Supporting Monitoring and Interruption Management in Complex Domains through Graded Multimodal Notifications

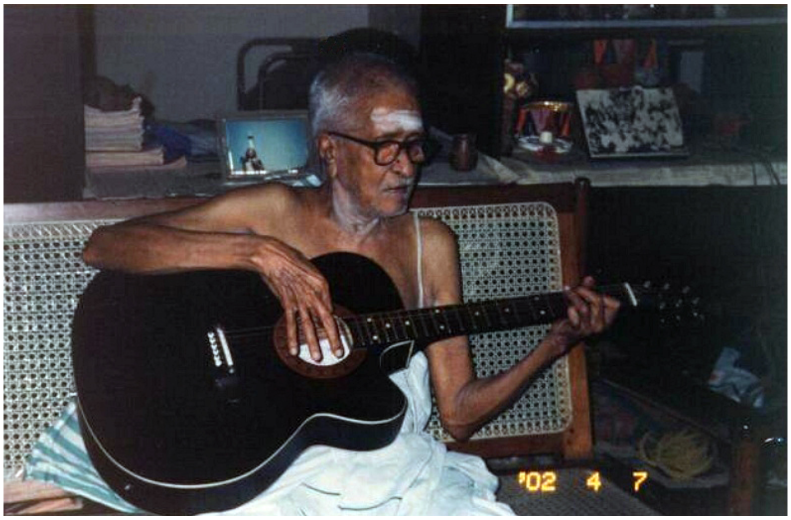
by

Swapnaa Jayaraman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2011

Doctoral Committee:

Professor Nadine B. Sarter, Chair
Professor Yili Liu
Professor Bernard J. Martin
Professor Colleen M. Seifert


'02 4 7

To *appamma*

Acknowledgements

My very first thanks go to *appa* and *amma* who have vicariously earned their doctorate! I thank them for being the most amazing parents on earth, for trusting my choices, for encouraging me to do anything (everything) I wished to, and for taking sheer pleasure in watching me do it all. Your unending love and support makes all the difference in the world.

I must thank my advisor Nadine Sarter for her patience and faith in me and for seeing my Ph.D. journey through to the very end. Nadine has been my greatest critic as well as my source of motivation. Her help in shaping my research ideas, her detailed feedback and her keen insights on my work have been invaluable. Her perseverance and hard work will continue to inspire me throughout my research career.

I would also like to thank my committee members who are all wonderful researchers and inspiring educators. I have been fortunate to be their graduate student even before they agreed to mentor me. I am grateful to Bernard Martin for being one of my first research inspirations and for persuading me to consider a career in academia. The delightful combination of his peculiar sense of humor and his passion for research and teaching and has helped me tide over some trying moments in school. I am also grateful to Yili Liu for enthusiastically sharing his worldly wisdom and experiences both concerning the subject of ergonomics and otherwise. I can only hope to someday inspire and captivate students the way he does. I also want to thank Colleen Seifert for her unrelenting support. She always made time to discuss my research over a cup of coffee anytime I asked. Her calming smile and strong presence have been a source of constant encouragement.

My graduate training and dissertation could not have been completed without the support of my funding sources. I am thankful to Thomas Finholt from the School of Information for trusting my abilities and providing me with my very first source of complete funding – a Graduate Student Research Assistant position – which helped me launch into my research path. I want to thank my department, Industrial and Operations Engineering, for supporting my research work through their generous fellowship. Funding acknowledgments would not be complete without thanking the National Science Foundation and the National Aeronautics and Space Administration for their project grants that gave me immense research latitude. I would also like to thank Rackham Graduate School, Ginsberg Center and Wallenberg Foundation for their generous fellowships which truly took my learning experience well beyond the classroom. Special thanks to the awesome individuals from Rackham, particularly to Jill McDonough and John Godfrey, who work so hard to keep up our celebrated “Michigan difference”.

I must acknowledge the support I have received from my department and my lab. My heart-felt thanks to Tina Blay for the warm teas, Michigan apples, aspirins and hugs that never failed to ward off the basement lab blues. I want to thank Tom Ferris for being an incredibly accommodating and supportive research partner, and for extending a helping hand whenever it was needed. His ability to be a dream graduate student while raising two babies and pursuing all his hobbies has always inspired me. I also thank Shameem Hameed for being an innovative lab programmer while also being a great research partner. With his technical and research skills, the sky is really the limit. Our other lab programmers Hari Molabanti, Joseph Phillips and Clif Reeder have been amazing in their abilities to quickly grasp our research requirements and create impressive simulations that resulted in several successful experiments. Many thanks in particular to Joe who graciously worked extra nights on my project to help me meet deadlines.

I must also thank some outstanding women from our lab who have contributed to my research in various ways: Wan-Ting Su for being a great friend and an invaluable research colleague, even after she moved to Taiwan; Melissa Ballard and Ashley Issa for their bright scholastic inputs; and Stephanie Widrig for being the best undergraduate lab partner one could ask for, and for repeatedly outdoing all expectations. My acknowledgements would not be complete without the mention of Huiyang Li, one of the most sincere and hardworking friends I am proud to say I have, and the rest of the gang - Richard Ho, Sara Lu, and Julie Prinet – for enlivening our existence in the basement with a constant supply of food, warmth, and cheer!

Two faculty members outside my department have been strong sources of support. I take immense pleasure in expressing my gratitude to my true *guru*, Michael Gordon, who has been my friend, philosopher, and guide. He regularly reassures me that one can pursue several interests and also combine them in a meaningful and socially relevant way. He lives his life as an example and constantly inspires me to follow my heart. I also want to thank Lalit Patil for his ready support and insightful advice every time they were sought.

My graduate life would be incomplete without my family and friends who were right by my side all the way. First thanks go to these true Wolverines, my very first roommates in Ann arbor - my sister Aneetha Jayaraman, my brother-in-law Vikalpa Jetly, and my brand new husband Shyam Venkatesan – who were there to celebrate every little milestone from my admission to my thesis defense over countless cakes, dinners and drinks. Thanks to my beloved brothers Samir Parekh, Deepak Haran, Arvind Haran, and Vasuki Seshadri for their loving long-distance encouragement. Thanks also to my grandparents, aunts, uncles, cousins and new in-laws for all the phone calls and emails bearing their best wishes at every turn. Every one of those wishes has touched my heart and inspired me to make them proud.

I also want to thank my dear friends, who were not only there when I needed them emotionally, but were also willing scapegoats for the incessant pilot experiments I conducted in my lab: Janani Mukundan for her unshakeable love for research and faith in me, Manickam Narayanan for his immense patience and unconditional love, Gaurav Rai for being a great editor and for constantly reminding me of the value of my research, Arvind Rao for his kindness and keen intellect which he shares so freely, Mekhala Raghavan for setting a great example as a researcher with a social conscience, Dinesh Krithivasan for his shining wisdom and clarity of thought, and the three stooges Sivakumar Natarajan, Saradwata Sarkar, and Balachandran Radhakrishnan for their constant presence, thoughtfulness, and all the laughter they provided during gloomy times. Special thanks to Arvind for giving me intellectually inspiring pep-talks when I needed them the most. An extra special thanks to Manickam for being the most wonderful friend and foster-parent one could ever ask for. I cannot count the number of days he has brought me food when I was working in hunger, times he has driven me to school when it was too dark or cold, and hours he has spent caring for me when I was physically or emotionally low. I give my heartfelt gratitude to him and his wife Aarthi for opening their home to me and Shyam when we needed it most. A very special mention goes to Maithreyi Gopalan Sandilya for putting up with me and my crazy lifestyle over the past seven years and for being the most special friend one could ever ask for!

For the one person without whom all this literally would not have been possible, I am at a loss of words to fully express my gratitude. Shyam has done for me all the wonderful things that the people mentioned above have done, and so much more. In all these years, not a single day has passed by without Shyam giving me his intellectual, emotional, sometimes unspoken, and perpetually dependable support. I could not have hoped to go through graduate school, or my life, with a better partner.

And now for that delectable phrase: I’m all done!

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Operators working in complex data-rich environments, such as air traffic control, need to cope with considerable and often competing attentional demands. They experience data overload in vision and audition, are required to timeshare tasks, and need to manage unexpected changes and events. Current technologies fail to support them in handling these challenges. This has led to breakdowns in performance and, in some cases, accidents. The goal of the present research was to develop novel informative types of notifications that minimize unnecessary attention switching and better assist operators in attention management. These notifications inform operators about the presence of an interruption as well as its urgency and location, thereby helping operators avoid performance costs associated with attention switching. Based on the assertion that information can be better processed in parallel if distributed across modalities, these notifications employ and combine two under-utilized modalities: touch and peripheral vision. They are graded, i.e., their salience varies over time, to reflect changes in the importance of attending to an interruption. Gradation was implemented either intra- or crossmodally (i.e., within or across peripheral vision and touch). The proposed designs were expected to improve the detection of unexpected events and the decision making about attention switching without significantly affecting performance on ongoing primary tasks. A series of studies were conducted to a) identify effective tactile notification designs, b) compare the effectiveness of peripheral visual and tactile notifications, and

c) evaluate 5 notification schemes that employ peripheral vision and touch as well as gradation. The findings from this research show significantly improved interruption management and overall task performance for all cued over uncued conditions, especially in the case of crossmodally graded notifications. They contribute to the knowledge base in multimodal information processing and display design as well as attention/ interruption management. This work goes beyond earlier studies by comparing the robustness of peripheral visual and tactile notifications under high workload and by exploring not only intra- but also crossmodal gradation in interruption cueing. At an applied level, it suggests ways in which future ATC operations can be supported more effectively to ensure the continued safety of the air transportation system.

Chapter 1 Introduction

Operators working in event-driven data-rich environments - such as aviation, process control, or military operations - need to cope with considerable often competing attentional demands. They are required to timeshare numerous tasks, process large amounts of data, and handle unexpected events at varying levels of urgency, often without proper technological support. As a result, operators struggle to manage tasks and interruptions in an effective manner. Part of the problem is that information about pending tasks and potential interruptions is typically presented in already overloaded sensory modalities (in most environments, vision and/or hearing). It tends to be presented once only and too late to allow for effective workload management and distribution. Also, the information is often not sufficient to help operators make an informed decision as to whether they should switch their attention immediately to the interruption or postpone and complete an ongoing task first. These shortcomings have been referred to as the "alarm problem" (Woods, 1995). A number of factors contribute to the alarm problem, including false alarms (alarm signals that are triggered not by an actual problem but because of a malfunctioning sensor or noise), nuisance alarms (alarm signals that are triggered by a correctly sensed off-nominal condition that does not represent a problem given the context in which it occurs), unspecified or ambiguous messages, and alarms that denote the status of the system rather than irregularities. Another problem is that alarm density is highest when the cognitive workload of the operator tends to be highest as well (Woods, 1995). During those situations, it would be of utmost importance that alarms do not distract or interrupt the operator unnecessarily. However, several incidents and accidents (e.g, Cook, et al. 1991; Moll van Charante et al. 1993; United States 1979, Federal Communications Commission 1991) as well as operational experience in a number of domains show that current context-insensitive alarm designs tend to exacerbate rather than eliminate the above problems. Nuclear power plant disasters like the Three Mile Island accident (Kemeny et al., 1979), aerospace incidents like the Apollo 12 lightning strike complications (Murray and Cox, 1990), and several medical accidents resulting in trauma or even death (Cook et al., 1991) can all be at least partially attributed to some kind of breakdown in attention and interruption management.

Woods (1995) proposed that the alarm problem can be tackled by improving a) the perceptual functions, b) the informativeness, and c) the attention directing capabilities of notifications. The purpose of the present research is to demonstrate how these goals can be achieved through the design and introduction of novel **multimodal and graded** types of notifications or interruption signals. Multimodal notifications are presented via multiple, preferably currently underutilized sensory channels – in this case: touch and peripheral vision (Oviatt, 2002; Sarter, 2002). The design of these notifications is based on the claim of Multiple Resource Theory (Wickens, 1984) that information can be better processed in parallel if it is distributed across sensory channels. Peripheral visual and tactile notifications are likely to be detected reliably without unnecessarily and unduly diverting the operator's attention from ongoing tasks (e.g., Sklar and Sarter, 1999; Hameed et al, 2006). Also, both types of cues can provide information about the location of an interrupting task or event on a display space which reduces search time once the decision is made to switch attention to the interruption (Ruz & Lupiáñez, 2002).

Graded notifications consist of a series of signals that vary over time in terms of their salience or intensity. These changes reflect the degree of urgency of dealing with an interruption (Lee et al., 2004; Sarter, 2005). Graded notifications can assist with attention management, workload distribution, and serve as reminders of pending tasks (Lee, 1992; Hess and Detweiler, 1994). They are beneficial in that they not only inform operators about the onset or presence of a task or event that may require their attention (as binary alarms do); they also provide partial information about (changes in) the urgency of dealing with the interrupting task or event. This allows operators to make better decisions about when to re-allocate their attention (Obermayer and Nugent, 2001) and avoid the performance costs associated with unnecessary or untimely task switching (Rogers and Monsell, 1995; Woods, 1995).

A series of simulation-based studies was conducted to develop and comparatively evaluate peripheral visual and tactile single-stage and graded interruption signals in terms of their ability to support interruption management and timesharing. The primary application domain for this research was Air Traffic Control (ATC): a data-rich, high-risk, and event-driven environment that imposes high cognitive and attentional demands on operators. The situation is expected to worsen due to the projected significant increase (two- to three-fold) in air traffic around the world. The expected introduction of new tasks and technologies implies that air traffic control will become even more data-rich and data-driven than it is today. To ensure and further improve the safety and efficiency of operations not only in air traffic control but also in other similar complex and event-driven domains, more effective support for attention management is needed in the form of well-designed alarms and notifications that reliably convey information about critical tasks and events without unnecessarily disrupting ongoing tasks – the focus of the present research which will contribute to the knowledge base in multimodal information processing, notification design and interruption management.

The following sections will discuss in more detail the processes and challenges involved in attention and interruption management and the associated problems with current alarm designs. Next, multimodal and graded notifications will be introduced as possible solutions to these challenges. Also, a brief introduction to the application domain, air traffic control, will be provided.

1.1 Attention and Interruption Management and the Alarm Problem

1.1.1 Attention and interruption management

William James (1890) remarked “Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought”. Attention has also been defined as the selective application of mental resources required to carry out mental operations, (Kahneman, 1973), or simply, the allocation of processing resources (Anderson, 2004).

Several forms of attention have been studied, most notably selective, focused, and divided attention (see Wickens et al., 1998). Selective attention involves attending to only one source of information out of many other possible sources. Focused attention underscores the active suppression of irrelevant signals such that only the relevant cues are processed. Divided attention is the most pertinent form of attention to the current research. It refers to the allocation of attentional resources to two or more sources of information or tasks either by switching back and forth between the sources or by attending to or performing both at the same time. The concept of divided attention is of great importance in data-rich dynamic domains in which operators are faced with a variety of potentially important and relevant sources of information at once.

To some extent, operators determine their attentional focus in an endogenous manner, i.e., driven by their needs and expectations. In other words, they engage in knowledge-driven or top-down information processing. Periodically, however, operators' attention will (need to) be captured by cues in their environment that notify them of unexpected and/or off-nominal changes and events. In those cases, they re-allocate their attentional resources in a data-driven or bottom-up fashion, which can be significantly affected by interface design. This process of attention allocation is described in Neisser's perceptual cycle (Neisser, 1976), emphasizing the interplay between top-down and bottom-up processes in *attention management*.

Attention management is concerned with allocating attentional resources to various potentially relevant and often simultaneous tasks and sources of information in a timely and efficient manner. This task is particularly challenging for operators in environments that are rife with interruptions and unpredictable events. In those domains, dynamic prioritization and rapid reorienting of attention are critical for effective attention management (Woods, 1995), i.e., for managing several threads of activities, either sequentially or in parallel.

Interruptions are known to have undesirable effects on the performance of both the interrupting and the interrupted task (e.g., Rogers and Monsell, 1995). Most notably, they induce an increase in errors on both tasks, such as skipping or repeating steps in a procedure (Gillie and Broadbent, 1989; Latorella, 1998). Performance tends to suffer especially in complex environments where operators have to attend to multiple tasks that have a high level of similarity, where interrupting tasks can be very complex, and where the timing of interruptions is not under operator control (e.g., Gillie and Broadbent, 1989; Czerwinski et al., 1991; Speier et al., 1997). Also, concerning the timing of cues, operators are known to be more interruptible when they transition between tasks or task sets (Miyata & Norman, 1986).

Latorella (1998, 1999) conducted a series of studies to explore interruption management on modern flight decks. The findings from this research led to the development of a qualitative model of interruption management, called the Interruption Management Stage Model (IMSM), which delineates the main outcomes of interruption management behavior: intentional dismissal and intentional integration on the positive side, and oblivious dismissal, unintentional dismissal and preemptive integration on the negative one.

Upon receiving a notification of a relatively unimportant interruption, an operator could actively choose to ignore it and continue performing the more critical primary task. This well-adapted behavior is known as an *intentional dismissal*. On the other hand, if the operator correctly considers the interrupting task important enough to warrant an immediate attentional shift, he/she intentionally, explicitly and strategically integrates the interruption into the ongoing task, i.e., he/she engages in *intentional integration*. These two cases of desirable interruption management can be supported through the design of informative interruption notifications that can be processed in parallel with the ongoing primary tasks so that the three undesirable outcomes of interruption management, - oblivious dismissal, unintentional dismissal and preemptive integration - can be minimized or avoided altogether,

An *oblivious dismissal* occurs when the stimulus announcing the interruption is not salient enough to be detected. As a result, the operator is unaware of the interruption and hence does not respond to it. An *unintentional dismissal* occurs when an interruption stimulus is detected but is not interpreted correctly and therefore dismissed. *Preemptive integration* of interruptions involves performing an interrupting task as soon as it is presented, even when this is not warranted, and thus preempting the ongoing task without the need to do so.

The current research will focus on two of these failures, oblivious dismissals and unintentional dismissals, since these are directly related to the quality of the interruption signal presented to the operator. For the sake of simplicity, from here on, an oblivious dismissal will be referred to as *miss* and an unintentional dismissal as a *misinterpretation*.

In addition to the failures described by Latorella, the current research includes another often observed symptom, *neglect*, or forgetting to attend to a postponed interruption.

Distinguishing between these different types of breakdowns in interruption management is difficult based on observations of overt behavior alone (see Figure 1-1) but it is critical for developing proper countermeasures. For example, an interruption signal may have been missed entirely, misinterpreted as being unimportant or less important, or interpreted correctly but postponed to complete an even more important ongoing task first. Presenting the same alarm again will be useful only in the first case. In the second case, the notification should not simply be repeated but needs to be conveyed in a clearer manner. In the third case, immediately repeating the notification is risky since it could unnecessarily disrupt ongoing tasks and functions, annoy the operator, and lead to poor performance on the primary task. Instead, sufficient time should elapse before reminding the operator of the postponed interruption. In the present research, an attempt will be made to account for and appropriately support all cases.

