those found in Experiment 2. Both the Tactors and No-Tactors conditions observed driving performance metrics where performance when using the Smartphone exceeded that of when using the mMCT. This could indicate at least one of three things:

(1) A plateau exists where increasing the screen size no longer results in performance gains and may result in performance loss;

(2) The increase in task difficulty and the increased level of interaction with the display compared to Experiment 2 impacted participant’s driving performance; or

(3) The differences in participants’ familiarity with the devices impacted performance, with greater device familiarity resulting in better performance (i.e., as participants use smartphones on a daily basis).

Number (3) is despite consistent training practices used when familiarizing participants with the mMCT.

The License Plate task took significantly less time to perform than the other two types of tasks. However, type of task did not impact any of the driving performance metrics, suggesting that the increased load from performing a secondary task alone is more influential on driving performance than the specific mental resources allocated to task completion (i.e., spatial versus categorical). This is consistent with findings from Experiment 1 and prior studies (e.g., Drews et al., 2009; Filtness et al., 2013; Tsimhoni & Green, 2001; Lyngsie, Pedersen, Stage, & Vestergaard, 2013; Yager, 2013) indicating that adding a secondary task impairs driving performance compared to baseline driving performance. These studies suggesty that engaging in a secondary task alone results in higher levels of structural and cognitive interference, thus decreasing SA and increasing MWL.

Limitations of this study involved the fidelity of the driving simulator, as in Experiment 1. However, in this case, workload from driving was designed to be consistent across the scenario, making the presentation of tasks potentially more representative of a real-world environment. Several participants noted the C-2 tactors caused a tickling feeling, but no participants ended the experiment due to discomfort from the tactors. The small map size for the Navigation task was difficult for some participants to read and may have superficially increased the response time to that task. As previously mentioned, participant's lack of prior experience with the mMCT program may have also impacted the results.

8. CONCLUSION

More than 1,060 people are injured daily in the United States in crashes involving distracted drivers, and interactions with in-vehicle technologies play a significant role in these crashes, with secondary tasks such as texting posing considerable concern (NHTSA, 2010; NHTSA, 2013). Similarly, distracted flying incidents involving the nonoperational use of Portable Electronic Devices (PEDs) have prompted the NTSB to issue a safety alert urging pilots to turn off such devices before entering the cockpit (NTSB, 2013). Moreover, secondary device interactions in both driving and general aviation can induce all three categories of distraction: visual, manual, and cognitive (NHTSA, 2010; NHTSA, 2013; Nall Report, 2015).

This dissertation discusses the impact of novel interface characteristics on performance and safety in dynamic, process control environments (i.e., driving and flying) using both situation awareness (SA) and mental workload (MWL) metrics. The goal was to determine which display characteristics have the greatest potential to reduce the level of interference between the primary and secondary tasks, which is indicated by redistributed or reduced MWL and increased SA. Figure 33 is a summary of the different interface characteristics and how they may decrease interference between the primary and secondary task. Figure 34 summarizes how each of the three experiments built on each other and highlights findings from each experiment. The following sections explore the impact of these interface characteristics in terms of the *Research Questions* posed.

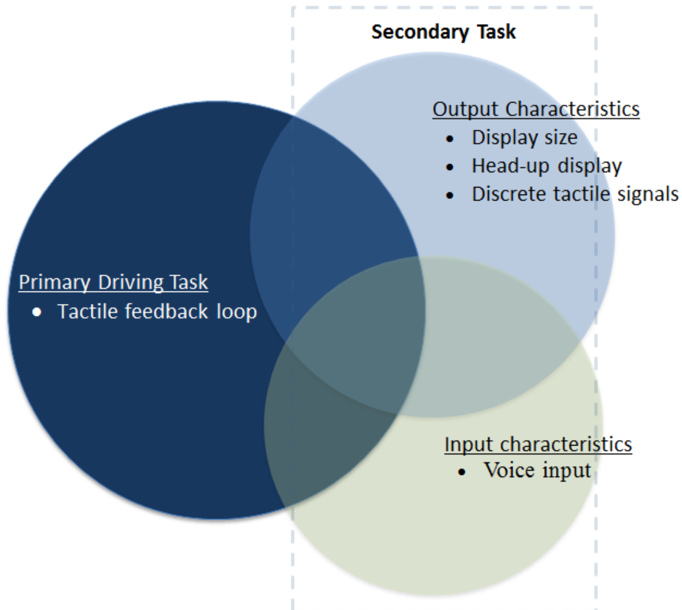

Figure 33. Representation of the different overlapping demands between performing the primary and secondary task and how different interface characteristics may reduce those overlapping demands.