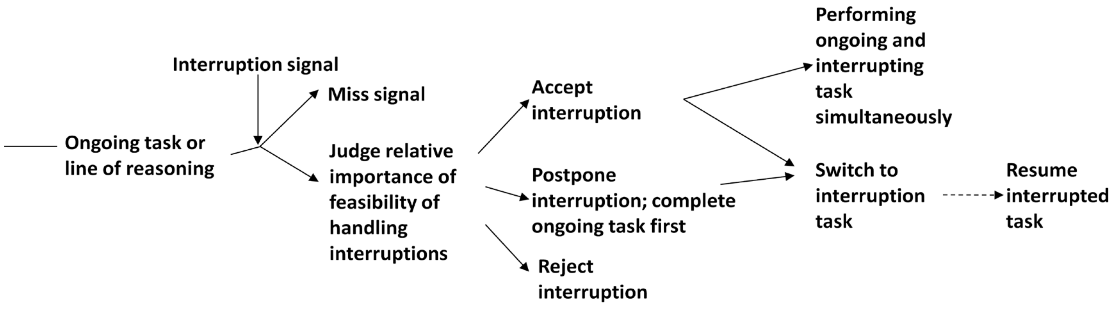


Figure 1-1: The process of interruption management (Sarter, 2005)

The process of interruption management is depicted in Figure 1-1. When an operator is presented with an interruption signal several outcomes, such as missing the signal and postponing the task, are possible. The situation can become even more complicated in very high workload conditions where an operator may face multiple simultaneous interruptions that need to be scheduled and coordinated. The figure below (Figure 1-2) outlines possible operator responses and performance outcomes in those situations. Broadly, the failures in interruption management are shown in red. These are instances when the operator will need more salient cues or repetitive reminders depending on whether the interruption stimulus was missed, misinterpreted or neglected. The desirable responses are shown in green, indicating that the interruptions are dealt with successfully. The blue instances (seen only in the decision-making stage) indicate that an action is postponed, calling for a reminder notification after a short period of time.

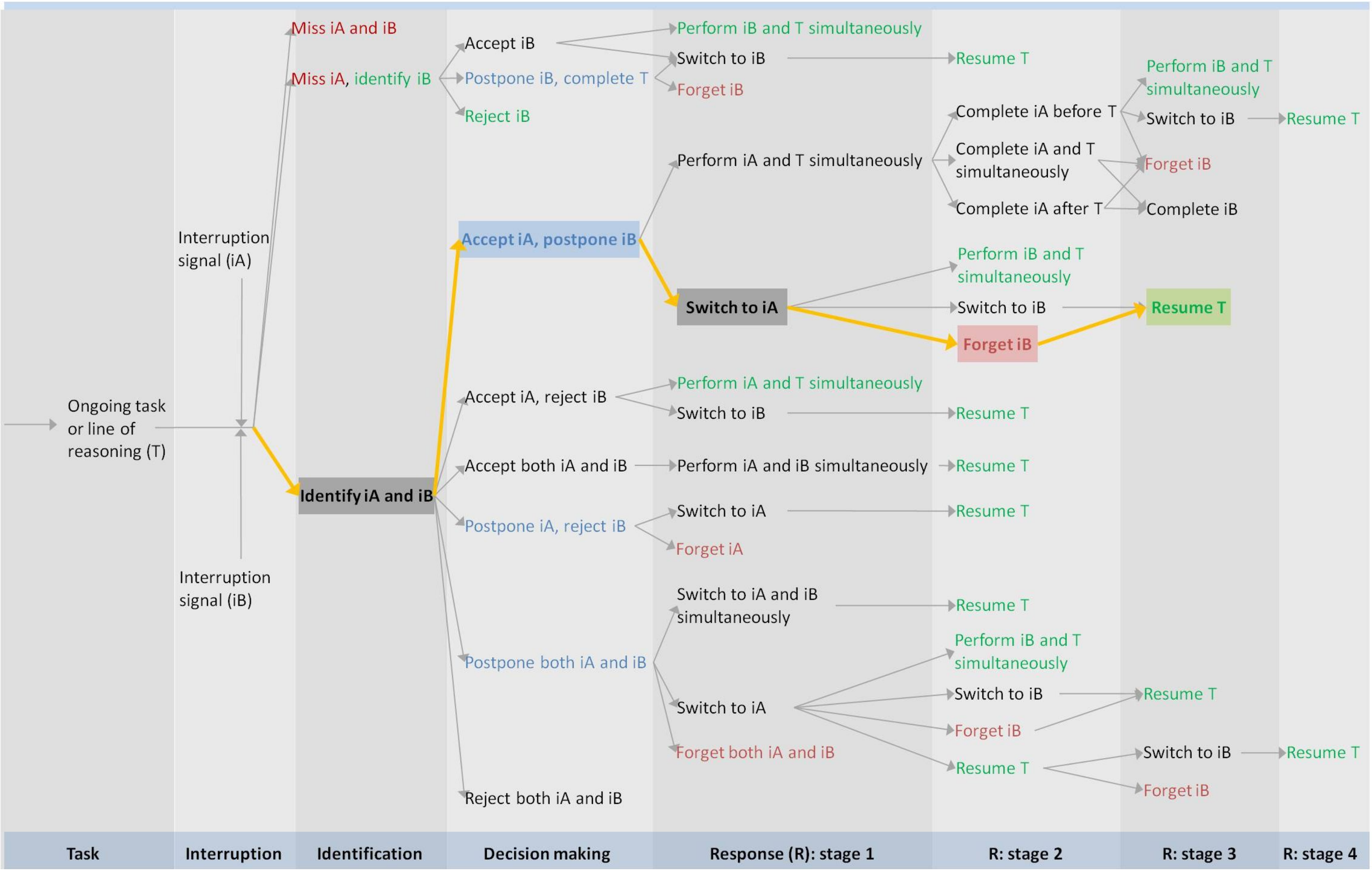


Figure 1-2: Potential outcomes of interruption management during simultaneous interruptions

In order to decide whether to timeshare and whether/when to switch from task to another, operators need to be informed about the nature and/or importance of interruptions without requiring a re-orientation of their attention in the first place and thus necessarily interrupting the performance of the ongoing task. The question is how we can resolve this apparent paradox. One possible approach is to exploit preattentive processes (e.g., Neisser, 1967; Neisser, 1976; Broadbent, 1977; Treisman, 1985; Treisman, 1986) and provide preattentive reference (detailed in section 1.1.2) by presenting at least partial information about upcoming and pending interruptions (Woods, 1995) that can be processed in parallel with ongoing tasks (Sarter, 2002; Woods, 1995).

1.1.2 Preattentive reference and the alarm problem

1.1.2.1 Preattentive processing and preattentive reference

The term preattentive processing was first used in the visual attention literature where it referred to the initial processing of the visual field to extract elementary features such as size, colors, grouping by similarity and grouping by proximity (Treisman, 1985; Treisman, 1986) and ascertain which parts of the visual field should be processed further (Broadbent, 1977). Preattentive processes are thought to be involved in orienting focal attention quickly to stimulating or interesting parts of the perceptual field (Rabbitt, 1984) and thus provide the information required for attentional selection.

The notion of preattentive processing is somewhat controversial, and the debate about the details and even the existence of preattentive processes continues even today, with ample empirical evidence both supporting and opposing the notion. Most of the debate revolves around the possibility of visual preattentive processes (e.g., Treisman, 2005; Driver, 1996; Mattingley et al, 1997; Wolfe et al.,2003) or the lack thereof (e.g., Joseph et al, 1997; Theeuwes et al, 1999). Some researchers have extended the discussion from visual information processing to tactile information processing, which will be considered as part of the present research (e.g., Deouell et al, 2000; Graham, 1997; Ohman, 1997; Zompa et al, 1995).

Given the controversial nature of the term “preattentive processing”, we will focus instead on “preattentive reference”, i.e., informative interruption signals that can be processed in parallel with ongoing activities without incurring significant performance costs if they are displayed via underutilized sensory channels. In defining preattentive reference, Woods (1995) lists the following criteria for attention-directing signals that: a) should be capable of being picked up by the operator in parallel with ongoing activities; b) should at least partially inform the observer about the interruption, without necessitating a shift of attention to itself; and c) do not require focal attention in order to be assessed.

Traditionally, alarm systems are single-stage warnings that are triggered only when the degree of threat exceeds a pre-determined threshold (Lee et al, 2004). Most alarm messages only announce that there is trouble without providing details about the anomaly or referring to the function or sub-system that is impaired (Woods, 1995). These alarms do not provide support for attention control, hence forcing the operator to shift attention away from the primary task to search for more information or to defer inspection of the alarm to a later period.

Attention-directing displays, on the other hand, support preattentive reference by providing at least partial information about the potential interruption. They can do so using various techniques. The technique of particular relevance to this research is to present information via under-utilized sensory channels so that the main foveal visual channel can still be devoted to the ongoing task. The use of modalities such as touch or peripheral vision in conjunction with traditional modalities such as foveal vision and audition is commonly known as multimodal information presentation (see section 1.2 for more on this topic). It is one possible means of supporting the process of interruption management which has been described by Latorella's (1998) IMSM model.

1.2 Multimodal Information Presentation

There are various modalities through which information can be presented: vision (comprising foveal vision and peripheral vision), audition (hearing), somatosensation (most notably, mechanoreception or touch), olfaction (smell) and gustation (taste). Among these, the two that have been employed the most to date by designers of human-machine systems are foveal vision and audition. These two channels have reached a saturation point, especially during non-routine situations when high task and cognitive loads tend to coincide with the busiest display of notifications (Woods, 1995). To offload these channels and better support time sharing and the concurrent processing of multiple sources of information, including notifications of interruptions, several authors have proposed to distribute information and tasks across additional channels, such as peripheral vision and touch (e.g., Oviatt, 2000; Sarter, 2000, 2002).

1.2.1 Multiple Resource Theory as the Basis for Multimodal Information Presentation

The main conceptual foundation for multimodal information presentation is Multiple Resource Theory (MRT) (Wickens, 1980, 1984). This framework was originally developed to explain empirical findings of improved timesharing under certain conditions. It proposes that attentional resources are not homogenous but differ along three dimensions (see Figure 1-3): 1) processing code (verbal versus spatial), 2) processing stage (perception versus cognition versus response selection/execution), 3) responses (manual spatial and vocal verbal) and 4) processing modality (vision and hearing). The extent to which tasks can be performed in parallel depends on the extent to which they draw from different attentional resources and thus avoid resource competition. This implies that one way in which parallel information processing can be supported is by distributing information across different sensory channels.

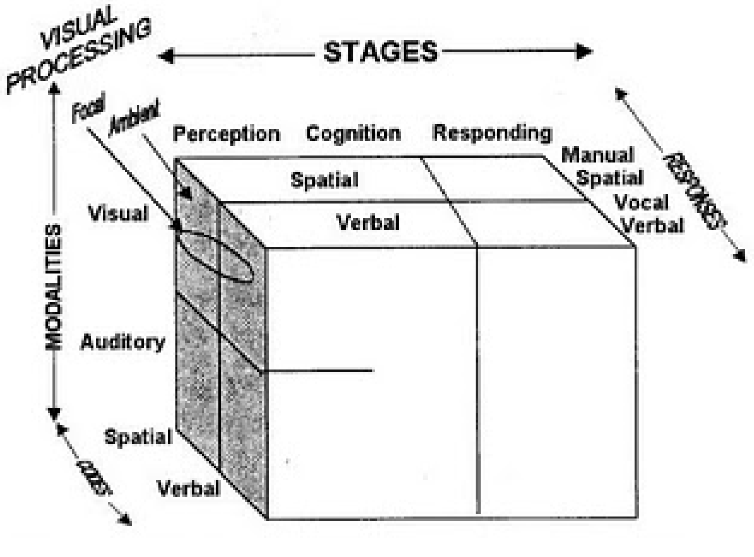


Figure 1-3: Structure of multiple resources (Wickens, 2000)

Several benefits to multimodal information presentation have been discussed, including an increased "bandwidth" of information transfer (Sarter, 2002; Oviatt, 2002), redundancy (i.e., providing the same information using different modalities), and synergy (i.e., using various modalities to present different chunks of information that need to be merged). Multimodal display systems have been shown to support attention management and time-sharing, and to improve the compatibility between the information medium and the content presented (e.g., Sklar and Sarter, 1999; Nikolic and Sarter, 2001). Clearly, multimodal information presentation is a promising means of presenting interruption notifications because operators should be able to process the information contained in these notifications while continuing to perform their ongoing tasks if they are presented via a different channel.

While it is true that information can be processed in parallel more easily if presented across different channels, modalities are not completely independent of one another. A series of highly controlled laboratory experiments by Spence and Driver (1997), Driver and Spence (1998), and Spence et al (2004) highlight that cross-modal spatial and temporal links in attention exist between vision, audition, and touch. These findings have been acknowledged by Wickens and colleagues (e.g., McCarley, et al, 2002) and resulted in modifications of MRT and the development of more complex models of multimodal processing, such as SEEV (Wickens et al., 2003) and N-SEEV (Wickens et al., 2009). Performance effects of crossmodal links have been replicated in studies employing more complex environments also (e.g., Ferris and Sarter, 2010).

Cross-modal spatial links in attention refer to the fact that the appearance of a signal in one modality in a particular location increases the readiness to perceive signals in a different modality in that same or similar location. The existence of these links contradicts early claims of modality independence by MRT and, to some extent, highlights limitations in human information processing. However, they can be exploited in the design of interfaces to guide operator attention across modalities. In the present research, for example, they will be used to guide operators' visual attention using tactile interruption notifications, thus avoiding adding additional visual information to often already cluttered displays. The location of these tactile notifications maps on to locations of interrupting events on the visual display, thus reducing search time for operators. Similarly, peripheral (as opposed to focal) visual cues will be employed to guide operators' attention to interrupting events.

1.2.2 Peripheral Vision

Peripheral visual cues such as abrupt visual onsets can be perceived in parallel with foveal visual cues and can function as powerful orientation mechanisms (Posner, 1980; Sarter, 2006). They are very effective in capturing attention in a bottom-up fashion (Jonides, 1981), making them promising candidates for supporting the reliable detection of, and guidance towards the location of, interruptions. Peripheral visual cues can be processed automatically (Muller and Rabbitt, 1989) and therefore do not differentially affect, or reduce performance on, concurrent tasks (Nikolic and Sarter, 2001). Early on, it was believed that abrupt visual onsets in the periphery attract attention rapidly and reliably in an involuntary manner (Yantis and Jonides, 1984). However, later studies showed that they do not necessarily capture attention irrespective of an observer's intentions and expectations (Yantis and Jonides, 1990). Instead, peripheral visual stimuli will be most effective in capturing attention if they match the mindset or expectations (a top-down influence) of the operator – a phenomenon referred to as 'contingent orienting' (Folk, et al., 1992).

One potential drawback of peripheral visual cues is that they are vulnerable to the phenomenon of attentional narrowing. The narrowing of the peripheral visual field occurs in response to high workload, the complexity of foveal tasks, and stressors (such as noise and fatigue (Mackworth, 1965; Easterbrook, 1959). More recent studies have indicated that the narrowing effect is not purely perceptual, but attentional as well (Hancock and Dirkin, 1983; Williams, 1988; Hancock and Hart, 2002). Therefore under high stress and workloads, it is possible that cues presented in the periphery are missed. The current research examines the robustness of peripheral visual notifications under varying workload conditions. In addition, it also explores one other channel that may not be affected by narrowing – touch – for the presentation of notifications.

1.2.3 Touch

Motivated by MRT, studies have shown that information presented via the tactile channel can be processed without interfering significantly with the performance of concurrent tasks in other modalities (e.g., Sklar and Sarter, 1999; Ho, Tan and Spence, 2005; Hameed et al, 2006; Ferris and Sarter, 2010).

There are several advantages to using the sense of touch as a channel of communication. The tactile channel is proximal and omnidirectional. It does not require a particular body or head orientation (like vision does) to be perceived. And its proximal nature makes the sense of touch an ideal channel for communicating information privately, without distracting other operators. Most importantly, tactile cues help offload the heavily used visual and auditory channels. Although the “bandwidth" of tactile information is small compared to the visual channel, it is possible to simultaneously present information about multiple aspects of an event by manipulating signal dimensions such as frequency, duration, pulse rate and location (Hameed, et al, 2006).

Tactile cues are typically applied using “tactors”, i.e., small vibrating devices similar to the ones used in pagers or cellular phones. When attached to suitable harnesses, they can be applied to a wide range of body locations such as wrists, arms, back, stomach, and thighs. This allows for the exploitation of cross-modal spatial links (e.g., Driver and Spence, 2004; Spence et al, 2004) by mapping locations in the environment to tactile signals in various body locations for the purpose of guiding visual attention.

In summary, both peripheral visual as well as tactile cues have unique advantages that can be exploited in the design of notification systems. These signals can be processed in parallel with tasks in other modalities; they are detected reliably and can convey information about tasks and events; and the existence of crossmodal links in attention make it possible to use cues in these modalities to guide operators’ focal vision to critical events. Thus, both cues have the potential to significantly improve the process of attention and interruption management.

1.2.4 Graded Notifications in Support of Interruption Management

One important consideration in the design of notifications that support interruption management is that an event (such as two aircraft slowly approaching each other) may be of relatively low importance when it is first being announced. However, as time elapses, its importance may increase (such as a pending collision) and exceed that of another ongoing task (such as assigning a new altitude to an aircraft for turbulence avoidance). This calls for the design of graded notifications that can be defined as a series of signals whose salience changes over time and is proportional to the changing degree of urgency of an interruption (Lee et al, 2004; Sarter, 2005). In contrast to traditional single-stage warnings which are presented only once and rather late in response to a preset danger threshold being crossed, the presentation of graded notifications begins earlier, as soon as an undesirable state or process begins to evolve, and before a truly dangerous condition has been reached. The early onset of graded notifications allows operators to decide whether they should switch their attention immediately to the interruption or postpone and complete an ongoing task first. It helps operators avoid the performance costs associated with unnecessary or untimely task switching (Rogers and Monsell, 1995; Woods, 1995), allows a higher degree of flexibility in scheduling tasks (Lee, 1992; Hess and Detweiler, 1994) and distribute workload. Gradations in notifications imply repetition of signals and thus also serve as reminders of pending tasks.

To date, very few studies have developed and evaluated graded notifications. For example, Sorkin et al. (1988) employed graded feedback to convey the degree of likelihood of an event and thus the need to orient attention towards that event.

Participants performed a continuous tracking task as their primary task. The task was a one-dimensional zero-order tracking task in which participants had to move a marker to follow a cursor at varying levels of difficulty and tracking interval. The secondary task was a diagnostic task requiring the detection of visual signals signifying a system failure. This task was augmented by an automated system providing visual or auditory alarms. RMS tracking error for the primary task and detection accuracy and response time for the secondary task were recorded as the dependent measures.

Two types of likelihood alarm displays informed the user if a system failure was unlikely, possible, probable or likely by altering the salience of the cues. In the visual alarm condition, the color of the cursor was altered from white (no danger), to green, yellow, and then magenta (likely danger) to indicate the likelihood of an actual system failure. In the auditory alarm condition, the danger levels were communicated verbally using a synthetic speech alarm. These two conditions were compared against 2-state alarms which featured either white or red cursor controls (for default or alarm, respectively) or speech alarms (where no message meant default and the message “check signal” called for the operator to check for an alarm). Alarm likelihood displays led to a significant reduction in response times to alarms. They improved the allocation of attention among tasks and aided decision-making without necessarily adding to the attentional load of the operator. Thus, an overall net gain in performance was achieved.

Ho and Sarter (2004), in an attempt to understand natural tendencies for modality usage, introduced the use of graded tactile cueing as one of the options for communication. In a multifaceted study exploring various aspects of information exchange between army officers, a computer-supported cooperative work environment featuring a dynamically updated shared map of the battlefield in simulated future US Army operations was used. Participants could communicate using visual, auditory and tactile means and could set up salerts for various events in these three modalities. The usage and affordances of modality integration, the utilization of the tactile modality, and the consistency of modality usage, among other measures, were studied. In the detailed analysis of tactile modality utilization, it was found that participants preferred to use the tactile modality to indicate the occurrence of critical events that required immediate attention. They also employed gradations in the tactile signals to effectively communicate the degree of urgency to each other by varying the number and the frequency of the tactile "buzzes".

Finally, Lee et al (2004) demonstrated that graded warning systems can provide greater safety margins and are received more positively than single-stage warnings. In a two-experiment driving simulation study, they used a warning system that notified participants of potential collisions based on the level of braking required – negligible, moderate and severe. The experiments compared traditional single-stage warnings and graded warnings in the auditory and tactile modalities. Auditory warnings were presented over standard speakers and tactile warnings through in-seat vibrations to the thighs. For the first experiment, the number of collisions and the adjusted minimum time to collision, or time taken for vehicles to collide at their relative position, were the dependent measures. The driver's response was measured also in terms of mean and maximum deceleration, as well as reaction time variables corresponding to acceleration, braking, and deceleration. In the second experiment, driver attitudes were measured using a series of subjective rating scales for annoyance, urgency, appropriateness, and trust in the system as well as self-confidence in detecting potential collisions.

The first experiment revealed that graded warnings led to a greater margin of safety than single-stage warnings which tended to induce a greater number of inappropriate braking responses. In the second experiment, participants also reported that graded tactile warnings were trusted more and were not more annoying than single-stage auditory ones. In general, tactile warnings were preferred to auditory warnings, a finding that deserves further investigation.

In carrying forward this line of work, the current research goes beyond earlier studies by enhancing graded notifications in two ways: 1) gradation across modalities (from peripheral visual to tactile cues) and 2) dynamic and simultaneous cueing.

1.2.4.1 Gradation across modalities (peripheral visual and tactile)

In addition to further exploring the benefits of intra-modal gradation, the current research also employs cross-modal gradation across peripheral vision and touch. Peripheral visual cues are presented first since they serve as effective attention capture and orientation mechanisms. They tend to be capable of accomplishing this function without undue distraction, even if presented at low salience levels. However, during periods of high cognitive loads, it is possible that the operator misses these cues due to attentional narrowing. Therefore, further increases in the importance of attending to the interruption are then communicated via a transition to tactile cues that are proximal and assumed to be more reliably detected as a result even during periods of high cognitive load. Combining both peripheral visual and tactile channels in the presentation of multimodal graded cues has never been attempted before but may allow for a broader and possibly more appropriate range of gradation.

1.2.4.2 Dynamic and simultaneous cueing

Earlier studies have focused on developing and testing graded notifications that inform operators of changes in the urgency of attending to a single problem or event (e.g., risk of colliding with another vehicle). However, in highly complex and dynamic environments like air traffic control, operators will likely face situations where multiple, sometimes related events of varying urgency may require their attention. For example, an aircraft deviating from its flight path can lead to reduced separation with multiple other airplanes. To evaluate the feasibility and effectiveness of graded notifications in situations faced by controllers in real-world event-driven domains, the current research explores the feasibility and effectiveness of presenting multiple graded notifications that change over time to accommodating multiple concurrent tasks/events at varying levels of urgency, and to assist with task sequencing.

1.2.5 The application domain: Air Traffic Control

Air Traffic Control (ATC) was chosen as the application domain for this research because it imposes considerable attentional demands on operators. Air traffic controllers engage in the processing of significant amounts of data and experience related difficulties with expectation-driven monitoring and monitoring for unexpected low-frequency events (Wickens et al., 1998; see Figure 1-4 for some of the many information sources in an

ATC station). Controllers are responsible for the safe and expeditious guidance of aircraft flying across a sector, a small area of airspace with horizontal and vertical boundaries. Through the analysis of visual information from the radar scope and the flight progress strips (providing data such as the type of aircraft, the intended route, altitude, speed, etc.), and based on auditory input from pilots and other controllers, the controller issues verbal instructions to pilots on the proper flight headings and altitudes to maintain separation from other aircraft, to space out arrivals and departures, meet flow control or metering restrictions, to avoid severe weather, and to remain clear of restricted flight areas (Nolan, 1994). Controllers solve four-dimensional problems (the fourth dimension being time) of varying complexity on their own or with the help of automated decision aids (such as URET – User Request Evaluation Tool) which help anticipate and detect likely conflicts between aircraft.

Figure 1-4: View of an air traffic control station showing the various sources of information

These systems use auditory and visual cues to alert controllers to potential problems. The extensive attentional demands in this domain generate workload bottlenecks and have contributed to breakdowns in attention management and resulting operational errors. Attentional demands are likely to increase further with the introduction of modern automation technologies, including electronic flight strips, more advanced position information, automated decision aids, and new tools for medium- and long-range conflict probes (Wickens et al., 1998). Also, the envisioned Next Generation Air Traffic System (NextGen; JPDO, 2007) – a completely transformed aviation system intended to accommodate the expected increase in traffic volume and complexity by 2018 is likely to create new challenges for air traffic controllers with pilots assuming more authority and autonomy in the sense that they will be allowed to change their flight path without prior approval by a controller and will be involved in traffic separation.

Support for effective attention and interruption management will therefore become even more critical in this domain (as in other domains that see more tasks and complex technologies being added). The following chapters will describe a series of 3 experiments that were conducted to 1) determine benefits of, and effective designs for, tactile interruption notifications that exploit crossmodal spatial links in attention, 2) compare how well peripheral visual versus tactile cues can support interruption management, and 3) evaluate a set of 5 notification schemes, including peripheral visual and tactile designs both with and without gradation, in terms of their benefits for supporting timesharing and interruption management.

Chapter 2 Guiding Visual Attention by Exploiting Crossmodal Spatial Links with the Sense of Touch: An Application in Air Traffic Control

The study described in this chapter was one of three studies that were conducted to ascertain the appropriateness of using multimodal notifications to support monitoring and interruption management. The current study examined how effectively crossmodal spatial links between vision and touch can be exploited to support attention allocation in the context of an air traffic control simulation. Response times and detection rates for unexpected critical events were compared between three conditions: 1) baseline with no cueing, 2) generic tactile cues that informed participants about the occurrence of a critical event and 3) location-specific tactile notifications that not only alerted but also guided visual attention to the location of the critical. The study also compares the effectiveness of providing location-specific tactile information at different levels of resolution by employing a 9- versus a 16- tactile element array. The baseline condition was expected to lead to the worst performance, followed by generic tactile cueing, and best performance with location-specific tactile notifications.

2.1 Method

Twenty-seven University of Michigan engineering students (mean age = 21.4 yrs, SD = 3.4 yrs) played the role of an air traffic controller in the context of a simplified Air Traffic Control (ATC) simulation. None of the participants had any prior air traffic control experience. Participants were instructed to monitor 40 aircrafts on a dynamic visual air traffic display for three events. A 'shape change' signified a potentially hazardous condition for a particular aircraft, which occurred with the disappearance of a wing or tail. An 'altitude change' request by an aircraft was indicated to the participants by the flashing of the aircraft icon (alternating yellow and green colors). A 'handoff' event occurred as an aircraft exited the sector boundaries of the designated airspace, indicated by the aircraft icon flashing alternating grey and green colors.

The participants responded to these events by right-clicking on the aircraft icon and selecting an appropriate action from a pop-up menu. Participants had 9 seconds to respond before the events timed out, after which the icons resumed normal shapes and color. A total of 40 visual events were presented on average once every 15 seconds. The events occurred either in isolation (single), or as two shape change events within a one second interval (double), or as an altitude change or handoff event followed by a shape change after one second (coupled). In the case of coupled events, the participants were instructed to first acknowledge the altitude change or the handoff before acknowledging the shape change. ATC audio recordings were played throughout the experimental scenarios to simulate typical background noise.

Performance results were compared between a baseline condition (no tactile cues) and the two different experimental conditions, uninformative and informative. In the uninformative condition, participants were presented with a tactile cue to their wrist which alerted them to just the occurrence of a shape change event on the visual display. This type of notification was presented using commercially available piezo-buzzers (tactors), which give a mild vibration of 150 Hz. In the informative condition, vibrotactile cues were presented to inform the participant about the occurrence and location of a shape change event. To this end, participants were wearing a tactile array on their back. The tactile array hardware is a modified version of a wireless tactile control unit developed by Jones et al. (2006). The tactile elements (tactors) were embedded in a flexible strap worn around the torso. The cues were presented to the back of the participant using an array of either 9 (3x3) or 16 (4x4) tactors. The tactors were positioned equidistant from each other, and their location mapped on to the ‘sub-sectors’ (9 or 16) of the visual display. Thus, a vibration cue in a particular location on the back would indicate a shape change occurring in the corresponding sector of the visual display.

A mixed factorial design was used, with the type of notification as the within-subject variable and the number of tactors in the informative condition as a between-subject variable. Sixteen participants wore the 9-tactor array while the other 11 participants wore the 16-tactor array. Participants experienced the various experimental conditions in randomized order.

The participants received general training on the three ATC tasks and responses. A minimum of 80% accuracy in detecting and identifying informative cues was required before the training was ended. In addition to the general training, participants also received specific training for each experimental scenario (baseline, uninformative and informative). The individual experimental scenarios each lasted 10 minutes. After undergoing general and scenario-specific training, each participant performed the three scenarios and went through a debriefing. In its entirety, the experiment took approximately 55 minutes to complete. Response times for the shape change events and response accuracy for all other events were logged. Synchronized eye-tracking data was also recorded by the ASL 6000 Series and the Remote eye tracker.

2.2 Results

2.2.1 Detection rates for shape change events

Overall, the participants detected 86.4% and 83.2% of shape changes in the 9- and 16-tactor informative cue conditions, respectively, compared to 58.8% in the uninformative condition and 50.2% in the baseline (No Cue) condition, as shown in Figure 2-1.

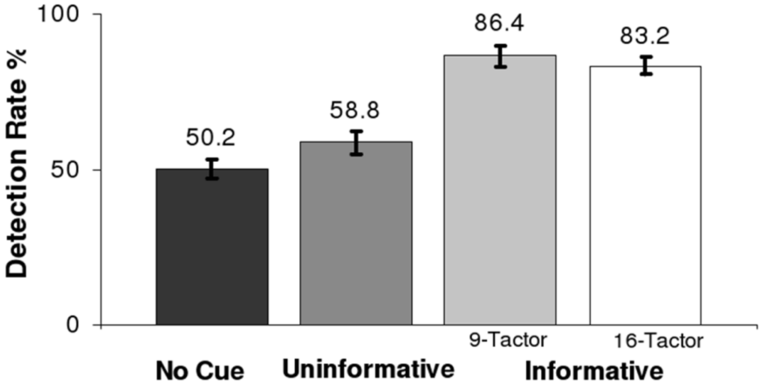


Figure 2-1: Detection rate for shape change

The detection rates for the shape changes were analyzed using paired t-test analysis to compare the mean rates of each participant. The detection rate in the informative conditions was significantly higher than in the uninformative condition (58.8%, t (21) = 7.25, $p < 0.001$) and the baseline (no-cue) condition (50.2, t (21) = 8.79, $p < 0.001$). There was no significant difference between the uninformative condition and the baseline condition.

As described earlier, the shape change events occurred either in isolation (single) or accompanied by a second shape change (doubled) or an altitude change or handoff event (coupled). This manipulation was introduced to examine performance under multitasking conditions with similar tasks (double) and dissimilar tasks (coupled). A paired t-test analysis revealed no significant differences between the baseline (no-cue) and uninformative condition in response to the single and coupled events. In the case of doubled event, there was a significant difference between the baseline (no-cue) condition and the uninformative condition (48.4% vs. 59.2%, t (26) = 3.010, p = 0.006). The informative condition was found to result in significantly higher detection rates than the uninformative condition for all three types of events (for 'single' events – 94.9% vs. 68.4%, t (26) = 6.819, p =0.000; for 'doubled' events – 80.8% vs. 59.2%, t (26) = 3.054, p = 0.005; for 'coupled' events – 82.9% vs. 49.1%, t (26) = 3.951, p = 0.001). The detection rates for shape changes under these conditions are shown in Figure 2-2.

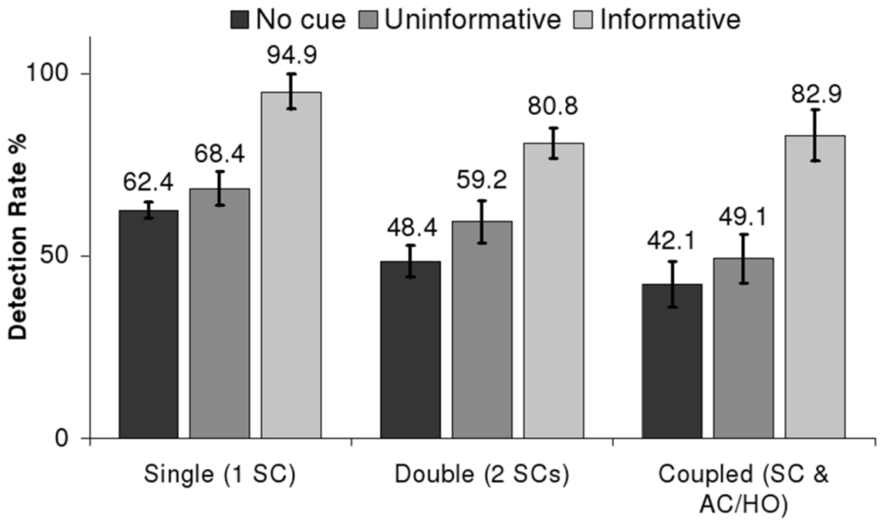


Figure 2-2: Detection rates for shape change events classified by event types

2.2.2 Response times to shape changes

The response times to shape changes improved significantly with informative cues about the target location (see Figure 2-3). A paired t-test analysis revealed statistical significance between the response times for the informative conditions (4.8s) versus the uninformative condition (5.4s, t (21) = 6.328, $p < 0.001$) and the baseline (no-cue) condition (5.4s, t (21) = 4.744, $p < 0.001$). There was no significant difference between the uninformative and the baseline (no-cue) condition.

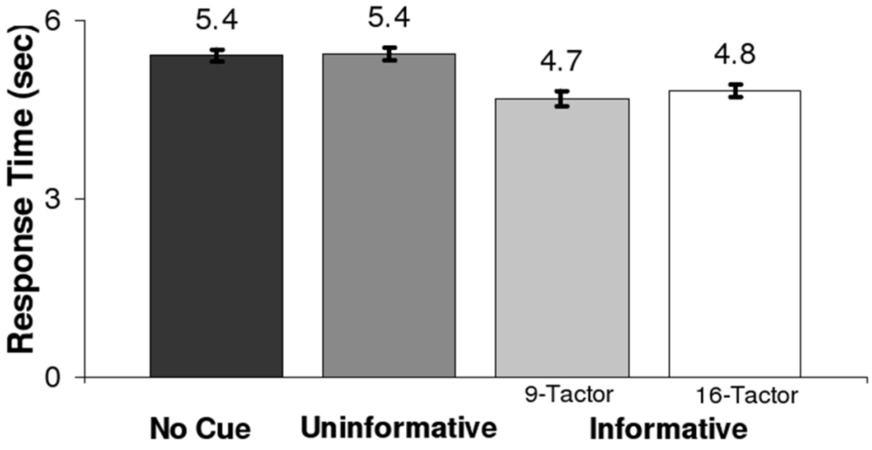


Figure 2-3: Response time for shape change events

2.2.3 ATC task performance

Performance on handoffs and altitude changes did not differ significantly between the various cue conditions. The completion rates for handoffs and altitude changes are as shown in Table 2-1below.

Conditions	**Handoff**	**Altitude change**
No cue	93.70%	79.80%
Uninformative	94.20%	85.50%
Informative	88.60%	76.40%

Table 2-1: Successful completion rates for handoffs and altitude changes

2.3 Discussion

This research examined the potential for exploiting crossmodal spatial links for providing effective attentional guidance in data-rich event-driven domains. Overall, the tactile cues lead to a 43% increase in the detection rate for shape changes, compared to the baseline condition. Between both tactile notification conditions, the informative location-specific cues produced an additional 44% increase in detection rate over the generic tactile notifications. This improvement holds up under both the single change event and the multi-change events (doubled and coupled). The informative tactile cues also lead to significantly faster response times to shape changes. Another important finding is that these benefits were achieved without performance costs to the other ATC tasks (handoffs and altitude changes). These findings highlight the promise of using tactile cues for guiding visual attention in environments that involve a high degree of competing visual demands and data overload.

Generic tactile notifications resulted in improved detection rates but did not affect response times. This finding confirms that general alerts are useful but not sufficient, especially in domains such as air traffic control where decisions and responses can be time critical. In contrast, in the case of informative tactile cues, the participants were provided with location information, and hence the area for visual search may have been reduced to 1/9th or 1/16th, respectively, of the original scan area which can explain the significantly reduced response times.

Increasing the number of tactors from 9 to 16, and thus attempting to increase spatial resolution, did not result in a performance benefit. In fact, the response time was increased by 0.1s for the 16-tactor condition. An explanation of this finding could be that the area occupied by the 16 tactors remained unchanged from that of the 9-tactor condition. This may have made it more difficult for participants to determine with any certainty the location of the tactile cue, and thus the corresponding sub-sector on the screen, based on an absolute judgment.

Overall, the findings from this study highlight the promise of using tactile cuing for providing attentional guidance in a variety of event- driven domains involving competing visual attentional demands. They informed the notification design for the final experiment in this line of research. As a next step in this line of research, a study was conducted to compare the effectiveness of peripheral visual and tactile cues for supporting interruption management.

Chapter 3 Informative Peripheral Visual and Tactile Cues in Support of Interruption Management

As discussed in previous chapters, there is a need for developing tools that help operators avoid unnecessary interruptions and resulting breakdowns in task performance and management. Given that operators' foveal visual and auditory channels are heavily taxed in most real-world domains, the peripheral visual channel and touch seem to be particularly promising candidates for presenting interruption cues (e.g., Hopp-Levine, Smith, Clegg, & Heggestad, 2006; Nikolic & Sarter, 2001; Sarter, 2002). These channels allow for the identification of a limited set of characteristic features (such as the brightness of visual cues or the frequency of vibrotactile cues) with minimal processing effort. Peripheral visual cues can be perceived in parallel with foveal visual cues and can serve as effective orientation mechanisms (Posner, 1980). The proximal nature of tactile cues allows for reliable attention capture, regardless of head or body orientation, and they are not as disruptive as auditory cues (e.g., Jones & Sarter, 2008; Sarter, 2002, 2007; Sklar & Sarter, 1999).