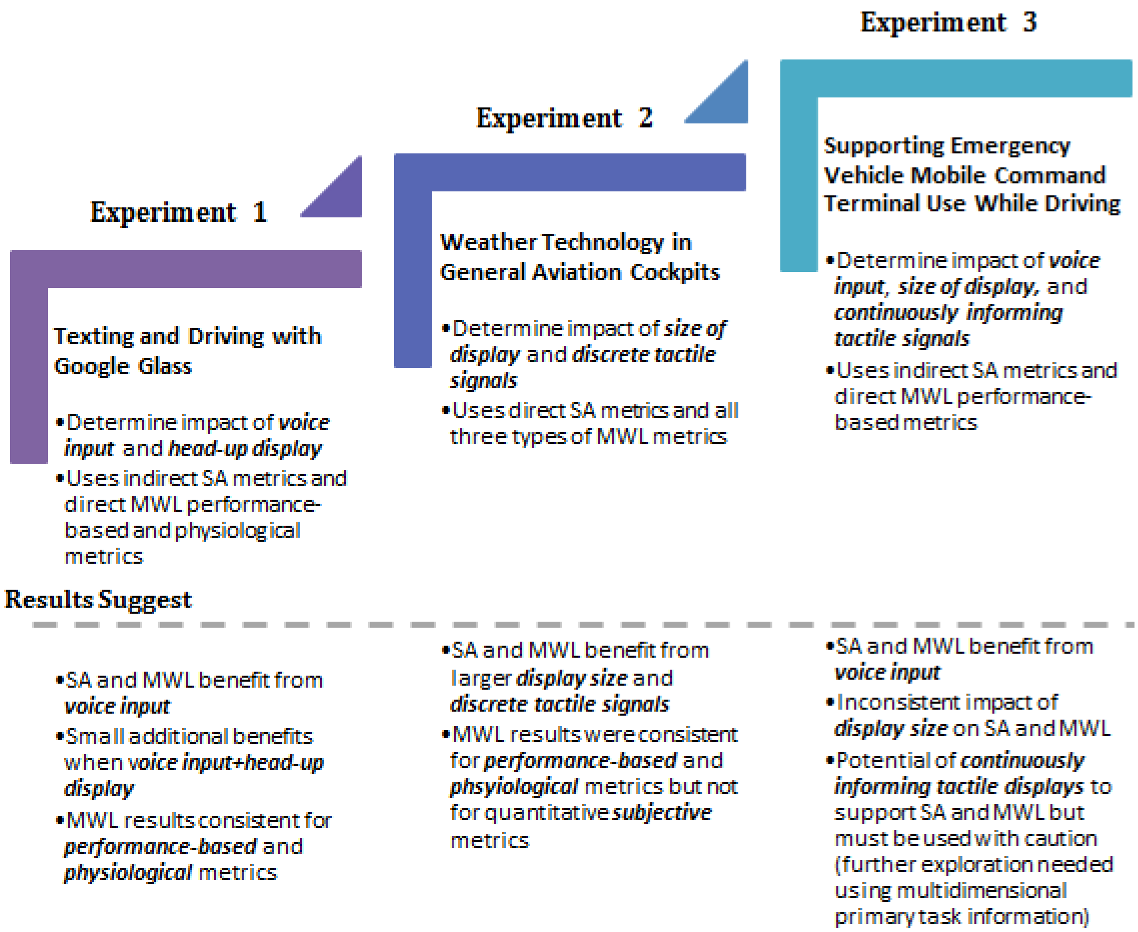

Figure 34. Summary of each of the three experiments and research findings.

8.1 Research Question 1: How Does Manipulating the Modalities Used in Presenting the Secondary Task Increase or Decrease the Mental Workload (MWL) of the Primary Task? And of the Task Set?

Analysis of mental workload (MWL) in the context of these studies is referring to the competition between the mental resources used to complete the two tasks (see Section 2.2 for full definition of MWL). The two tasks together are theorized to exceed the resource capacity of the participant and thus exceed the participant's potential level of attentional demand. This resource competition is commonly referred to within the framework of Multiple Resource Theory (Wickens 1980, 2002) and interference theory (Ivry, Diesrichson, Spencer, Hazeltine, & Semjem, 2004). The goal of these studies was to determine which interface characteristics may reduce the resource conflict between the two competing tasks, thus reducing or redistributing MWL and potentially increasing multitasking safety.

The modalities manipulated when performing the secondary task in this set of studies were input method (i.e., manual versus voice input), display size and orientation (i.e., tablet versus smartwatch and head-down versus head-up display), and use of discrete tactile signals (i.e., no vibration versus single vibration versus graded vibrations). The first research question addresses whether or not manipulating these interface characteristics results in reduced or redistributed MWL. If interface characteristics reduce or redistribute MWL, then this indicates that developers should potentially incorporate these design characteristics into new technologies entering automotive and aviation cockpits.

8.1.1 Input Method

Use of a voice input resulted in MWL benefits, regardless of whether or not changing input methods significantly impacted secondary task response time. These benefits, which were observed in Experiments 1 (Texting and Driving with Google Glass) and 3 (Supporting Emergency Vehicle Mobile Command Terminal Use While Driving), correspond with previous findings about the relationship between voice input and MWL (e.g., Tsimhoni & Green, 2001; Horrey & Wickens, 2007; Sawyer et al., 2014; Beckers et al., 2014; He et al., 2015).

The performance benefits of voice input over manual input can be in part explained via Multiple Resource Theory (Wickens, 1980; 2002), as using speech input for a secondary task reduces competition for the manual resources also used in driving, thus mitigating this potential source of structural interference (e.g., Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Hurts et al., 2011; Sawyer, Finomore, Calvo, & Hancock, 2014). In Experiment 1, the time required to verbally enter text messages was faster than manual entry and involved less focal visual reorientation during text entry and verification, resulting in fewer glances away from the road exceeding the critical safety duration of 1.6 seconds (Horrey & Wickens, 2007). A significant time difference was, however, not found when performing the mock MCT secondary tasks in Experiment 3, with a small amount of evidence suggesting voice input takes longer and thus imposes a higher MWL than manual entry, particularly in the No Tactors condition. However, the subjective workload measures indicated voice over manual input more effectively redistributed MWL.

The difference in whether or not the duration to complete the secondary task was reduced by using voice input is likely due to the differences in the complexity of the tasks and the user's familiarity with each interface prior to the experiments (e.g., Lansdown, Brook-Carter, & Kersloot, 2004). The different tasks in Experiment 3 did indeed use different processing resources and differentially impacted Standard Deviation in Longitudinal Speed (SDLS). Processing code resources are along a different dimension than modalities in Multiple Resource Theory (Wickens 1980; 2002) (see Figure 2 in Section 2.4); hence, while information may be presented in the visual and auditory modalities, a spatial processing task, for example, has a limited bandwidth across both modalities. Developers should hence account for spatial versus symbolic processing resources across modalities when designing interfaces for drivers and pilots.

8.1.2 Display Size and Orientation

Use of head-up orientation and a larger display size were both found to result in reduced MWL for both drivers and pilots. Experiments 1 (Texting and Driving with Google Glass) found evidence of the benefit of head-up displays, and Experiment 2 (Weather Technology Characteristics in General Aviation Cockpits) found evidence of benefits of increased screen size. However, Experiment 3 (Supporting Emergency Vehicle Mobile Command Terminal Use While Driving) found no evidence of a difference due to screen size. More specifically, the physiological performance data in Experiments 1 (i.e., eyes-off-road counts) and 2 (i.e., fNIRS data) both suggest a performance benefit for the use of head-up display and larger displays. This benefit likely reflects the smaller amount of visual reorientation required to switch between the primary and secondary tasks when using either of these display characteristics.