The present study evaluated and compared the effectiveness of informative peripheral visual and tactile cues for supporting task and interruption management in the context of a simulated supervisory control task. More specifically, this research sought to address the following questions: (a) How reliably and accurately can participants detect and interpret peripheral visual and tactile interruption cues while performing a demanding visual task? (b) Do participants make correct decisions about attention switching based on the information encoded in the interruption cues? and (c) How much does the presentation of interruption cues affect performance on the ongoing visual task?

Detection rates were expected to be higher for peripheral visual and tactile interruption cues than for the existing notification scheme, which involves a small embedded visual cue only. The modalities and eccentricities of the experimental cues were chosen to support parallel processing with the ongoing visual task. Because the experimental cues were more clearly separated from other display components than the embedded cue of the existing scheme, they were also somewhat more salient. The interpretation accuracy for experimental cues was expected to depend heavily on the encoding method (i.e., the proper mapping of content-to-signal parameter).

Information regarding the importance of the interrupting task was anticipated to have the strongest effect on task-switching behavior, both intrinsically and because participants were provided with decision rules for this parameter. Finally, ongoing visual task performance in the two experimental conditions was expected to suffer the least in the tactile cue conditions, given that peripheral visual cues may still interfere slightly with foveal visual task processing.

3.1 Method

3.1.1 Participants

Thirty students from the College of Engineering at the University of Michigan participated in the experiment. Their average age was 25 years (SD = 3 years). All participants were right-handed. Participation was voluntary, and participants were compensated $30 for approximately 90 min of their time.

3.1.2 Tasks

The participants played the role of supervisory controller of a simulated water control system in space shuttle operations. They completed two tasks that were representative of the cognitive demands encountered in this domain. In particular, they were responsible for (a) a continuous arithmetic task (a representation of the controller's ongoing visual-manual task demands) and (b) a discrete interrupting task (handling deviations from the desired water level in the system).

Arithmetic task. Participants evaluated the correctness of a series of simple-integer arithmetic equations, which were presented visually. The four components of each equation appeared one at a time in the following sequence: first, a two-digit operand for 1 s; then an arithmetic operator (+ or -) for 1 s; the second two-digit operand for 2 s; and, finally, an integer solution (which was either the correct or an incorrect sum or difference) for 2 s. Participants had another 4 s to make a response. The full arithmetic equation display and the required button response constituted one experimental trial lasting 10 s. Participants indicated the correctness of the solution by clicking on the appropriate button (✓ when correct, × when incorrect) at the bottom of the screen. A progress bar below the response buttons displayed the time remaining for each trial (Figure 3-1), and a response error was recorded if time ran out.

Interrupting task. During 50% of arithmetic task trials, participants were notified of an interrupting water control task. They were instructed to use the information encoded in a peripheral visual or tactile notification signal to decide whether, and when, to switch their attention to the interrupting task. The interrupting task, once initiated, replaced the arithmetic task on the screen and required the participant to control the flow of water by clicking on a valve symbol (and thus opening the valve) until a preset water level was reached. With continuous water flow, the water level reached the required height in the holding tank in either 5 s (for the short duration tasks) or 10 s (long duration). Once the task was complete, participants could return to the next arithmetic trial.

3.1.3 Notifications

Three different types of notifications were used to notify participants of a pending water control task: a baseline “uninformative” visual notification and two types of “informative” notifications, presented via either peripheral vision or the tactile modality. The notifications were presented either at the beginning of the arithmetic task trial (coinciding with presentation of the first operand) or 4 s into the trial (coinciding with presentation of the solution). The baseline visual notification was modeled after the existing notifications used in the Distributed Collaboration and Integration (DCI) system (Martin et al., 2003). It consisted of an exclamation mark that appeared over a clock icon in the controller display to announce a pending interrupting task without providing any additional information about the nature of the task (Figure 3-1). The exclamation mark remained superimposed on the clock icon for 6 s before disappearing.

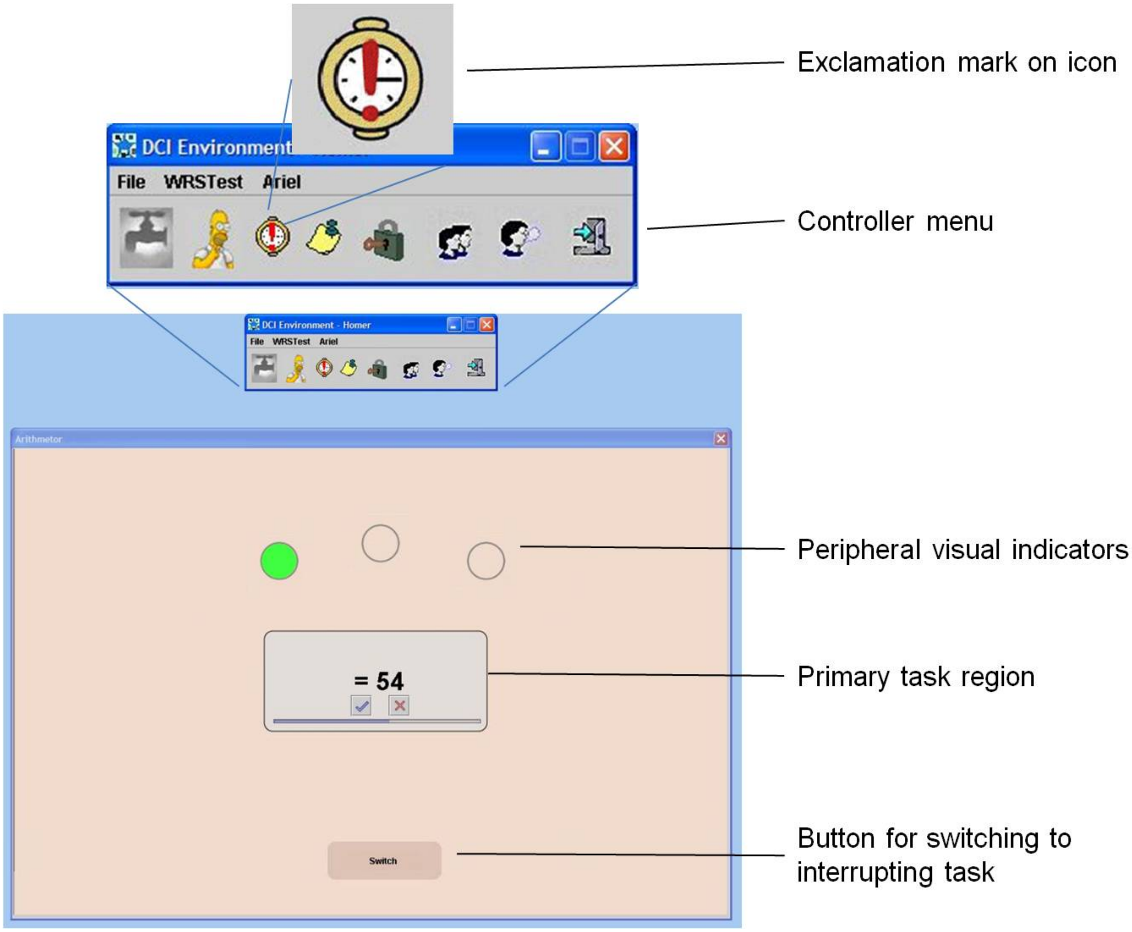


Figure 3-1: Design and location of baseline visual and peripheral visual interruption cues as well as arithmetic task.

The informative peripheral visual notifications were presented in the form of three circles above and approximately 2.5 inches radial distance (6° of visual angle) from the center of the arithmetic task display area. A notification involved one of the circles flashing bright green at a particular frequency for a particular duration, depending on the characteristics of the interrupting task (for details, see the encoding scheme section).

The informative tactile notifications consisted of vibrations that were presented to the participants' fingers using three "tactors": 1" × 1/2" × 1/4" piezo-electric devices (Audiological Engineering Corporation; http://www.tactaid.com/) that were affixed to a glove. The tactors were positioned directly over the proximal phalanges of the index, middle, and ring fingers. To ensure a consistent spatial mapping, participants were instructed always to rest their hands palms down on the table. Pilot testing was used to select 150 Hz as the operating vibration frequency of the tactors, which is easily perceptible but not uncomfortable. Cue presentation involved periodic activation and deactivation (a regular series of pulses) of one of the three tactors with a specific pulse frequency for a specific duration (for details, see the encoding scheme section).

Encoding scheme. Three types of information about the interrupting task were encoded in the peripheral visual and tactile cues. The *domain* of the interrupting task referred to which one of three water control subsystems required an intervention. This information was communicated by cue location (left, middle, or right). The *duration* of the event referred to the amount of time required to complete the interrupting task (short or long), and it mapped to the duration of cue presentation. Cues were presented for approximately

3 s to communicate a short task duration and approximately 6 s to communicate a long task duration.

The *importance* of the interrupting task (low, medium, and high) was communicated through the characteristic frequency of the cue (flashing of the green circle or vibration pulse frequency). Low task importance was represented by a low pulse frequency (approximately 1 pulse/sec), whereas medium and high importance levels were represented by higher cue frequencies (5 pulses/sec and 15 pulses/sec, respectively). Extensive pilot studies suggested that these cue parameters mapped well onto the relevant information regarding the interrupting task.

Instructions. In the experimental conditions, participants were instructed to immediately switch attention for high-importance interrupting tasks, use their discretion for medium-importance tasks, and never switch for low-importance tasks. No explicit switching instructions were given for tasks of different durations; however, participants were informed that switching for long-duration (as opposed to short) interrupting tasks would have a higher cost on performance in the arithmetic task because the equation display would be hidden for a longer period.

Although noting the domain of the interrupting event was important to addressing the correct subsystem, we did not expect the domain information to affect switching decisions. This information was included primarily to test whether participants were able to extract several dimensions of the signal reliably. In the baseline condition, because no interruption information was encoded in the signal, participants were asked only to note whether a cue was present or not—no task switching was required.

3.1.4 Procedure

After giving written consent to participate in the study, participants were familiarized with the interface, tasks, and required responses during a general training session. Three 15-min experimental sessions followed, each presenting one of the three interruption cues (baseline visual, informative peripheral visual, and informative tactile) in a randomized order. There were 27 arithmetic task trials in the baseline condition and 72 trials each in the informative peripheral visual and tactile cue conditions. Participants completed session-specific training prior to each experimental session. Short breaks (approximately 3 min) between sessions served to minimize fatigue. The entire experiment took a total of approximately 90 min to complete.

3.1.5 Data Collection

Accuracy on the arithmetic task was measured as the percentage of trials in which correctness of the equation was judged accurately. The simulation software discriminated between response errors (which included inaccurate responses and failures to respond within the time limit) and consequential misses (which occurred when participants switched to the interrupting task and did not have a chance to judge the equation). This distinction was important, as misses that were the result of correct decisions to switch to the interrupting task should not reflect negatively on arithmetic task performance.

Following each arithmetic task trial, participants were asked to complete a short on-screen questionnaire that required them to indicate whether they detected a cue and, if so, to identify the perceived domain, duration, and importance of the interrupting task (for the informative cues). Responses were used to calculate detection rates and interpretation accuracy for each of the three parameters. The questionnaire served to distinguish missed or misinterpreted cues from deliberately ignored interruptions. The participants' decision to switch or not switch to the interrupting task was also recorded by the software.

3.1.6 Experimental Design

The study employed an unbalanced nested design. The main independent variable was cue presence (whether or not a cue was presented during an arithmetic task trial). Within cued trials, the type of interruption cue was varied (baseline visual, informative peripheral visual, and informative tactile), as was the timing of the cue with respect to the onset of the arithmetic equation display (at the beginning or delayed by 4 s). Within the informative cuing conditions (peripheral visual and tactile), three task parameters were varied: interrupting task domain (left, middle, right), duration (short, long), and importance (low, medium, high).

The recorded performance measure for the arithmetic task was the percentage of trials for which an accurate response was given. The dependent measures regarding interruption cuing were detection rate (percentage "detected" responses in the on-screen questionnaire following cue presentation) and interpretation accuracy for each experimental cue parameter (domain, duration, and importance). Finally, the accuracy of participants' task-switching decisions was also recorded in the experimental conditions.

3.2 Results

The data were analyzed in repeated-measures ANOVAs, and Bonferroni corrections were used in post hoc analyses (Bonferroni, 1936).

3.2.1 Cue Detection Rate

A significant difference was found between the percentage of cues detected in each cue condition, F(2, 58) = 9.27, p = .0003. Post hoc pair-wise comparisons (with Bonferroni correction) revealed that the baseline visual cues showed lower detection rates (83.3%) than for both the peripheral visual (99.2%), t(58) = 3.73, p = .0004, and tactile cues (99.2%), t(58) = 3.73, p = .0004. No difference in detection rate was found between peripheral visual and tactile cues (Figure 3-2).

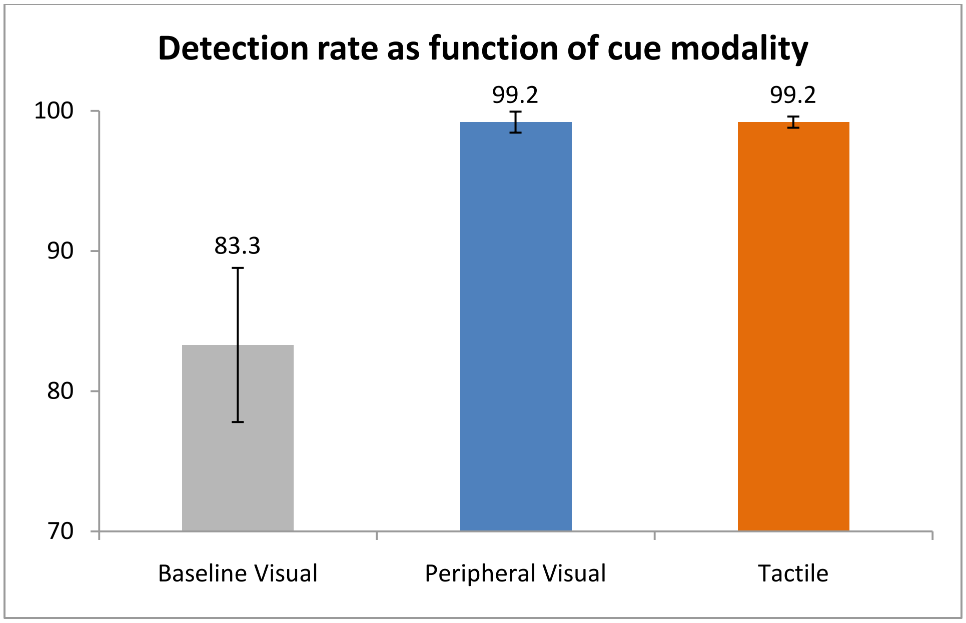


Figure 3-2: Detection rate as a function of cue modality.

3.2.2 Interpretation Accuracy

The complete set of information encoded in the informative interruption cues (domain, importance, and duration) was correctly interpreted in 70.7% of all cue presentations. The interpretation accuracy for each individual cue parameter differed significantly, F(2, 58) = 28.65, p < .0001 (Figure 3-3), with the highest accuracy observed for the identification of domain (95.1%), followed by event importance (88.2%) and duration (83.1%).

Between the two informative cue types, the likelihood of correctly interpreting all three encoded parameters in a cue did not differ. For interpretation of domain information, peripheral visual cues showed a higher interpretation accuracy (96.5%) than tactile cues

(93.7%), F(1, 29) = 8.21, p = .0077. No significant difference was found for the interpretation of task importance or duration (Figure 3-3).

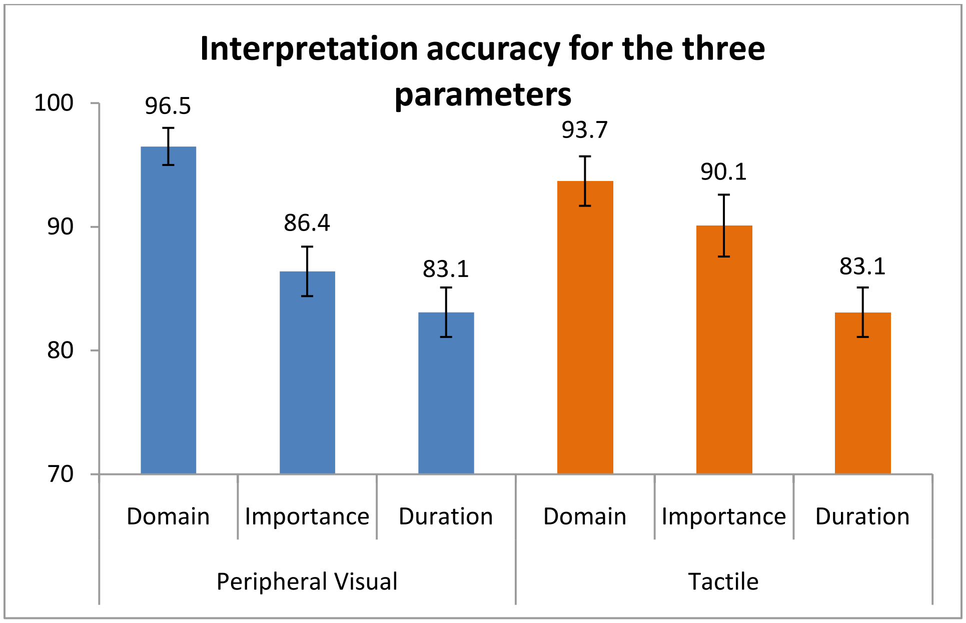


Figure 3-3: Overall interpretation accuracy for each parameter encoded in informative cues.

The timing of the cue with respect to the arithmetic task trials had a significant effect on the interpretation of duration information, F(1, 29) = 21.862, p < .0001. The interrupting task duration was interpreted with 79.9% accuracy when the cue was presented at the beginning as opposed to 86.4% accuracy when the cue was presented in the middle of a trial. The cue timing did not have an effect on interpretation of the other two cue parameters.

3.2.3 Task-Switching Behavior

For the analysis of task-switching behavior, data from one participant were removed (misinterpreted switching instructions). In general, participants based their switching decisions on task importance, F(2, 56) = 116.60, p < .0001, with 93.1%, 46.9%, and 1.3% switching rates for high-, medium-, and low-importance interruptions, respectively (see Figure 3-4). Switching errors are defined as trials in which participants correctly identified the importance of an interrupting task but failed to apply switching rules (always switch for high-importance cues and never switch for low-importance cues). Switching error rates were significantly higher for high-importance cues (6.9%) than for low-importance cues (1.3%), F(1, 28) = 6.29, p = .0182. No specific switching rules were given for cues with medium importance; therefore, switching error data were not collected for these cues.

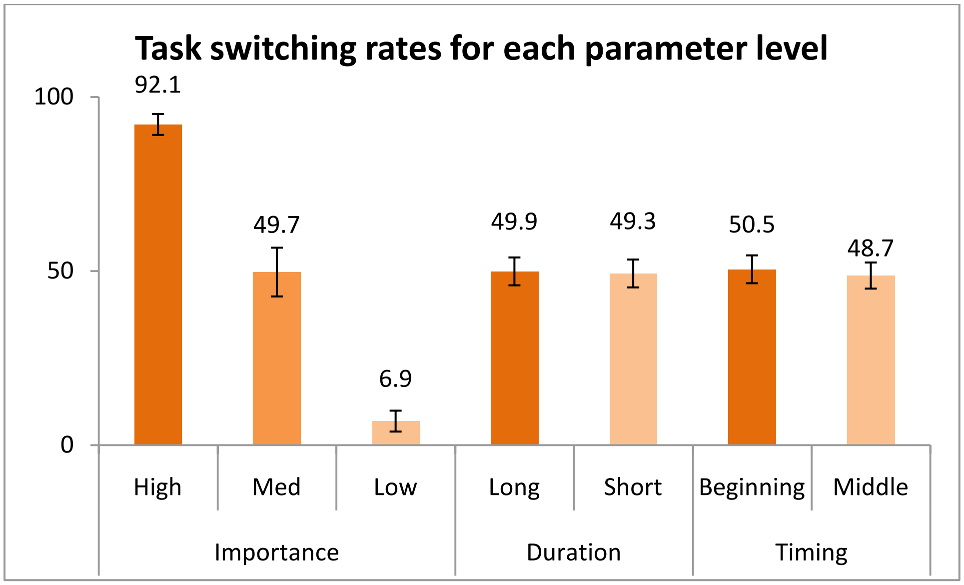


Figure 3-4: Task-switching rates for each parameter level (correctly interpreted informative cues only)

3.2.4 Performance on Arithmetic Task

Cue presence, type, and presentation timing had significant effects on arithmetic task accuracy. The peripheral visual condition showed a significant difference in arithmetic performance between cued (trials with a peripheral visual interrupting cue) and un-cued trials (93.8% vs. 97.4%), F(1, 29) = 6.35, p = .0175; see Figure 3-5. Cue type also had a significant effect, F(2, 58) = 6.99, p = .0019, on arithmetic task performance. The baseline condition (in which participants never switched tasks but simply needed to judge the presence of a cue after each trial) resulted in higher accuracy rates (98.1%) than did both the peripheral visual (95.6%) and tactile (96.4%) cues. No significant difference was found between the peripheral visual and tactile cues. The timing of cue presentation affected arithmetic task accuracies as well, F(1, 29) = 5.435, p = .027. Lower accuracy was found when cues were presented at the beginning of a trial (86.0%) than when they were presented mid-trial (87.9%).

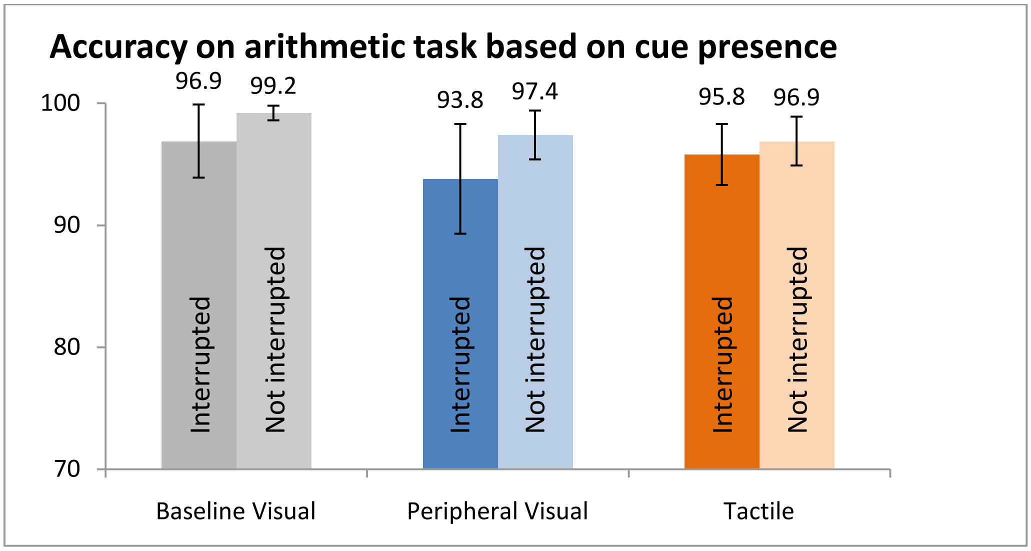


Figure 3-5: Accuracy on arithmetic task for cued and uncued trials as a function of cue type.

3.3 Discussion

The goal of the present study was to examine the effectiveness of informative peripheral visual and tactile cues for supporting interruption management, a challenge in many data-rich, event-driven domains. Overall, the findings show that both types of informative cues resulted in significantly higher detection rates than in the baseline condition. In fact, both conditions led to almost perfect detection performance. This benefit can be explained for the peripheral visual cues by their higher salience (because of size, color coding, and location relative to primary task display; Nikolic, Orr, & Sarter, 2004; Wickens, 1992). In the case of tactile cues, the use of an otherwise underutilized channel likely resulted in the observed benefit (e.g., Sarter, 2006).

The increase in detection rates as a result of increased salience and redistribution of information to other channels may not be quite as dramatic in some real-world environments as it was in the current experiment (99.2% for both peripheral visual and tactile cues). Participants' relatively stable head position and their visual attentional focus on the center of the screen likely contributed to their near-perfect detection performance for peripheral visual cues. Also, tactile cues were presented to the resting fingers of participants with virtually no incidental vibrotactile noise, which could be present in some real-world domain environments and which could mask these cues to some extent. Finally, the participants were interrupted during 50% of the trials in this experiment to create the number of trials necessary to meet the statistical demands of the study. The performance advantage may vary in domains with lower interruption rates. Still, these conditions are unlikely to entirely account for the observed performance improvements—

some domains that stand to benefit from the proposed notifications involve isolated periods of high interruption frequency, and not all involve environmental vibrations or allow for more widely distributed visual attention.

In the majority of cases (70.7% of all cases), participants were able to interpret the information encoded in the interruption cues accurately. Observed differences in interpretation accuracy for individual parameters may be the result of two factors: the need for absolute judgments in some cases and the relative appropriateness of the visual and tactile modalities for conveying certain types of information. Information on task importance and expected completion time required participants to make absolute judgments. By presenting a reference stimulus and thus requiring relative (rather than absolute) judgments, we could improve the interpretability of these parameters.

The task domain, which was encoded spatially, showed the highest interpretation accuracy. This was expected, given the high spatial discrimination capabilities of vision and touch (Intriligator & Cavanagh, 2001; Johnson & Phillips, 1981; Moy, Singh, Tan & Fearing, 2000; Sarter, 2002). In contrast, two forms of temporal encoding were used to communicate task importance (pulse frequency) and expected time to task completion (duration of cue presentation). The tactile modality is better suited for interpreting temporal information than the visual modality (which is reflected in the trend for accuracy of importance interpretation); however, both vision and touch, compared with audition, require larger differences for reliable temporal, as opposed to spatial, discriminations (Wright, Buonomano, Mahncke, & Merzenich, 1997).

Task-switching behavior. The availability of information about task importance, in combination with decision rules, allowed participants to make more informed and accurate decisions about task switching. Participants almost always (93.1%) switched appropriately when cues announced high-importance interrupting tasks (*intentional integration* in Latorella's [1999] Interruption Management Stage Model [IMSM]). Cases in which participants failed to switch to a highly important task can be explained, for the most part, because they missed or misinterpreted the cue (oblivious and unintentional dismissal breakdowns, respectively, in IMSM; Latorella, 1999). They avoided switching for 98.7% of cues that communicated low importance for the interrupting task (intentional dismissal). Very few participants (1.3%) detected and interpreted the low-importance cues correctly, and yet they incorrectly switched to the interrupting task (preemptive integration). Thus the informative cue designs used in this study seemingly support the desirable responses to interrupting cues, intentional dismissal and intentional integration.

Participants' compliance with task-switching instructions resulted in their improved overall performance, as it allowed them to optimize their performance on the secondary task, minimize unnecessary inattention to the ongoing arithmetic task, and minimize performance decrements associated with the act of switching attention between tasks. These so-called switching costs (e.g., Rogers & Monsell, 1995) have been demonstrated consistently, even for highly practiced tasks, and can increase with task complexity. The switching cost can be significantly reduced if at least partial information about an upcoming task is provided through informative cueing (Meiran, 1996).

The expected time to complete the interrupting task (communicated as duration of cue presentation) appears not to have affected task-switching behavior, even though the amount of time away from the ongoing arithmetic task should affect performance on that task. This lack of consideration may be explained, in part, by the fact that decoding this information was apparently more difficult (lowest identification rate of all parameters; see Figure 3-3), and thus perceived uncertainty may have limited reliance on this parameter. It is also possible that the relative performance costs associated with time spent away from the arithmetic task may not have been sufficiently clear to participants; therefore, they tended not to base task-switching decisions on this information.

Ongoing task performance as a function of cue type. A trade-off was observed between performance on the primary arithmetic task and the processing of interruption-related information in the informative peripheral visual condition. A similar pattern was observed in the baseline condition between arithmetic task performance and cue detection. In contrast, processing informative tactile cues minimally affected performance on the primary task. The fact that performance on the arithmetic task suffered most when peripheral visual interruption cues were presented may be explained by the onset of proximal visual stimuli capturing and leading to a reorientation of foveal visual attention, which could have interfered with the visual attentional requirements of the arithmetic task (Nikolic & Sarter, 2001). With additional training, participants may be able to suppress this visual reorientation to a larger extent.

In the baseline condition, the lower detection rate for interruption signals and the fact that less processing was required per cue presentation (because they did not encode any information about the interrupting task) may explain the fact that the arithmetic task performance decrement in cue trials did not reach significance.

Overall, the findings from this study demonstrate the benefits of using peripheral visual and tactile cues to support attention management. They highlight the effectiveness of these cues not only for reliably notifying users of a (potential) interruption but also for conveying partial information about this interruption, even under high attentional load. These findings inform the design of multimodal and graded notifications that will be elaborated upon in the following chapters.

Chapter 4 Design of Multimodal Graded Notifications

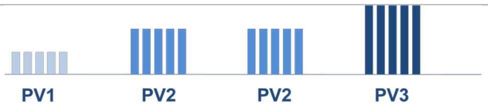
The overarching goal of this research is to design, develop, and evaluate graded multimodal notifications that support attention and interruption management by providing support for (a) reliable detection of interruptions, (b) accurate identification of the source(s) and nature of interruptions and (c) prioritization of tasks based on information about interruptions. To this end, the following types of multi-stage notifications were created and compared in the final study of this line of research: a) four-stage peripheral visual notifications, b) four-stage tactile notifications, c) graded peripheral visual notifications, d) graded tactile notifications, and e) crossmodally (combining peripheral visual and tactile) graded notifications. The four-stage notifications are enhanced versions of traditional alarms, in that they employ signals at only one level of salience and inform the operator only that an incident has occurred, but this information is presented four times over a given period of time in the form of four identical cues to increase the likelihood of detection. The location of the incident is also communicated via the site of the cue source. Graded notifications are also presented as four cues but they begin at a low level of salience and get stronger with time if they are not responded to. While the four-stage and graded peripheral visual and graded tactile notifications comprise of a single modality, graded multimodal notifications employ two peripheral visual cues followed by two tactile cues at increasing levels of salience. The following sections will describe in more detail the steps and decisions involved in designing these notifications.

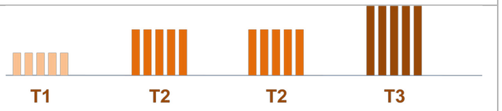
4.1 Basic design requirements

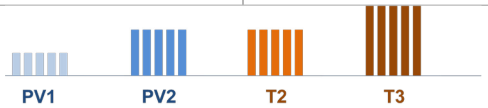
For the purpose of this study, a notification is defined as ‘a set of cues presented over a period of time with the goal of providing an operator with partial information about the (possibly changing) nature of an event that might require the his/her attention’. Thus, notifications are different from traditional alarms which are single-stage warnings and inform an operator only that an anomaly has occurred without presenting any additional details to aid in deciding whether/when to attend to the event (Woods, 1995). As mentioned above, the following types of notifications were created and compared in this study: a) four-stage peripheral visual, b) four-stage tactile, c) graded peripheral visual, d) graded tactile and e) graded multimodal (combining peripheral visual and tactile). Notifications were presented in peripheral vision and via the sense of touch to exploit the orienting power and ability of these two channels to guide visual attention to relevant information. All notification cues were presented up to four times, unless responded to at an earlier stage. In the case of non-graded notifications, this simply served to remind operators of the event if they decided to postpone shifting their attention away from the ongoing task. Graded cues were also presented up to four times but at 3 levels of salience to reflect the increasing urgency of attending to the event. The entire set of notifications that were tested is depicted in Table 4-1.

<table>
<tr>
<td>4-stage</td>
<td>

(a)</td>
<td>

(b)</td>
</tr>
<tr>
<td>Graded</td>
<td>

(c)</td>
<td>

(d)</td>
</tr>
<tr>
<td>Graded multimodal</td>
<td colspan="2">

(e)</td>
</tr>
</table>

Table 4-1: Set of notifications: a) four-stage peripheral visual cues, b) four-stage tactile cues, c) graded peripheral visual cues, d) graded tactile cues, and e) crossmodally graded cues. Increase in depth of color and increase in height of bars indicate an increase in salience.

One of the primary requirements for an effective notification is that it is reliably detected without being intrusive to the point of distracting an operator from ongoing tasks. In this study, reliable detection was defined as a notification being perceived at least 90% of the time at the lowest salience level and in the absence of any concurrent workload.

To ensure comparability across the different notification designs, the salience levels of peripheral visual (pv1, pv2, pv3) and tactile cues (t1, t2, t3) were chosen such that they fulfill the criteria listed in Table 4-2. Note that a) the salience levels of cues across modalities were matched and b) in the case of crossmodal gradation, the second peripheral visual signal and the first tactile signal were presented at the same salience level (PV2=T2) to be able to isolate the effect of modality switching on noticeability.

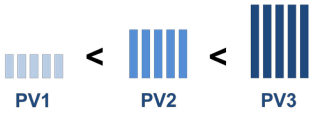
Graded peripheral visual cues	
Graded tactile cues	
Relationship between peripheral visual cues and tactile cues	

Table 4-2: Criteria used to design salience of cues that are combined to form graded notifications. Increase in depth of color and increase in height of bars indicate an increase in salience.

Also, it was important that the peripheral visual and tactile cues differed only in terms of the presentation modality but were subjectively comparable in terms of how they presented information. For example, if the peripheral visual cue for a certain type of interrupting event is presented at the rate of 3 onsets per second, the tactile cue for the same event were also presented at 3 vibrations per second. Lastly, the interface was designed such that it would not interfere with normal operations since this would make it unacceptable for most real-world domains.

4.2 Notification properties

The notifications were designed with the following goals in mind:

1. Reliable detection

    a. Notifications were designed such that they were perceived at least 90% of the time at the lowest salience level and in the absence of any concurrent workload (see section 4.2.2).

    b. Cue salience, encoded in the brightness of the peripheral visual cues and the frequency of the tactile cues, fulfilled the three criteria $p1<p2 < p3$, $t1< t2< t3$, and $p2=t2$.

2. Accurate interpretation

    a. Location of event. To ensure accurate and reliable localization of the interrupting event, the peripheral visual cues were presented around the sector in which the event occurred and the location of the tactors on the participant’s abdomen mapped onto the 8 sectors of the visual display, thus exploiting crossmodal spatial links in attention (see section 4.2.1).

    b. Importance of event. Cue patterns are designed to represent the level and changes in importance of attending to interruptions (see section 4.2.3).

Cue parameters were chosen based on a review of the relevant literature and extensive pilot testing. The following sections elaborate on the process of developing the various cue parameters.

4.2.1 Arrangement of cues

The arrangement of peripheral visual and tactile cues is crucial to the effectiveness of notifications. Ideally, peripheral visual cues should overlay the visual display so that they guide attention to the location of importance, and tactile cues should be arranged in a layout similar to the elements on the primary visual display such that they can be easily and naturally mapped. In this experiment, the arrangement of cues mimics the 8 sectors around the hand-off region on the ATC radar screen. Extensive pilot testing was conducted to ascertain the location of both peripheral visual and tactile cues that would be most beneficial to the operator.

4.2.1.1 Peripheral visual cue arrangement

The peripheral visual cues consisted of flashing corner segments outlining the outer vertices of the sector in which the interruption occurred along the edges of the screen (see Figure 4-1). The cues were presented at a visual angle of 19° vertically and 28° horizontally, measured from the center of the hand-off region, as shown in Figure 4-2 .

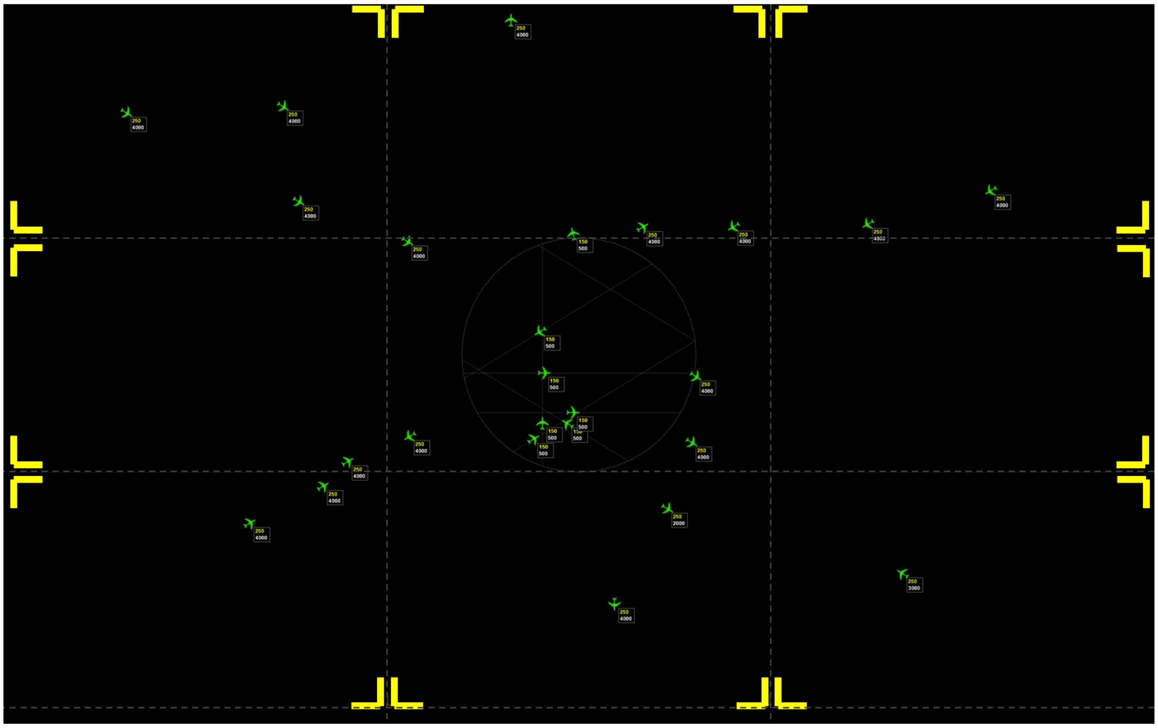


Figure 4-1: Peripheral visual cues for each of the eight sectors on the ATC screen.
Note: Size and brightness of cues are exaggerated for better visibility.