Experiment 1 quantified the amount of visual reorientation (i.e., eyes-off-road time) via eye movement analysis, showing Google Glass' HUD only required participants to change their gaze direction and did not require a more disruptive change in head or body posture to view text on the device. This finding highlights the HUD's ability to allow the tasks of driving and texting to be conducted within a more proximal visual field, which directly reduces structural interference and may indirectly reduce cognitive interference (Ivry et al., 2004) as disruptive postural changes inhibit cognitive activities and larger distances between device locations increase the effects of selective attention. The N-SEEV model supports these findings, suggesting additional theoretical underpinnings as to why the use of head-up displays in the same design category as Google Glass result in reduced attentional shifts compared to tradition head-down smartphone interactions when performing multitasking activities (Steelman-Allen, McCarley, Wickens, Sebok, & Bzotek, 2009)

While use of the Windows tablet in Experiment 2 did require participants to alter their body posture, this adjustment was less intrusive than the altered body posture taken to view the Pebble smartwatch. Building on this, the larger screen and consolidation of information within the map on the screen kept users better informed of situational details. This supposition is supported by data in the post-experiment questionnaire, which indicated pilots had a significant preference for the Windows tablet over the Pebble smartwatch.

The potential benefits of HUDs and increased screen size are, however, not without reservation. Prior studies concerned with HUD use note that affording a more "shareable" visual field with the external environment has the potential to result in attentional detriments, such as visual and cognitive capture (Tufano, 1997; Yantis & Egeth, 1999). For example, visual cues on the HUD that are salient or task-relevant, such as the onset of an icon communicating that a text message has arrived, will be more likely to lead to a reorientation of visual attention away from the roadway. The results of Experiment 1 suggest that when using Glass, this type of captured visual attention was quickly be reoriented back to the road with relatively minimal consequences to driving. The eye glance analyses indeed confirmed that when a text message arrived on Glass, the driver reoriented their attention to the screen; however, the number of safety-critical glances (i.e., those that exceed the 1.6 second threshold for increased safety risk (Horrey & Wickens, 2007)) while using Glass for messaging was significantly reduced compared to when using the Touch_Keyboard in all stages.

Experiment 2 found evidence of MWL benefits of larger screen size while Experiment 3 found no difference in MWL due to screen size. As Experiment 2 compared a Pebble smartwatch against a Windows tablet, this inherently included a larger change in visual orientation that may have confounded the results compared to those of Experiment 3, which tested an iPhone 4S smartphone against a touchscreen laptop. Alternatively, this may indicate that while the Pebble smartwatch is sufficiently small to induce a performance decrement, the iPhone 4S may be large enough to avoid such decrements. Of note, however, is that users were less familiar with operating the in-house MCT mockup than with the iPhone 4S, which may have resulted in superficially higher performance in some cases when using the iPhone 4S.

Based on evidence from all three studies, the findings suggest that developers should account for display size and potential changes in the driver's or pilot's visual and postural orientation when designing interfaces to reduce multitasking performance decrements.

8.1.3 Discrete Tactile Signals

The use of discrete tactile signals resulted in MWL benefits, which may reflect the redistribution of mental resources away from the visual channel. In Experiment 2 (Weather Technology Characteristics in General Aviation Cockpits), participants had a significantly better reaction time to ADAPs when receiving either of the vibration conditions (i.e., SingleVibe or GradedVibe), suggesting that the vibrotactile alerts attracted the pilot's attention and speeded alert identification. This attention management and alert identification aid would then result in pilots undertaking fewer visual attention shifts likely because they felt more informed about the situation and less need to repeatedly sample the weather data.

By providing notifications in an alternate modality, the discrete tactile signal offloaded the mental workload required to repeatedly visually sample the weather data and added a smaller amount of mental workload for tactile signal detection. Hence, the use of tactile signals likely reduced and redistributed the mental workload required for the overall weather task. This logic corresponds with inferences from Multiple Resource Theory (Wickens 1980; 2002).

The lack of difference between the single and graded vibration conditions is in contrast to prior research (e.g., Hameed et al., 2009; Ferris & Sarter, 2011; Lee, Hoffman, & Hayes, 2004) and indicates that in this study a single vibration was sufficient to aid pilots in determining when an attentional shift is necessary. Graded alerts are a form of Likelihood Alarm Display (LAD). LADs encode information in the signal sent to the operator, which prior studies indicate can improve multitasking performance, including attention allocation and information integration across tasks, without adding to the operator’s attentional load (Sorkin, Kantowitz, & Kantowitz, 1988). This indicates that the alert urgency that was encoded in Experiment 2 should have resulted in faster ADAPs response times. The lack of difference found may be due to the small number of alerts used in each scenario as the scenarios only contained one alert of each urgency level. Alerts were also presented with increasing urgency in both the Alaska and New Mexico scenarios, potentially prompting expectations about the alert urgency across the scenarios. Developers should hence take into account the number and urgency of alerts presented in the environment when determining whether or not to use LADs.

8.2 Research Question 2: How Does Increasing or Decreasing the MWL Impact the Situation Awareness (SA) of the Primary Task?

Analysis of situation awareness (SA), in the context of these studies, refers to how the participant perceives and strategically manages information (see Section 2.1 for full definition of SA and Section 2.3 for full explanation of the relationship between SA and MWL). This information is gained via mechanisms that increase or decrease MWL, such as resource allocation and levels of attentional demand, suggesting that a direct relationship should exist between increases and decreases in MWL and SA. Figure 35 is an adaptation of Figure 1 that highlights the direct mechanisms that influence SA and how these mechanisms relate back to MWL through information perception, attention, and strategic management.