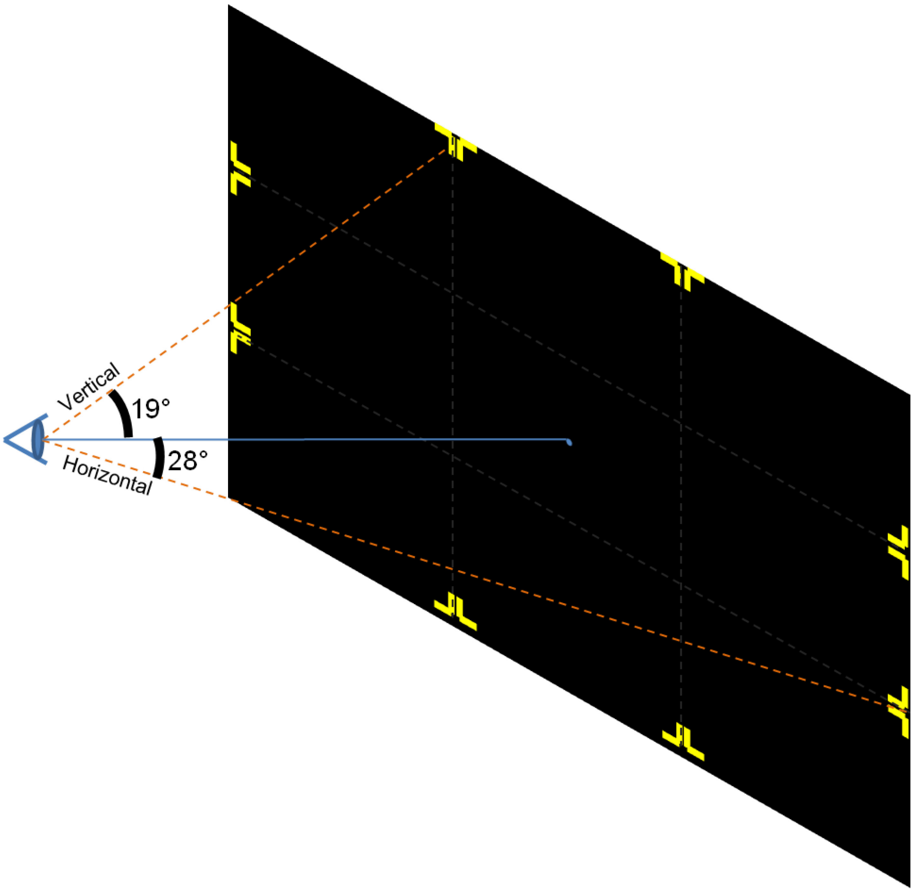


Figure 4-2: Vertical and horizontal visual angles at which peripheral visual cues were presented

The choice of these specific vertices of the eight rectangular sectors was made based on two guidelines. Firstly, all the chosen vertices must be visible within the binocular field of view so that none of them will have the potential disadvantage of being seen only in the uniocular fields (Henson, D. B., 1993) (see Figure 4-3). Second, the vertical and horizontal angles distended by the chosen vertices must be uniform along the height and the width of the screen (see Figure 4-2). The vertices for the four corner sectors that lie on the vertices of the screen would have violated both of these guidelines.

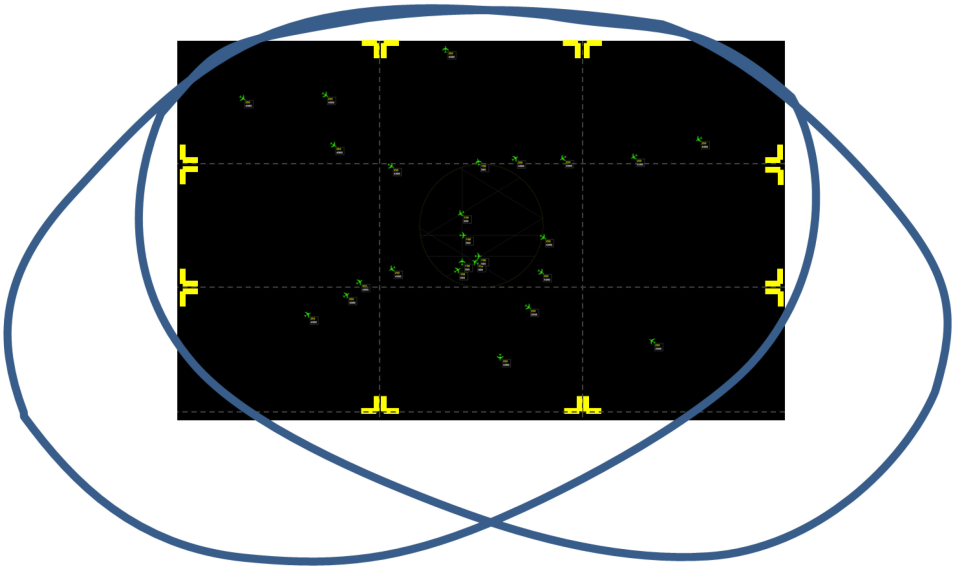
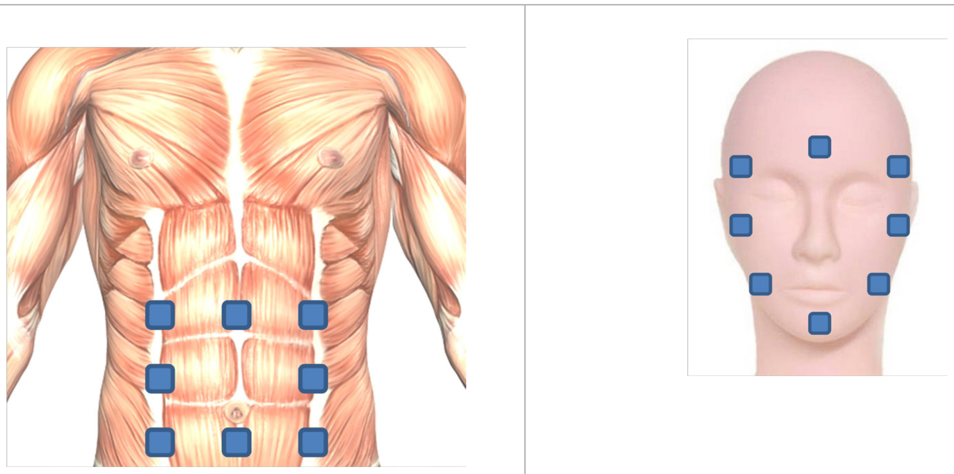


Figure 4-3: Arrangement of peripheral visual cues within the binocular field of vision

4.2.1.2 Tactile cue arrangement

The tactile cues were presented using tactors, i.e., devices that present vibrations to the skin (in this case: Engineering Acoustics C-2 tactors, 250 Hz, 0.3” dia.) at a frequency range of 30 – 350 Hz and a gain of 2. The tactors were arranged such that the location of the cues can be easily mapped to the locations of the eight sectors on the radar screen, thus exploiting crossmodal spatial links (e.g., Driver and Spence, 2004; Spence et al, 2004; Sklar and Sarter, 1999; Ho, Tan and Spence, 2005; Hameed et al, 2006). An initial review of literature on tactile sensitivity of various body parts suggested the back of the abdomen, the front of the abdomen, and the face as candidate locations (Weinstein, 1968; Cholewiak, Brill and Schwab, 2004). The back of the abdomen had been tested in our laboratory in earlier studies (see Chapter 2 for example). Tactile harnesses were fashioned in the laboratory to test the feasibility of using the face and the front of the abdomen as tactor sites (see Figure 4-4). The results from the pilot study suggested that the vibrotactile sensitivity of front of the abdomen and the back of the abdomen were similar.

However, the choice of the face as a tactor site was not desirable due to (a) high levels of discomfort to participants, (b) complexity of maintaining tactor contact on the skin, and (c) the requirement of the absence of surface irregularities such as facial hair.



Front of abdomen

Face

Figure 4-4: Placement of tactors on various test locations

The two remaining body sites, front of the abdomen and back of the abdomen, were both well suited for the purposes of presenting information using tactors. The sensitivity to, and localization of, stimuli at these sites have been thoroughly explored in early psychophysical studies (e.g., Weber, 1826/1978; Penfield and Rasmussen, 1950; Weinstein, 1968). However, experiments in these studies measured the ability to detect gaps in stimuli pressed to the skin or to localize taps or probes. These findings cannot be applied to the current research since the response to touch and the response to vibrations are vastly different. One reason is that tissue movements due to vibratory signals are large (Franke, et al, 1951). Another reason is that this movement can occur deep in the dermis and epidermis where many of the vibratory receptors exist, thus creating a significant difference between a simple touch and a vibration signal (Bolanowski, Gescheider & Verrillo, 1951; Greenspan & Bolanowski, 1996).

More recent studies have explored the effects of *vibrotactile* stimuli on the body, especially on the abdomen (Dobbins and Samways (2002); Priplata, et al, (2003); Rupert (2000); Tan, Lu, and Pentland (1997); Van Erp (2000)). Of particular interest and relevance to this current research is the extensive set of experiments conducted by Cholewiak and colleagues on the sensitivity and accurate localization of vibrotactile stimuli on the abdomen (Cholewiak, Brill and Shwab, 2004). These experiments served as the basis of selecting the tactor sites for this study.

The experiments conducted by Cholewiak and colleagues tested the sensitivity and accurate localization of vibrotactile stimuli presented on tactor sites around the abdomen, starting from above the navel and going around the trunk on both sides to the spine. The results of these experiments show that (a) localization is significantly better near the navel and the spine than on the sides of the body, (b) identifying the locus of the stimulus was near perfect at the navel and spine locations, (c) no significant difference in localization was observed between the back and the front of the abdomen, nor between the different heights of the locations above the navel. Another study by Van Erp and Werkhoven (1999) found that the vibrotactile sensitivity along the center of the torso, both in the front and at the back, was higher than on the sides.

These findings suggest that, in principle, there is no significant difference in perception and performance between using the front of the abdomen and the back of the abdomen as tactor sites. However, proper mapping seems to be critical. For example, recent studies have shown faster response times to vibrotactile cues that are presented from the location or direction in which a corresponding critical event occurs (e.g., Ho et al., 2005). For the purpose of this research, the front of the abdomen was therefore chosen as it faces the screen in front of the air traffic controller where all relevant information is presented.

Given that both body sites were equally effective in locating vibrotactile cues, and given that the back of the abdomen has already been tested in earlier studies, the front of the abdomen was chosen for the current research. This choice provided the potential of answering an additional question: how will the mapping of tactile vibrations on the front of the body to the screen compare to mapping vibrations on the back of the body?

4.2.2 Cue salience

The appropriate cue salience was determined after the user interface, cue placements and cue patterns had been established. Determination of cue salience is critical to this study for several reasons. First, it is important to determine the lowest salience level that will lead to reliable detection of the cue. Secondly, the upper bound needs to be determined to make sure the signal is not too intrusive and will not, in a mandatory fashion, capture and redirect attention. Finally, crossmodal matching of cue salience for the peripheral visual and tactile cues in this study was critical to avoid a confound between modality and salience. Since four levels of salience were needed for the graded design, the two intermediate levels had to be chosen to ensure that they were equivalent to each other and distinguishable from the lowest and highest acceptable salience value. The peripheral visual and tactile salience levels had to meet the criteria: $p1 < p2 < p3; t1 < t2 < t3; p2 = t2$. The following sections describe a pilot study and three experiments that were conducted to determine the appropriate salience values.

4.2.2.1 Salience Determination – A pilot study

The first step in determining the appropriate salience levels for notifications was to identify perceptual and comfort thresholds in the current simulation set up. Since the experimental apparatus and cue design in this study are unique, a pilot test had to be conducted to identify perceptual thresholds. Seven participants (4 male, 3 female; age: 23-28 yrs) were seated in front of the black simulation screen which did not display any airplanes. Participants were instructed to focus their vision on a dim blue dot of 0.5" radius and 20% of maximum screen brightness that was displayed in the middle of the 2.5" radius handoff region. They were asked to look at, or immediately around, that dot to avoid "seeing spots". They were presented with peripheral visual cues that ranged from 1% to 20% of maximum screen brightness (which was 400cd/m^2) and lasted 3 seconds. The cues consisted of yellow blinking segments outlining the vertices of the sector in which a potential interruption would be present, as described in section 4.2.1. Participants were asked to respond verbally as soon as they noticed a cue. Four out of the seven participants perceived all cues whose brightness was at or above 8%; two noticed cues at or above 7% of maximum brightness; and one participant noticed them at 9% or above. Based on these observations, the lowest brightness level for peripheral visual cues in the current simulation was set at 8%.

To determine the tactile detection and comfort thresholds, participants were fitted with a harness to which 8 tactors were attached and placed over their abdomen. They were presented with medium intensity vibrations (3 pulses per second) that lasted for 3 seconds. Initially, vibration frequencies ranging from 50Hz to 350Hz were presented one after another at intervals of 10 seconds and increments of 50Hz. Participants reported that frequencies of the vibration sensation increased steadily from 50Hz to 250Hz after which it began to decline. Another set of tests was conducted with frequencies between 100Hz to 250Hz to determine whether any of those frequencies felt too intense or unpleasant. Five participants reported that they began feeling uncomfortable when signal frequencies of 200Hz and 210Hz were employed. Two other subjects reported that frequencies above 220Hz resulted in an unpleasant “tingling” sensation. Based on these reports, the discomfort (upper) threshold for tactile cues was set at 180Hz.

These pilot tests served to identify upper and lower bounds for the peripheral visual and tactile signals in the context of the experiment set up that will be used in subsequent experiments. Following these pilot tests, three experiments were conducted to identify the salience levels within the acceptable ranges that can be distinguished reliably and can therefore be used for the design of graded notifications. An additional goal was crossmodal matching of salience, i.e., the identification of salience levels for peripheral visual and tactile cues that are perceived to be equal and can therefore be employed when transitioning between the two modalities as part of the crossmodal gradation. The first experiment identified the lowest salience levels for peripheral visual cues that can be identified in the absence of any concurrent tasks or workload. These are different from the values obtained in the pilot study in which only a black screen with a blue dot was used with the cues to measure perception thresholds. The screen in the first experiment displayed airplanes moving across the screen, similar to what will be seen in the final study. The second experiment determined the upper bounds of tactile cue levels that can be reliably detected in high workload situations without interfering significantly with the ongoing tasks. The lower bounds of the tactile cues were determined by the third experiment. The third experiment was a crossmodal matching study that determined the salience levels of peripheral visual and tactile cues that are perceived to be equal.

4.2.2.2 Experiment 1 – Lowest acceptable salience for peripheral visual cues

This experiment was conducted to determine the lowest salience levels for peripheral visual cues that can be identified in the absence of any concurrent tasks or workload.

Method: 18 students participated in this experiment (7 females, 11 males; aged 19-32 yrs) and were compensated for their time. All participants had normal or corrected-to-normal vision and no experience with air traffic control. Participants were seated in front of the ATC simulation screen and were asked to focus their vision on the blue dot within the hand-off region. Twenty-five airplanes flew across the screen at the speed of 7 seconds per inch to simulate a low workload scenario. Peripheral visual notifications indicated the presence of an incident in one of the 8 sectors. The cues were presented in randomized order at varying levels of salience (10%, 11%, 12%, 13%, 14% brightness) and three frequencies (1, 5, and 10 pulses per second). Participants were instructed not to look for, or directly at, cues and to respond to cues that they noticed in the periphery by tapping the spacebar on the keyboard. Detection rates and response times were recorded.

Results: To identify the lowest salience level at which the signals could be detected reliably, a cross-tabulation analysis was employed that compared the detection rates for the various salience levels. The results of the cross-tab analysis confirm that, as salience increases, the detection rate increases correspondingly. The Pearson's R (0.526, N = 450) also showed a significant positive correlation between salience and detection rates. A Chi-Square Goodness of Fit test indicated that this relationship is highly significant (χ^2 (4, N = 450) = 173.883, $p < 0.001$). The lowest salience level at which p-v cues were detected at least 90% - our preset criterion - of the time was 13% (see Figure 4-5).

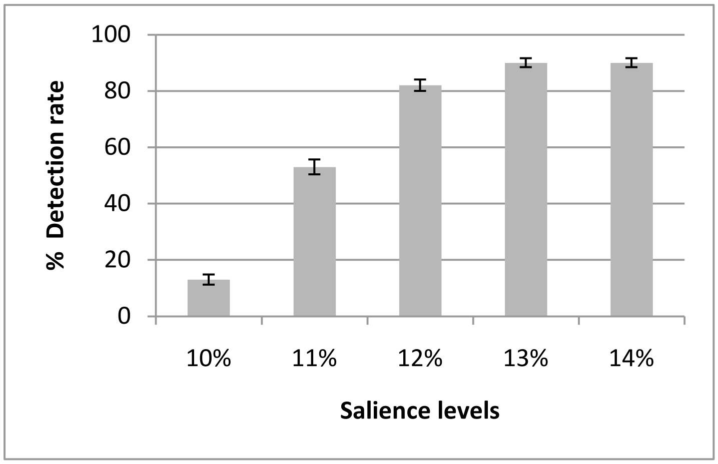


Figure 4-5: Mean detection rates and standard error for various levels of salience

4.2.2.3 Experiment 2 –Upper bounds of tactile cue salience

This experiment was conducted to determine the upper bounds of tactile cue levels that can be reliably detected in high workload situations without interfering significantly with the ongoing tasks.

Method: Seven students (1 females and 6 males; aged 21-28 years) participated in this experiment and were compensated for their time. Participants were again seated in front of the ATC radar screen which displayed a circular hand-off region at the center surrounded by 8 sectors (see section 6.2). In this experiment, ninety airplanes moved across the airspace at a speed of 7 seconds per inch to simulate a high workload scenario. Participants' main task involved closely monitoring airplanes that entered the hand-off circle at the center of the screen for two possible events: airplanes flying at the wrong speed and airplanes flying on a wrong heading. If an airplane was flying at the wrong speed, this could be detected based on information in the data block. Participants were instructed to left-click on the airplane and select the speed correction factor from a drop-down menu. If an airplane was flying on a wrong heading, the top down view on the radar scope would show it deviating from its prescribed path which was indicated by a grey line. Participants had to left-click on the airplane and select the "rectify heading" option from a drop-down menu. One of these events occurred approximately once every 7 seconds. In addition, approximately once every 42 seconds, an "interruption" would occur in the form of an airplane flying below the recommended altitude of 4000 ft. Altitude levels were displayed at the bottom of the airplane data block in white. An altitude deviation was considered gentle if the altitude dropped to 3000 ft, moderate when it was 2000 ft and severe when it was 1000 ft. When an interruption was detected, participants had to left-click on the airplane and select the correction factor from a drop-down menu to instruct the airplane to increase its altitude by that amount.

Participants were fitted with a harness around their waist with an array of eight tactors on the front of their abdomen that mapped directly onto the 8 sectors on the screen. Tactile cues were presented to indicate the occurrence of an interruption in one of the 8 sectors. Since prior testing showed that cue frequencies of 200Hz and above crossed the comfort threshold, the next acceptable frequency - 180Hz - was employed in this experiment. Pilot tests were conducted to ensure that cues at this frequency could be perceived 100% of the time. Participants placed their dominant hand over a mouse and their non-dominant hand over the spacebar on the keyboard. They were instructed to perform their main tasks (above) while responding as quickly as possible to tactile cues.