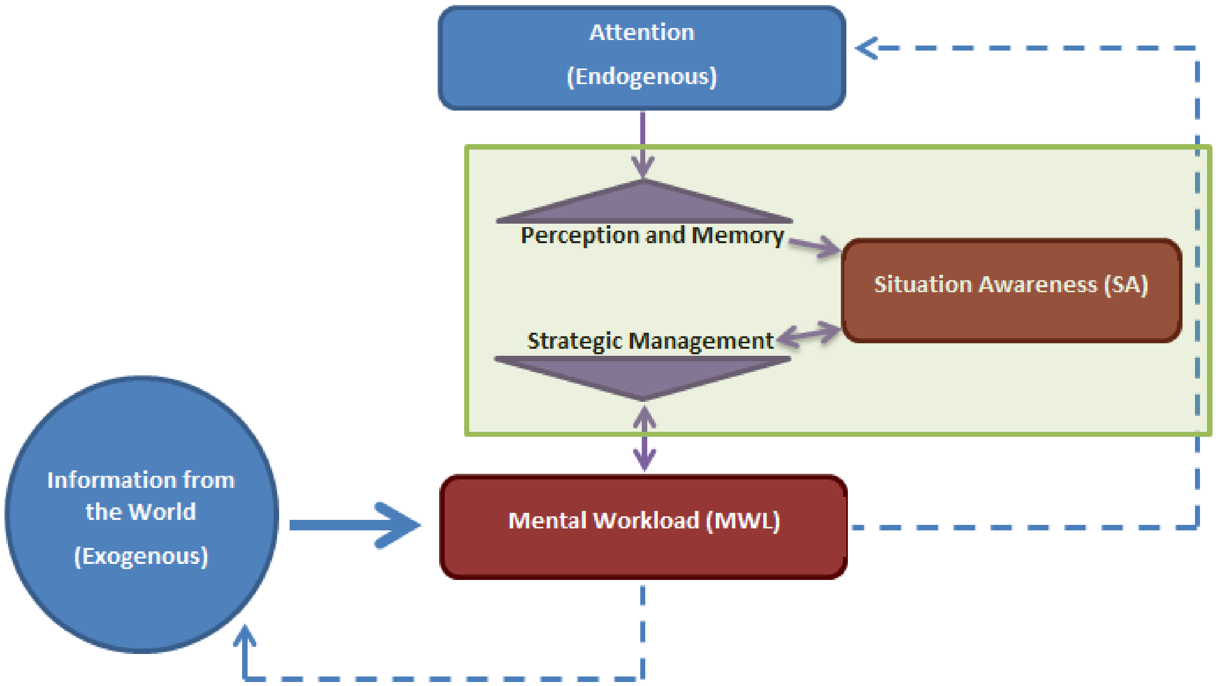

Figure 35. Representation of the relationship between SA and MWL (adapted from Tsang & Vidulich, 2006). The green box emphasizes how this relationship is directly impacted by the exogenous perception of information and the endogenous management of that information.

Again, the modalities manipulated when performing the secondary task in this study were input method (i.e., manual versus voice input), display size and orientation (i.e., tablet versus smartwatch and head-down versus head-up display), and use of discrete tactile alerts (i.e., no vibration versus single vibration versus graded vibrations). The second research question addresses whether or not manipulating these interface characteristics results in increased or decreased SA. If interface characteristics increase SA, then this indicates that developers should potentially incorporate these design characteristics into new technologies entering automotive and aviation cockpits.

8.2.1 Input Method

Use of a voice input resulted in SA benefits in cases when Tactors were not used. Experiments 1 (Texting and Driving with Google Glass) resulted in SA benefits, and Experiment 3 (Supporting Emergency Vehicle Mobile Command Terminal Use While Driving) resulted in SA benefits in conditions when the tactors were not being used.

This reflects the impact of decreased MWL found when answering Question 1. In Experiment 3, speed maintenance performance (i.e., driving) was, however, worse in the voice input than in the manual input condition when using the Tactors, indicating a SA potential bottleneck between exogenous perception and endogenous strategic management when information is being presented via both on the auditory and tactile channels. This aligns with prior research into the auditory and tactile modalities, which emphasizes how both these modalities are temporal and thus have a combined capacity limitation (Gallace, 2007; Lu et al., 2013). Developers should hence proceed with caution when designing interfaces for drivers and pilots that involve both auditory and tactile components.

8.2.2 Display Size and Orientation

Use of head-up orientation and larger display size indicate confounds in the potential benefits of screen size that contrast to the MWL findings, suggesting that the endogenous strategic management of information in the visual channel may be more limited than the capacity to perceive such exogenous information. Experiments 1 (Texting and Driving with Google Glass) and 2 (Weather Technology Characteristics in General Aviation Cockpits) found evidence of potential benefits of these interface characteristics while Experiment 3 (Supporting Emergency Vehicle Mobile Command Terminal Use While Driving) found specific cases with potential costs. Overall, however, the post-experiment surveys consistently suggested that participants preferred larger display sizes and head-up display options.

The combined head-up (HUD) and voice input display provided by Google Glass did not result in any additional SA benefits over solely using smartphone voice input. While HUDs improve the ability to share visual resources between driving and secondary tasks, drivers may still be susceptible to other perceptual and attentional phenomena, such as change blindness and inattentional blindness (Galpin, Underwood, and Crundall, 2009; Simons, 2000). Prior studies show how operators can become fixated on HUD-displayed elements and can miss major visual events even within the immediate field-of-view, especially under higher workloads (Haines, 1991; Ververs & Wickens, 1998). Moreover, despite the fact that multitasking with Glass was less detrimental than multitasking with the other devices, performance in the Glass condition was still worse than performance in the Baseline (no-texting) condition. This finding is consistent with other studies whose results indicate that the best overall driving performance occurs in driving-only, no-texting conditions (e.g., Drews et al., 2009; Filtness et al., 2013; Tsimhoni & Green, 2001; Lyngsie, Pedersen, Stage, & Vestergaard, 2013; Yager, 2013).

Experiments 2 and 3 found contradictory evidence about the impact of display size on SA, further supporting the non-generalizability across experiments when evaluating MWL. Experiment 2 found the reaction time to Situation Awareness Probes (SAPs) was better with the larger screen size; this matches the evidence from Experiment 3 in the No Tactors condition, which suggests using the larger mMCT display results in better driving performance than using the smaller iPhone 4S display. However, in the Tactors condition, driving performance using the iPhone 4S was better than when using the mMCT. This may again reflect a potential bottleneck in perceptual resources for incoming exogenous information. Alternatively, processing resources required to perform the tasks in Experiment 2 versus the tasks in Experiment 3 may have resulted in an interaction effect between the simulation environment or the device used to perform the secondary task and the processing (i.e., spatial versus symbolic) resources required to perform the different types of tasks (Ferris & Sarter, 2011; Ardoin & Ferris, 2014). This may have resulted in greater resource competition and thus increased structural and cognitive interference. The most likely explanation of this opposing evidence is, however, that participants in Experiment 3 were all familiar with the iPhone 4S prior to the start of the experiment. Both observational data and post-experiment surveys noted that the familiarity with the iPhone 4S exceeded the level of competency gained with the mMCT through training, despite participants plateauing in performance using the mMCT prior to the start of the data collection scenarios.

SA findings across the three studies provide additional evidence that developers should consider display size, orientation, and task processing resources when designing interfaces for drivers and pilots.

8.2.3 Discrete Tactile Signals

The use of discrete tactile signals in Experiment 2 (Weather Technology Characteristics in General Aviation Cockpits) resulted in similar SA benefits as found when evaluating MWL, indicating that the reduction or redistribution of MWL through the perceptual channel may have aided in the endogenous strategic management of incoming information. In contrast to previous research, the results of pilot response time to SAPs suggest that a single vibration may be more beneficial than a graded vibration; this aligns with the MWL findings and may again reflect the limited number of alerts that pilots received. The post-experiment survey corroborated these findings of potential discrete signaling benefits, indicating the participants found the vibration to be useful, particularly when associated with an alert appearing on the Windows Tablet (Kaber & Endsley, 2004). As the SA and MWL findings align, these results again suggest developers should take into account the number and urgency of alerts presented in the environment when determining whether or not to use Likelihood Alarm Displays.

8.3 Research Question 3: How Does Giving the User Additional Information for the Primary Task (in Alternate Modalities) Influence MWL of the Primary Task and of the Task Set as well as SA of the Primary Task?

Use of a continuously informing tactile displays to support speed maintenance in Experiment 3 (Supporting Emergency Vehicle Mobile Command Terminal Use While Driving) resulted in potential SA and MWL benefits between speed maintenance and the secondary task. Conversely, the continuously informing tactile display may have weakened overall driving performance by producing decrements in other driving control activities, such as maintaining lane position. As previously noted, some differences were observed in the Tactors versus No Tactors condition, indicating an interaction between the use of Tactors and the potential benefits or costs of different input methods and display sizes. Both quantitative and qualitative results from this study were, however, not sufficient to determine whether or not the benefits in supporting speed maintenance were due to the reduction or redistribution of mental resources to the tactile channel or to participants being able to “preattentively” interpret information from the tactile signal (Woods, 1995). By reducing this workload and creating trust between the user and the continuously informing tactile display, this may have prompted the user to over-rely on the display for gaining information on overall driving performance, especially when performing a secondary task. As the display only reflected one dimension of driving performance, this would have resulted in other driving performance decrements and likely decreased the response time to the secondary tasks. Alternatively, the "preattentive" interpretation implies that users did not need to consciously process (or may have automatically processed) information from the continuously informing tactile display, reducing the mental workload required to maintain a constant speed, which again Experiment 3's results support but do not confirm (Watson & Sanderson, 2004; Ferris & Sarter, 2011)).

The findings for this study strongly suggest that researchers need to further explore the use of continuously informing tactile displays in multitasking environments before making design recommendations to developers.

8.4 Summary of Findings

Table 20 overviews the findings from all three studies in the same format at Table 7 from Section 3. Table 20 thus lists each novel interface characteristics (i.e., mediating measures) addressed in this dissertation. The numbers in the cells are again the numbers of the experiments that address the proposed novel interface characteristic. Green cells indicate that experimental results suggest benefits for that interface characteristic while yellow-orange cells indicate that experimental results either cautiously suggest benefits for that interface characteristic or are confounded, presenting both positive and either negative or null results for that interface characteristic.

Table 20. *Dissertation framework as summary of findings*

	Problem States (Types of Interference)					
	Aiding Primary Task		Aiding Secondary Task			
			Not Embedded		Embedded	
Solution States (Mediating Measures)	Structural interferenc e	Cognitive interference	Structural interference	Cognitive interference	Structural interference	Cognitive interference
Voice Input			**1**		**3**	
Head-up display			**1**			
Size of Display					**2, 3**	**2**
Discrete tactile signals					**2**	**2**
Continuousl y informing tactile signals	**3**					

8.5 Limitations

Each study in this dissertation had its own limitations, most of which focused on the fidelity of the simulation environment and difficulties encountered when coding video recordings of data. The pool of participants for Experiments 1 and 3 was limited to a engineering students at Texas A&M, which are not representative of the general population. Experiment 2, however, used real pilots and was demographically representative of the overall population of General Aviation pilots. In addition, this study neither tested for nor accounted for potential device "super users" (i.e., a special group of people that can multitask without performance decrements). The overarching limitations of this set of studies were the difficulty in distinguishing which metrics best reflected SA versus MWL and difficulty in determining how to distinguish the impact of structural versus cognitive interference. Both limitations raise questions that may be addressed in future research.

8.6 Practical Applications

Knowledge of the individual and interacting roles of interface characteristics can inform designers and policymakers as they seek to support driver performance and safety in an increasingly multitask-oriented driving environment.

Based on both SA and MWL metrics, evidence suggests that voice input presented benefits for multitasking performance when the primary task has a high level of visual demand. However, developers should proceed with caution when designing interfaces that involved both auditory and tactile components as these channels have a combined limitation. Display size and the amount of re-orientation required to view the display should attempt to make the secondary task display on a more "sharable" field-of-view with the primary task display. The use of tactile displays is beneficial for effectively reorienting the operator's attention to the secondary task when necessary; however, the additional benefits previously found when using Likelihood Alarm Displays (LADs) may reflect the number and urgency of alerts presented in the environment, which developers should hence to into account. Additionally, special attention should be given to whether to tasks the equipment is being designed for uses spatial versus symbolic processing resources. Finally, evidence suggests the potential of using continuous informing displays to provide additional primary task information, but further research must be conducted before making recommendations about this design feature to developers.

8.7 Future Research

Future research in the exploration of this set of input and output interface characteristics should include more varied age groups within the experiments. In this dissertation, Experiment 2 was the only study that included a representative age range when compared to the general public. Relatedly, as additional studies should be conducted using head-up displays that focus on the impact of tasks that are more objectively beneficial to the driver than texting, such as GPS navigation tasks. Additional research also needs to be conducted to address the alternate uses of wrist-based devices in continuous, dynamic process control environments, both for tactile signaling and for determining at what point reducing the information displayed might make the technology a viable outlet for visual information. Lastly, building on the potential benefits observed when using continuously informing tactile displays for “multi-tracking” tasks (i.e., tasks that involve tracking multiple elements at once, such as speed and lane maintenance in driving), additional studies should be conducted using multi-dimensional tactile signals that give information both on speed and on lane deviation. If the representation of this information can be internally integrated by the user into one stream of information, this type of display may pose the greatest potential for offsetting the structural and cognitive interference in multitasking environments. This work should also be extended to explore supporting the primary task using other device features, such as head-up display (i.e., displays embedded in the windshield providing driving performance information) and voice input (i.e., controlling the vehicle using voice commands).