The experiment was divided into four sessions that participants completed in randomized order: two with and two without tactile cues. When tactile cues were included, they were presented at 180Hz in one of the eight locations on the participant's abdomen at varying intensities (gentle at 1, moderate at 3, and severe at 10 pulses per sec) in randomized order. Participants responded to cued interruptions with their non-dominant hand by tapping on the spacebar - once for gentle, twice for moderate and three times for severe cases - while performing their main tasks. Performance was measured using hit rates for main tasks and correct detection rates for interrupting cues. Performance on the main task was compared between the interrupted and uninterrupted sessions.

Results: The rates of accurate detection of interruptions was compared between cued and uncued conditions using repeated measures ANOVA and found to be significantly different ($F(1,6)=417.684$; $p < 0.001$). Accuracy on the cued trials was 92.3%, almost three folds the accuracy on the uncued trials (33.0%) (see Figure 4-6). The performances on the main tasks during the cued and uncued trials were also compared. Performance in the cued trials was 46.3% which was less than performance in the uncued trials (50.9%), although the two were not significantly different (see Figure 4-7). The results imply that the presentation of tactile cues at a salience level of 180Hz vastly improved the accuracy of detecting and identifying interruptions without significantly affecting main task performance, despite being detected correctly.

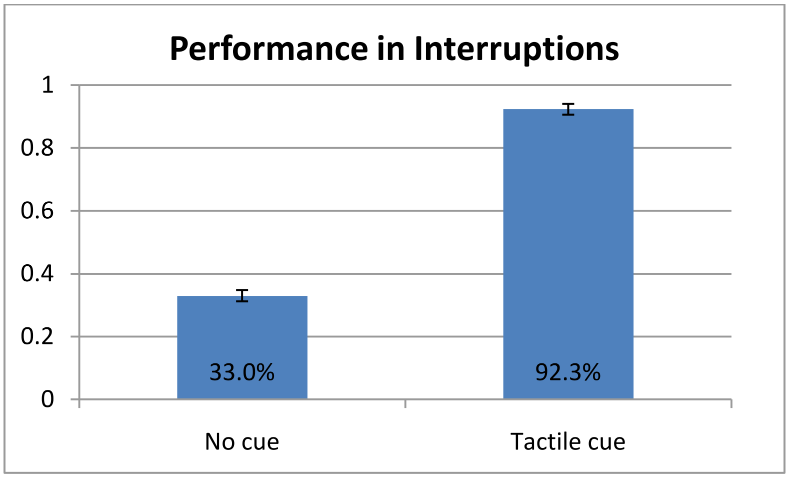


Figure 4-6: Accuracy of detecting interruptions (mean and SE) in uncued and cued trials

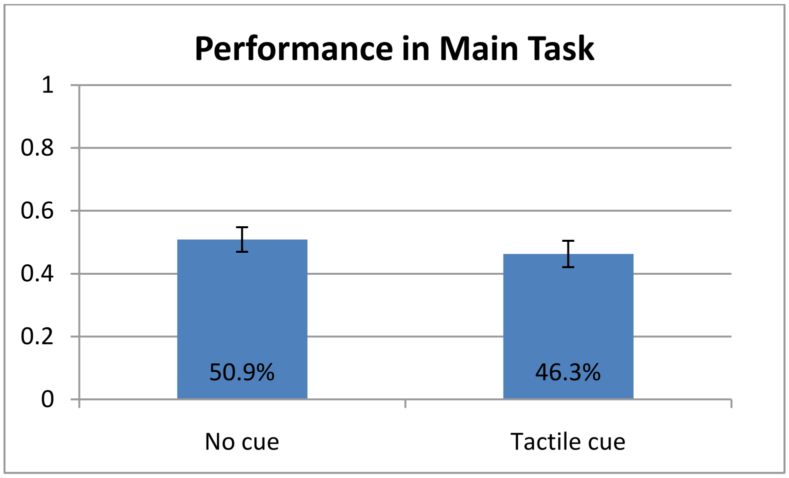


Figure 4-7: Performance in main task (mean and SE) in uncued and cued trials

4.2.2.4 Experiment 3 – to determine equivalent peripheral visual and tactile cues

Experiment 3 was conducted for the purpose of crossmodal matching, i.e., to determine the setting at which peripheral visual cues and tactile cues were perceived to be equally salient. These settings were then used for p2 and t2 in the crossmodal gradation condition to ensure that an increase in detectability can be attributed to the switch in modality rather than an increase in salience.

Method: A total of 18 participants (7 females, 11 males, average age 23.5) participated in this experiment and were compensated for their time. Participants were seated in front of the ATC screen and provided with a keyboard with the up-arrow, down-arrow and return key highlighted. The screen displayed the 8 sectors, the hand-off region and the dim blue dot described in Experiment 1. While participants focused their vision on the center of the screen, they were presented with simultaneous peripheral visual and tactile cues. All cues were of moderate intensity (5 pulses per sec). Salience levels were selected based on a combination of piloting and results from experiments 1 and 2 (13%, 16%, 19%, 22% for peripheral visual and 70Hz, 100Hz, 130Hz and 160Hz for tactile cues). The experiment involved two sessions: one for the purpose of peripheral visual adjustment (PVA) and another for tactile adjustment (TA), presented in random order. In the PVA session, participants were asked to use the arrow keys to increase or decrease the salience of the peripheral visual cues to "match" the salience of the tactile cues and then strike the return key. In the TA session, participants adjusted the salience of the tactile cues to match the salience of the peripheral visual cues. Participants were allowed to take as long as they needed to determine the salience levels at which the cues matched. The matching values were recorded in both modes.

The results from the two sessions were compared by calculating the median values (Bond & Stevens, 1969). The comparison showed that the 16% brightness peripheral visual cue was perceived to be equivalent in salience to the 105Hz frequency tactile cue, and the 19% brightness PV cue matched the 130Hz tactile cue (see Figure 4-8). These matching values were then tested in a setup similar to the one in experiment 1 to ensure that they can be detected 100% of the time in the absence of any workload.

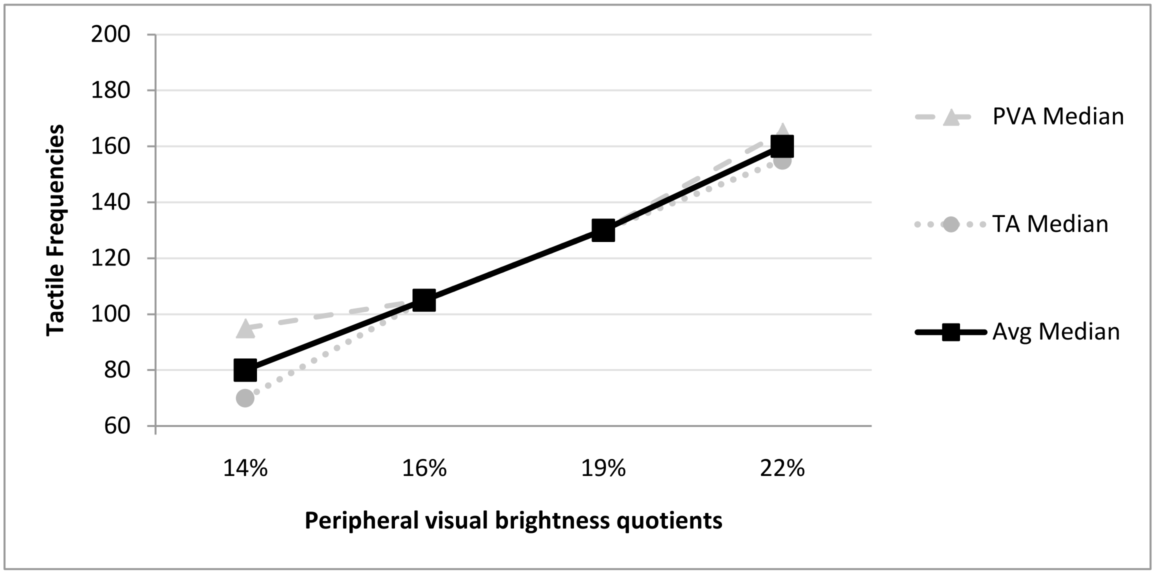


Figure 4-8: Median values for crossmodal matching of peripheral-visual and tactile cues

The experiments described above revealed the detection thresholds that are unique to the current experiment setup and the crossmodal matching values that are critical for the final experiment. These experiments resulted in the establishment of the following cue values employed in the final notification design:

pv1 = 13%	Pv2 = 16%	Pv3 = 22%
t1 = 70Hz	t2 = 105Hz	t3 = 180Hz

4.2.3 Cue patterns

Supporting preattentive reference requires that partial information is provided about the nature of the interrupting task or event. One important aspect that needs to be conveyed is the urgency of attending to the impending interruption. As demonstrated in previous studies (Chapters 3 and 4), modulating the pulse rate of the cue is the most effective means for achieving this goal. Such modulation can be applied to peripheral visual cues as well as tactile cues. In this study, pulse rates of 1 per second, 3 per second, and 10 per second were used to represent a gentle, moderate or severe deviation from normal operations (Figure 4-9). Extensive pilot studies revealed that pulse rates of less than 1 per second are perceived as too far apart to be part of a single cue, while pulse rates of more than 10 per second begin to be perceived as one continuous signal.

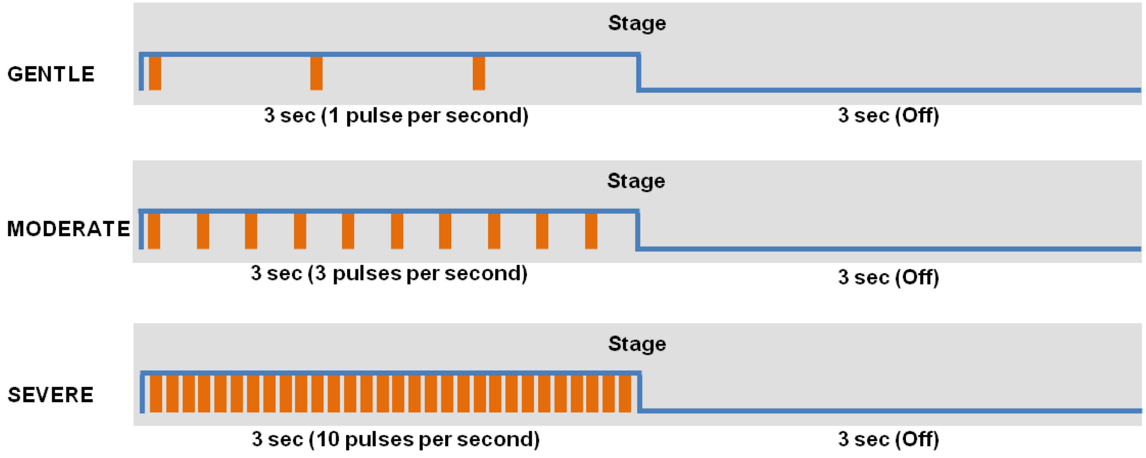


Figure 4-9: Pulse rates used in gentle, moderate and severe cues

Interruption notifications consisted of four cues: stages a, b, c, and d. Cues had an on- and off-time of three seconds each, resulting in a total notification duration of 24 seconds. An additional six-second period was provided for participants to respond to the interruption (stage e). Figure 4-10 depicts the cue stages and cue durations within a notification.

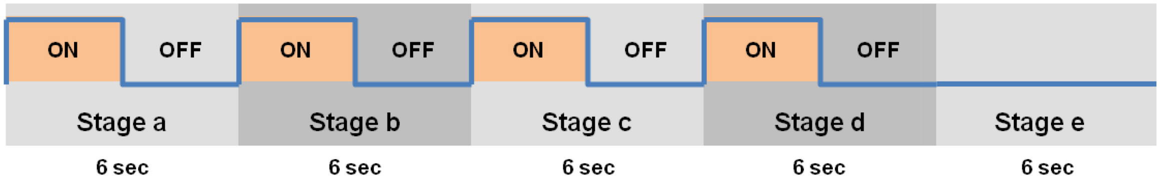


Figure 4-10: Cue stages and cue durations within a single notification

In this chapter, the basis for and process of designing five different types of notifications were described. The spatial arrangement of these notifications, the salience levels of the individual cues, and the patterns employed to convey information were described. Findings from preliminary evaluations of the various notifications were presented. The following chapter focuses on the final experiment in this line of research which served to evaluate comparatively these notifications in the context of a simulated air traffic control task set and environment to examine their robustness and ability to support monitoring and interruption management.

Chapter 5 Evaluation of Multimodal Graded Notifications

5.1 Purpose, goals and expectations

The purpose of this final experiment is to assess and compare the effects of graded and non-graded peripheral visual and tactile notifications on the performance of concurrent tasks in a dynamic, event-driven environment (in this case, simulated simplified ATC operations). The choice and design of these notifications followed from the findings of the earlier experiments described in this document. The main goal of these notifications is to support operators in attention and interruption management. They are expected to:

1. Improve performance in detection of, and response to, interruption tasks
2. Improve decision making about attending to interruptions
3. Improve overall performance under high cognitive load
4. Maintain or improve performance on primary tasks
5. Improve timesharing performance and management of primary and interrupting tasks

Specifically, the results of the experiments are expected to show the following:

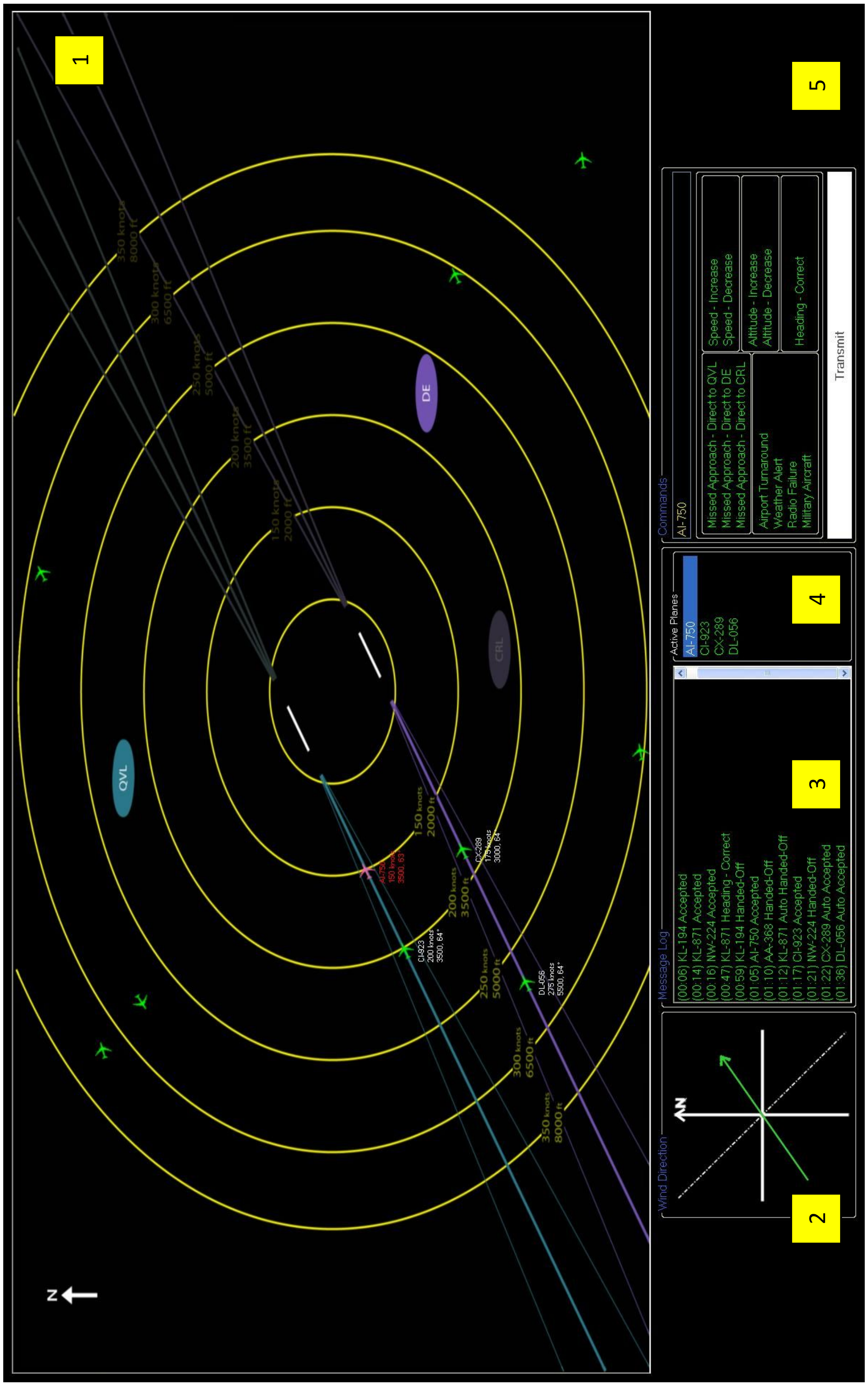
1. Peripheral visual and tactile interruption notifications do not adversely affect primary (focal visual) task performance (both detection rates and accuracy)
2. Response times to primary task events are faster in all cued conditions
3. Detection rates and accuracy for interrupting tasks are higher, and response times are faster, in all cued conditions
4. The highest accuracy and detection rates are observed with graded cues
5. Graded interruption cues result in better performance (detection rates and accuracy) than non-graded ones, especially during high workload
6. Graded notifications result in better performance (detection rates and accuracy) than other cued or uncued conditions when multiple interruptions coincide
7. Response times to altitude deviations will be inversely proportionate to the severity of the deviations with graded notifications
8. Multitasking performance is better in cued than un-cued conditions

5.2 Identification of Promising Tasks/Events

The primary goal of this pilot study was to explore when and why breakdowns in monitoring are likely to occur in a simulated ATC task. The aim was to identify the types of events/notifications that are most likely to be missed by controllers, especially under high workload conditions, and that should therefore be employed in this final experiment as tasks and interrupting events for which notifications will be designed and tested.

5.2.1 Method

20 UM students (age: 19-30 yrs) with no experience in ATC operations participated in this study. Lack of experience was acceptable since this study involved only a subset of actual ATC tasks and focused on perceptual aspects of the task rather than the complex decision making that experienced controllers accomplish. The participants completed a background questionnaire at the beginning of the experiment, followed by three 10-minute training sessions to familiarize themselves with the interface, the associated cues, and their tasks. After reaching the performance criterion of 70% accuracy on their tasks, participants completed a 90-minute experimental session. Their main task was to accept and hand off of airplanes entering and leaving the sector. To manipulate operator workload, the number of airplanes on the screen was gradually increased from 2 to 4 (low workload), 4 to 6 (medium workload), 6 to 10 (high workload) and then decreased back to 2 airplanes. Each airplane took 40 seconds to pass through the ATC sector.



```
(00:06) KL-194 Accepted
(00:14) KL-871 Accepted
(00:16) NW-224 Accepted
(00:47) KL-871 Heading - Correct
(00:59) KL-194 Handed-Off
(01:05) AI-750 Accepted
(01:10) AA-368 Handed-Off
(01:12) KL-871 Auto Handed-Off
(01:17) CI-923 Accepted
(01:21) NW-224 Handed-Off
(01:22) CX-289 Auto Accepted
(01:36) DL-056 Auto Accepted
```

Figure 5-1: Desktop ATC simulation. Panel 1: Main radar screen with runways (center), speed zones (concentric circles), landing wedges (diagonal lines), way points, airplanes with their data blocks. Panel 2: Wind wane. Panel 3: Message log. Panel 4: Active airplanes. Panel 5: Interruption commands and “transmit” button.