REFERENCES

Angell, L. S., Auflick, J., Austria, P. A., Kochhar, D. S., Tijerina, L., Biever, W., ... & Kiger, S. (2006). *Driver workload metrics task 2 final report* (No. HS-810 635).

Ardoin, W. J., & Ferris, T. K. (2014, September). Redundancy Gain or Cost? Performance Effects of Redundantly-encoded Tactile Displays in Multitask Scenarios. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 1711-1715). SAGE Publications.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., & Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Frontiers in Human Neuroscience, 7(871).*

Beckers, N. B., Schreiner, S., Bertrand, P., Reimer, B., Mehler, B., Munger, D., & Dobres, J. (2014). Comparing the Demands of Destination Entry using Google Glass and the Samsung Galaxy S4. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58,* 2156-2160.

Boer, E. R., Ward, N. J., Manser, M. P., Yamamura, T., & Kuge, N. (2005, June). Driver performance assessment with a car following model. In *Proceeding of the third international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 433-440).

Bogler, C., Mehnert, J., Steinbrink, J., & Haynes, J. D. (2014). Decoding vigilance with NIRS.

Bunce, S. C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K., & Onaral, B. (2011). Implementation of fNIRS for monitoring levels of expertise and mental workload. In *Foundations of augmented cognition. Directing the future of adaptive systems* (pp. 13-22). Springer Berlin Heidelberg.

Burke, E. M. (2015). *The Wearable Instrument Approach For Pilots (Master's Thesis).* Texas A&M University.

Bustamante, E. A. (2008). Implementing likelihood alarm technology in integrated aviation displays for enhancing decision making: A two-stage signal detection modeling approach. *International Journal of Applied Aviation Studies*, *8*(2), 241-262.

Byrne, E. (2015). Commentary on Endsley's "Situation Awareness Misconceptions and Misunderstandings". *Journal of Cognitive Engineering and Decision Making*, *9*(1), 84-86.

Ciftçi, K., Sankur, B., Kahya, Y. P., & Akin, A. (2008). Multilevel statistical inference from functional near-infrared spectroscopy data during stroop interference. *Biomedical Engineering, IEEE Transactions on*, *55*(9), 2212-2220.

Colle, H. A., & Hiszem, K. J. (2004). Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference. *Ergonomics*, *47*(13), 1406-1423.

Donmez, B., Boyle, L., & Lee, J. D. (2009). Designing feedback to mitigate distraction.

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text Messaging During Simulated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *17*, 762-770.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current opinion in neurobiology*, *7*(2), 255-261.

Durso, F. T., Dattel, A. R., Banbury, S., & Tremblay, S. (2004). SPAM: The real-time assessment of SA. *A cognitive approach to situation awareness: Theory and application*, *1*, 137-154.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64.

Ferris, T. K., & Sarter, N. B. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 17-26.

Ferris, T. K., & Sarter, N. (2011). Continuously informing vibrotactile displays in support of attention management and multitasking in anesthesiology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 600-611.

Fitch, G. M., Hanowski, R. J., & Guo, F. (2015). The risk of a safety-critical event associated with mobile device use in specific driving contexts. *Traffic injury prevention*, *16*(2), 124-132.

Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., ... & Dingus, T. A. (2013). *The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk* (No. DOT HS-811 757).

Flannagan, M.J., and Harrison, A.K. (1994). The effects of automobile head-up display location on younger and older drivers (UMTRI -94-22). Ann Arbor, MI, USA: The University of Michigan Transportation Research Institute.

ForeFlight. (2007). Website <http://www.foreflight.com/>.

Gallace, A., Tan, H. Z., & Spence, C. (2007). The body surface as a communication system: The state of the art after 50 years. *Presence: Teleoperators and Virtual Environments*, *16*(6), 655-676.

Galpin, A., Underwood, G., & Crundall, D. (2009). Change blindness in driving scenes. *Transportation Research Part F: Traffic psychology and behaviour*, *12*(2), 179-185.

Garmin. (2013). *D2 Portable Avionics Watch*. Website <https://buy.garmin.com/en-US/US/in-the-air/portable-gps/d2-/prod148289.html>.

Girouard, M. M., Rae, M. M., Croll, J., Callaghan, J. P., McKinnon, C. D., & Albert, W. J. (2013, September). Police Officer Discomfort and Activity Characterization During a Day Shift and a Night Shift. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1820-1824). SAGE Publications.

Green, M. (2000). "How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times. *Transportation Human Factors*, *2*(3), 195-216.

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

Gupta, R., Laghari, K., Arndt, S., Schleicher, R., Moller, S., & O'Shaughnessy, D. (2013). Using fNIRS to characterize human perception of TTS system quality, comprehension, and fluency: preliminary findings. In *International Workshop on Perceptual Quality of Systems (PQS).*

Haines, R. F. (1991). A breakdown in simultaneous information processing. In G. Obrecht & L. W. Stark (eds.), *Presbyoia Research From Molecular Biology to Visual Adaptation*, 171-175.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage Publications.

He, J., Chaparro, A., Nguyen, B., Burge, R. J., Crandall, J., Chaparro, B., Ni, R., & Cao, S. (2014). Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis & Prevention*, *72*, 287–295.

He, J., Choi, W., McCarley, J. S., Chaparro, B. S., & Wang, C. (2015). Texting while driving using Google Glass™: Promising but not distraction-free. *Accident Analysis & Prevention*, *81*, 218-229.

He, J., Ellis, J., Choi, W., & Wang, P. (2015). DRIVING WHILE READIND USING GOOGLE GLASS VERSUS USING A SMARTPHONE: WHICH IS MORE DISTRACTING TO DRIVING PERFORMANCE. *Proceedings of the Eighth International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design*. Salt Lake City, UT.

Helton, W. S., Warm, J. S., Tripp, L. D., Matthews, G., Parasuraman, R., & Hancock, P. A. (2010). Cerebral lateralization of vigilance: a function of task difficulty. *Neuropsychologia*, *48*(6), 1683-1688.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2013). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Frontiers in human neuroscience*, *7*.

Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J. K., Sassaroli, A., Tong, Y., & Fantini, S. (2007). Human-computer interaction and brain measurement using functional near-infrared spectroscopy. In *Proceedings of the ACM UIST'07 Symposium on User Interface Software and Technology*. ACM Press.

Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Horn, M., Shaer, O., ... & Jacob, R. J. K. (2009). Using brain measurement to evaluate reality based interactions. In *Proc. CHI 2009 Workshop on Challenges in the Evaluation of Usability and User Experience in Reality Based Interaction* (Vol. 19).

Hoshi, Y., Tsou, B. H., Billock, V. A., Tanosaki, M., Iguchi, Y., Shimada, M., ... & Oda, I. (2003). Spatiotemporal characteristics of hemodynamic changes in the human lateral prefrontal cortex during working memory tasks. *Neuroimage*,*20*(3), 1493-1504.

Horrey, W. J., & Wickens, C. D. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. *Transportation Research Record: Journal of the Transportation Research Board*, *2018*(1), 22-28.

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*, 67-78.

Hosking, S. G., Young, K. L., & Regan, M. A. (2009). The Effects of Text Messaging on Young Drivers. *Human Factors, 51,* 582-592.

Hurts, K., Angell, L. S., & Perez, M. A. (2011). The Distracted Driver Mechanisms, Models, and Measurement. *Reviews of Human Factors and Ergonomics*, *7*(1), 3-57.

Ivry, R., Diedrichsen, J., Spencer, R., Hazeltine, E., & Semjen, A. (2004). A cognitive neuroscience perspective on bimanual coordination and interference. In *Neuro-behavioral determinants of interlimb coordination* (pp. 259-295). Springer US.

Jeon, M., Walker, B. N., & Gable, T. M. (2014). Anger effects on driver situation awareness and driving performance. *Presence: Teleoperators and Virtual Environments*, *23*(1), 71-89.

Joslin, R. E. (2013, September). Human Factors Hazards of IPADS In General Aviation Cockpits. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 56-60). SAGE Publications.

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

Kahneman, D., "Attention and Effort," Prentice Hall, 1973.

Kawachi, I. (1997). Heart Rate Variability. The John D. and Katherine T. *MacArthur Foundation: Research, allostatic load notebook.* Retrieved from <http://www.macses.ucsf.edu/research/allostatic/heartrate.php>.

Kiefer, R.J. (1991). Effect of a head-up versus head-down digital speedometer on visual sampling behavior and speed control performance during daytime automobile driving. SAE 1991 Transactions, 100, Section 6 (Journal of Passenger Cars), pp. 82-93, SAE Paper #910111, Warrendale, PA: Society of Automotive Engineers.

Kiefer, R.J., and Gellatly, A.W. (1996). Quantifying the consequences of the "eyes-on-road" benefit attributed to head-up displays (SAE Paper #960946). Warrendale, PA, USA: Society of Automotive Engineers.

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics*, *47*(1), 91-104.

Lee, J. D., Hoffman, J. D., & Hayes, E. (2004, April). Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 65-72). ACM.

Lee, J.D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

Liu, Y. C., & Wen, M. H. (2004). Comparison of head-up display (HUD) vs. head-down display (HDD): driving performance of commercial vehicle operators in Taiwan. International Journal of Human-Computer Studies, 61(5), 679-697.

Lu, S. A., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2013). Supporting Interruption Management and Multimodal Interface Design Three Meta-Analyses of Task Performance as a Function of Interrupting Task Modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(4), 697-724.

Lyngsie, K., Pederson, M. S., Stage, J., & Vestergaard, K. F. (2013). Don't Text while Driving: The Effect of Smartphone Text Messaging on Road Safety during Simulated Driving. *Human-Computer Interaction – INTERACT 2013,* 546-563.

MacKenzie, I. S., Sellen, A., & Buxton, W. A. (1991, April). A comparison of input devices in element pointing and dragging tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 161-166). ACM.

Matthews, M., Bryant, D., Webb, R., & Harbluk, J. (2001). Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems. *Transportation Research Record: Journal of the Transportation Research Board*, (1779), 26-32.

McKendrick, R., Ayaz, H., Olmstead, R., & Parasuraman, R. (2014). Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. *Neuroimage, 85*, 1014-1026.

Menhour, L., Lechner, D., & Charara, A. (2009). Steering control based on a two-level driver model: experimental validation and robustness tests. In *Control Applications,(CCA) & Intelligent Control,(ISIC), 2009 IEEE*, 125-130.

Merzagora, A., & Izzetoglu, K. (Website 2016). FNIR FAQ. BioPac: Knowledge Base. Retrieved from <https://www.biopac.com/knowledge-base/fnir-faq/>

National Aeronautics and Space Administration (NASA). (2008). *Solstice Moon Illusion*. Science News. Retrieved from <http://science.nasa.gov/science-news/science-at-nasa/2008/16jun_moonillusion/>.

National Highway Traffic Safety Administration. (2010) Distracted driving research. Retrieved from http://www.distraction.gov/content/get-the-facts/index.html

National Highway Traffic Safety Administration. (2014). *distraction.gov*. U.S. Department of Transportation. Retrieved from http://www.distraction.gov/

National Highway Traffic Safety Administration (NHTSA), "Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices", 2013. Docket No. NHTSA–2010–0053, Document ID: NHTSA-2010-0053-0060.

National Highway Traffic Safety Administration. (2013). What is distracted driving. Retrieved from http://www.distraction.gov/content/get-the-facts/facts-and-statistics.html.

National Transportation Safety Board (NTSB). (2013). *Safety Alert: Distracting Devices? Turn Them Off!* Retrieved from <http://www.ntsb.gov/safety/safety-alerts/Documents/SA_025.pdf>.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why pilots miss the green box: How display context undermines attention capture. *The International Journal of Aviation Psychology*, *14*(1), 39-52.

Okabayashi, S., Sakata, M., Furukawa, M., and Hatada, T. (1990). A heads-up display performance in Automotive use. Society for Information Display, 3, 255-261.

Olson, R. L., Hanowski, R. J., Hickman, J. S., & Bocanegra, J. L. (2009). *Driver distraction in commercial vehicle operations* (No. FMCSA-RRR-09-042).

Owners, A., & Pilots Association. (2015). 24th Annual Joseph T. Nall Report: General Aviation Accidents in 2012.

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

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140-160.

Pashler, H. E. (1998). *Attention*. Psychology Press.

Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J., & Chang, R. (2013, April). Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proceedings* *of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 473-482). ACM.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research.*Encyclopedia of human computer interaction, 1*, 211-219.

Power, S. D., Kushki, A., & Chau, T. (2011). Towards a system-paced near-infrared spectroscopy brain–computer interface: differentiating prefrontal activity due to mental arithmetic and mental singing from the no-control state. *Journal of neural engineering*, *8*(6), 066004.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation Theory and Applications*. Mahwah, NJ: Erlbaum.

Robert. (2008). Light Flight: Free Online Aviation Courses. Retrieved on October 2, 2015 from <http://blog.chron.com/lightflight/2008/06/free-on-line-aviation-courses/>.

Rodriguez Paras, C., Yang, S., Tippey, K., Ferris, T. K. (2015). Physiological indicators of the cognitive redline. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting.* Los Angeles, CA. October 26-30, 2015.

Rosenthal, T. J. (1999). STISIM drive user's manual. Hawthorne, CA: Systems Technology.

Sawyer, B. D., Finomore, V. S., Calvo, A. A., & Hancock, P. A. (2014). Google Glass: A Driver Distraction Cause or Cure? *Human Factors, 56 (7)*, 1307-1321.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147-155.

Skin Conductance Explained. *Psychlab*. Retrieved October 5, 2015 from <http://www.psychlab.com/SC_explained.html>.

Sklar, A. E., & Sarter, N. B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*(4), 543-552.

Solovey, E. T., Lalooses, F., Chauncey, K., Weaver, D., Parasi, M., Scheutz, M., ... & Jacob, R. J. (2011). Sensing cognitive multitasking for a brain-based adaptive user interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383-392). ACM.

Sojourner, R.J., and Antin, J.F. (1990). The effects of a simulated head-up display speedometer on perceptual task performance. Human Factors, 32, 329-340.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *30*(4), 445-459.

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). 2 Cognitive Distraction While Multitasking in the Automobile. *Psychology of Learning and Motivation-Advances in Research and Theory*, *54*, 29.

Steelman-Allen, K. S., McCarley, J. S., Wickens, C., Sebok, A., & Bzostek, J. (2009, October). N-SEEV: A computational model of attention and noticing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 12, pp. 774-778). SAGE Publications.

Tests and Procedures: EEG (electroencephalogram). In *Mayo Clinic Patient Care & Health Info*. Retrieved October 5, 2015 from < http://www.mayoclinic.org/tests-procedures/eeg/basics/definition/prc-20014093>.

Tippey, K. G., Ritchey, P., & Ferris, T. K. (2015). Crossover-Repeated Measures Designs: Clarifying Common Misconceptions for a Valuable Human Factors Statistical Technique. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. *Handbook of Human Factors and Ergonomics*, *Third Edition*, 243-268.

Tsimhoni, O. & Green, P. (2001). Visual Demand of Driving and the Execution of Display-Intensive In-Vehicle Tasks. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, 45*(23), 1586-1590.

Tufano, D. R. (1997). Automotive HUDs: The overlooked safety issues. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 303-311.

Ververs, P. M. & Wickens, C. D. (1998). Head-up displays: Effect of clutter, display intensity, and display location on pilot performance. *International Journal of Aviation Psychology*, *8(4)*, 337-403.

Vidulich, M. A., & Tsang, P. S. (2015). The Confluence of Situation Awareness and Mental Workload for Adaptable Human–Machine Systems. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 95-97.

Valasek, J., Ferris, T., Brown, L., Rantz, B., & Whitehurst, G. (2015). Weather Technology in the Cockpit (WTIC) Project C: General Aviation Weather Alerting Phase 1 Final Report. Submitted to the United States Federal Aviation Administration, April.Ward, N. J. (2000). Automation of task processes: An example of intelligent transportation systems. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *10*(4), 395-408.

Watson, M., & Sanderson, P. (2004). Sonification supports eyes-free respiratory monitoring and task time-sharing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(3), 497-517.

Wickens, C. D. (2001). Workload and situation awareness. In P. A. Hancock & P. Desmond (Eds.), Stress, workload, and fatigue (pp. 442–450). Mahwah, NJ: Lawrence Erlbaum.

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

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and Performance, 8*, 239-257.

Wickens, C.D. (2002). Multiple Resources and Performance Prediction. *Theoretical Issues in Ergonomic Science*, 3(2), 159-177.

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

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 397-403.

Wickens, C.D., Hollands, J.G., Parasuraman, R., & Banbury, S. (2012). *Engineering Psychology and Human Performance (4th Edition)*. New York, NY: Pearson.

Wickens, C. D., & Horrey, W. J. (2008). Models of attention, distraction, and highway hazard avoidance. *Driver distraction. Theory, effects and mitigation*, 249-279.

Wickens, C. D., & Huey, B. M. (Eds.). (1993). *Workload Transition:: Implications for Individual and Team Performance*. National Academies Press.

Wiener, E. L., & Nagel, D. C. (Eds.). (1988). *Human factors in aviation*. Gulf Professional Publishing.

Williams, S. R., Shenasa, J., & Chapman, C. E. (1998). Time course and magnitude of movement-related gating of tactile detection in humans. I. Importance of stimulus location. *Journal of Neurophysiology*, *79*(2), 947-963.

Wittmann, M., Kiss, M., Gugg, P., Steffen, A., Finka, M., Poppela, E., & Kamiya, H. (2006). Effects of display position of a visual in-vehicle task on simulated driving. *Applied Ergonomics, 37,* 187–199.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, *38*(11), 2371-2393.

Yager, C. E. (2013). Driver Safety Impacts of Voice-to-Text Mobile Applications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57,* 1869-1873.

Yager, C., Dinakar, S., Sanagaram, M., & Ferris, T. (2015). Emergency Vehicle Operator On-Board Device Distractions. Prepared for *Intelligent Transportation Society of America (ITS America)*

Yang, S. & Ferris, T.K. (submitted). Cognitive efficiency in human-machine systems: metrics of display effectiveness for supporting multitask performance. Submitted to *Human Factors*.

Yang, S., Nevins, L., & Ferris, T. K. (2015). Investigating Redundant Multimodal Speedometer Displays for Supporting Concurrent Lane and Speed Tracking. In *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting.* Los Angeles, CA.

Yang, S., You, N. Y., & Ferris, T. K. (2013, September). Supporting drivers in concurrent lane and speed tracking tasks with novel visual, auditory, and tactile speedometer displays. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1918-1922). SAGE Publications.

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

Young, K., Lee, J. D., & Regan, M. A. (Eds.). (2008). *Driver distraction: Theory, effects, and mitigation*. CRC Press.