To replicate some of the attentional requirements of an ATC controller, three frequent and five rare interrupting events were presented at random to add to the variety and reduce the predictability of interruptions. In the experiment scenario, frequent events (altitude deviation, speed deviation and airplane deviation) appeared 16 times each and rare ones (missed approach, radio failure, bad weather, wind direction change and military aircraft intrusion) appeared 4 times each. Participants were not aware of the frequency of occurrence of the events in advance. Each event was indicated using subtle (level 1) or salient (level 2) changes in visual attributes of the airplane or display object in question. These changes were broadly classified into blink, dim, directional change, numerical change and onset (see Table 5-1: Cue types and visual attributes of eventsTable 5-1). Throughout the experiment, participants performed their main task of monitoring of the airspace to accept and hand off of airplanes entering and leaving the sector. When presented with an interrupting event, participants were required to detect and accurately identify it correctly by clicking on the element of interruption (from panel 1 in Figure 5-1), confirming it from a list of given interrupting events in the command window (panel 5 in Figure 5-1) and then clicking on a response button ("transmit" button in panel 5 in Figure 5-1). For example, if a participant detected an airplane deviating from its path, s/he should click on the airplane, confirm it as "plane deviation" from the list of events and then click on the transmit button. The dependent measure for the events was rate of accurate detection. The influence of workload and the types of events on interruption detection accuracy were measured.

Cue Type	**Event**	**Visual attributes**
Blink	Missed approach - Level 2	Data block blink (8 per sec)
	Radio failure - Level 2	Dim and airplane blink
Dim	Bad weather - Level 1	Gradual onset
	Radio failure - Level 1	Dimming of data block
Directional	Speed deviation - Level 1	Numerical change in data block
	Wind direction - Level 1	Directional change (far)
	Wind direction - Level 2	Directional change (near)
Numerical	Altitude Deviation - Level 1	Numerical change in data block
	Plane deviation - Level 1	Directional change (5°)
	Military aircraft - Level 1	Numerical change in data block
Onset	Altitude Deviation - Level 2	Abrupt color onset
	Bad weather - Level 2	Abrupt onset
	Plane deviation - Level 2	Directional change (20°)
	Missed approach - Level 1	Data block blink (2 per sec)
	Military aircraft - Level 2	Onset of larger image
	Speed deviation - Level 2	Abrupt color onset

Table 5-1: Cue types and visual attributes of events

5.2.2 Results

A repeated-measure ANOVA was used to analyze the effect of workload and event type on detection rate. A main effect of workload (F(2,38)=30.893, $p < 0.001$) resulted in significantly lower detection rates under medium and high workload (both 45%; $p<0.001$), as compared to low workload (62.9%) (see Figure 5-2).

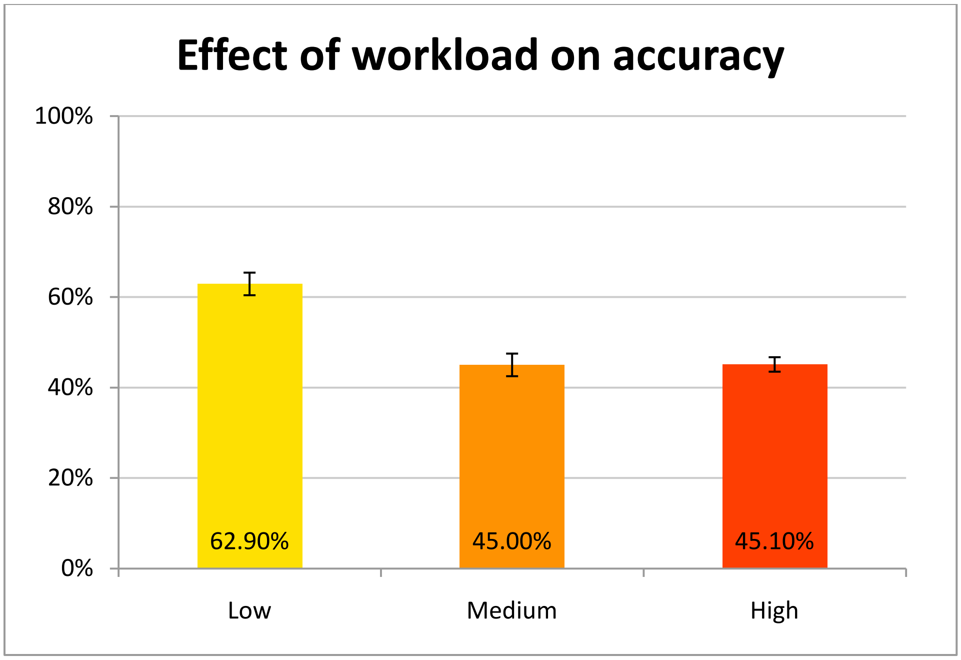


Figure 5-2: Accurate detection rates (mean and SE) under low, medium and high workload

There was also a main effect of cue type (F(4,76)=123.255, p (< 0.001)). All pairwise comparisons (Tukey's LSD post-hoc tests) between the 5 cue types revealed significant differences in detection performance (p = 0.004 for blink and onset; p = 0.021 between dim and numerical; p < 0.001 for all other combinations). The lowest detection rate was observed for directional cues (8.9%), followed by numerical cues (25.9%) (see Figure 5-3). Dimming effects were noticed in 41% of all cases. And the highest detection rates were found for blinking cues and cues that featured an abrupt onset (85.5% and 76.8%, respectively).

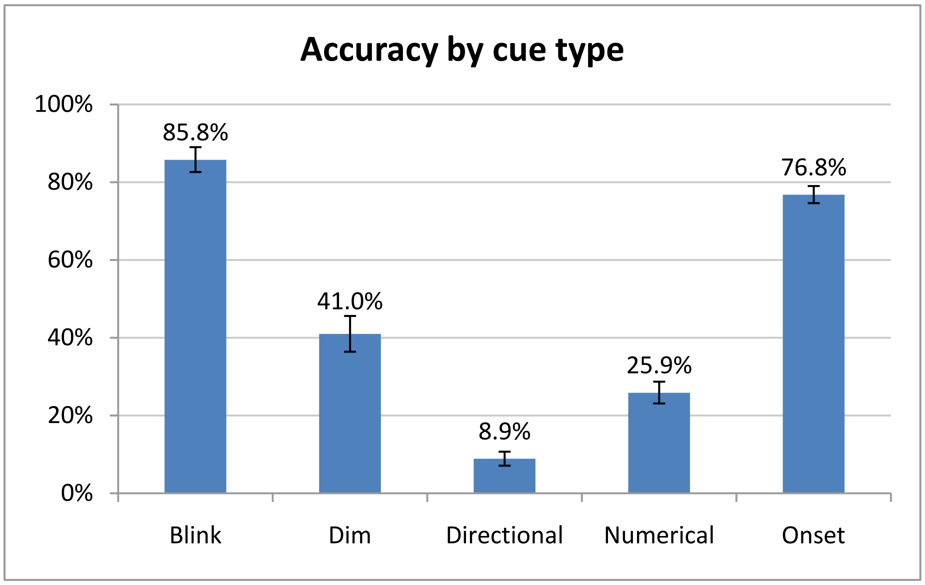


Figure 5-3: Accuracy of detection (mean and SE) for every cue type

The findings from this study were used to inform the design of tasks and events in the final experiment. In particular, directional and numerical changes which were most often missed in this study, were promising candidates for demonstrating the benefits of improved notification designs.

5.3 Participants

Thirty one students from the University of Michigan (11 females, 20 males; aged 19-30) participated in this study. All participants had normal or corrected-to-normal vision and no known disorders that may have affected tactile sensitivity or peripheral visual abilities. Two participants were color-blind (which was acceptable since interface components were not distinguished solely based on color) and one was a certified pilot who had prior knowledge about air traffic control operations. Participants were required to have had sufficient rest the night before the experimental sessions. During each day of the training and experimental sessions, they were offered one serving of a caffeinated beverage to help maintain their alertness.

5.4 Apparatus

The experimental setup was designed to recreate important elements of a real ATC workstation, including the main monitoring task, workload fluctuations, and unpredictable or random interruptions. The ATC simulation environment consisted of a computer with a 32-bit color monitor (25.25”x 15.8”) with a 178° horizontal and vertical viewing angle. The monitor resolution was 2560x1600 pixels at a 60Hz refresh rate. The monitor was placed on a desk facing an adjustable chair (Figure 5-4) . The workstation was equipped with a keyboard and mouse. Only the space bar, Ctrl keys and arrow keys on the keyboard and the left click button on the mouse were used to provide input. The lights in the room were turned off during all experimental sessions to ensure that the brightness levels of peripheral visual cues were accurately replicated.

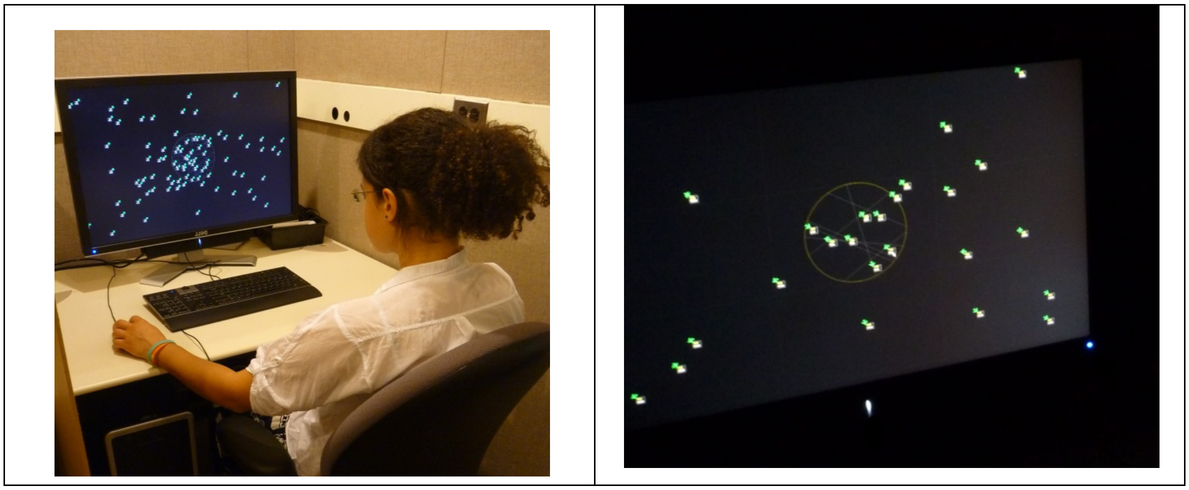


**Figure 5-4: Left: Participant seated at ATC simulation.
Right: ATC screen as seen with ambient lights turned off**

The monitor displayed the radar scope of an ATC approach controller, depicting all aircraft in his/her sector. The screen was divided into a matrix of 9 sectors and a small horizontal bar at the bottom. Each sector measured 8.4”x5.2”, and the horizontal bar measured 25.2”x1”. The sector at the center contained a yellow circle of 2.5” radius, called the hand-off region. Within this region, the participant had to ensure that airplanes were flying at the correct speed so that they could be handed off to the tower controller who would then provide guidance through landing. The eight sectors surrounding this hand-off region were demarcated by grey dashed lines (see Figure 5-5).

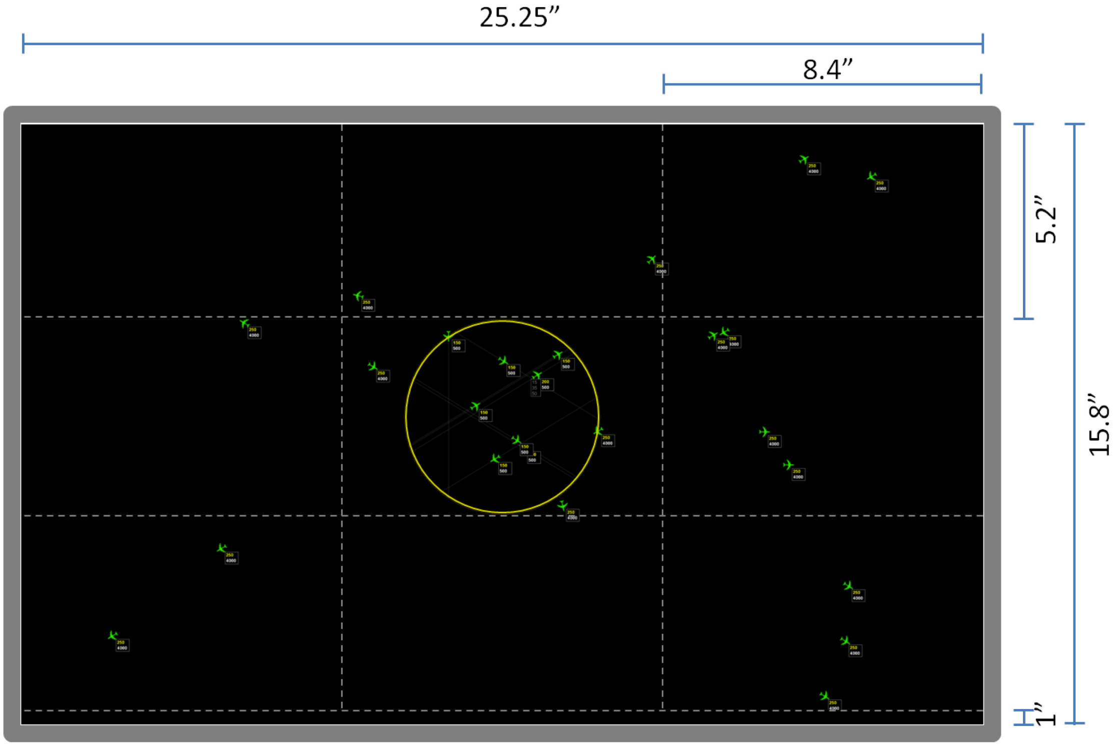


Figure 5-5: Screen dimensions (thickness of sector lines and yellow circle is exaggerated for visibility)

Airplanes were represented by a green 0.3”x0.3” airplane icon with a data block measuring 0.36”x0.33” pixels diagonally below and to the right of the plan icon. Data blocks displayed the speed (in knots) and altitude (in feet) that the airplane was flying at (see Figure 5-6). In addition, for airplanes inside the circular hand-off region, their prescribed heading was depicted by a thin grey line. All airplanes were moving across the screen at a speed of 7 inches per second in different directions. The number of airplanes on the screen was manipulated over time for the purpose of varying controller workload. In low workload conditions, 25 airplanes were present on the screen, and under high workload, 90 aircraft were depicted. The workload conditions did not affect the frequency of hand-offs or interruptions in any scenario.

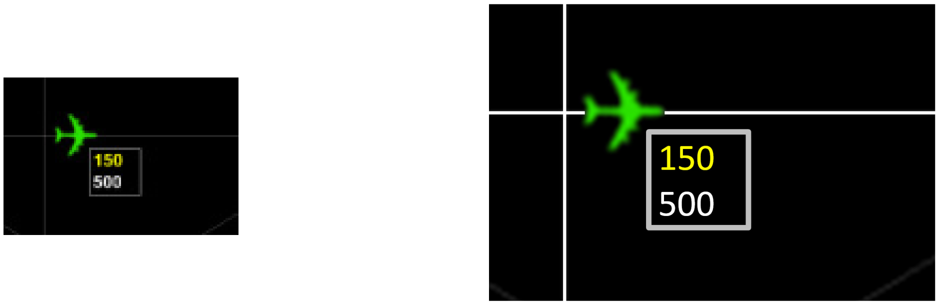


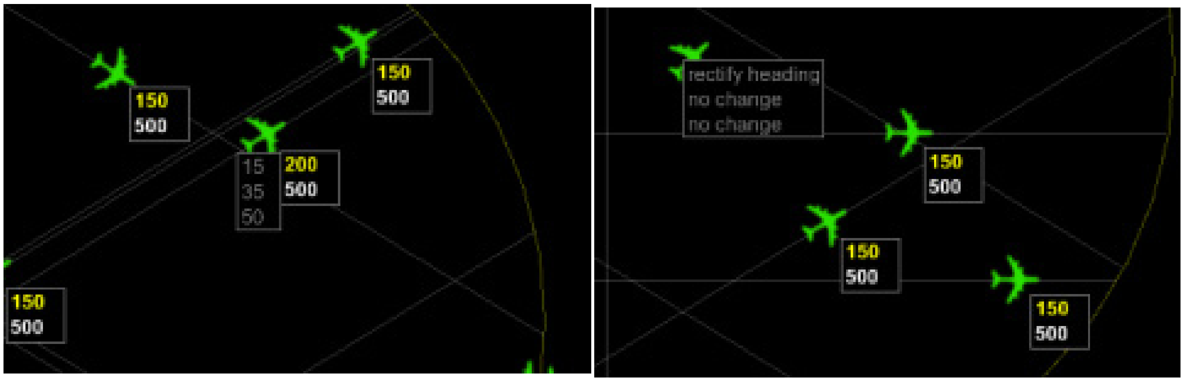
Figure 5-6: Left: Screenshot of airplane with heading (horizontal line), and data block displaying speed (top, yellow) and altitude (bottom, white). Right: Enlarged and re-touched to show heading and data block

Participants were seated in front of the ATC workstation at a distance of approximately 2ft from the radar screen. They were asked to rest against the seat back so that their relative position to the monitor remained constant throughout the sessions. Participants were fitted with an elastic neoprene wrap that held tactors firmly in contact with the front of their abdomen. They were asked to wear a layer of thin cotton clothing on which the tactors were placed. The placement of the tactors was adjusted according to the size and shape of the participant's abdomen, with a minimum of 65mm of vertical and horizontal separation between the center of the tactors. This ensured that even with different seating postures, the distance between tactors would not fall below the two-point discrimination threshold of 36mm (Weinstein, 1966).

5.5 ATC scenario: tasks

5.5.1 Primary tasks

The main task of the participants was to closely monitor the hand-off region and ensure that all airplanes in that region flew at the prescribed airspeed of 150 knots and on the correct heading. If an airplane was flying at the wrong speed, this could be detected based on information in the data block which showed 165, 185 or 200 knots. Participants had to left-click on the airplane and select the speed correction factor from a drop-down menu: 15, 35 or 50 (see Figure 5-7-a). If the airplane was flying on a wrong heading, the top-down view on the radar scope would show it deviating from its prescribed path which was indicated by a grey line. Participants had to left-click on the airplane and select the "rectify heading" option from a drop-down menu. The other two menu items were named "no change" (see Figure 5-7-b). The position of the "rectify heading" option in the menu was randomized so that the reaction time to choose the correct menu option was more comparable to the reaction time taken to choose the speed correction factor for the other main task. Airspeed or heading deviations occurred approximately once every 7 seconds.



(a)

(b)

Figure 5-7: Drop-down menus for (a) rectifying airspeed and (b) heading deviations

5.5.2 Interruption tasks

All airplanes in the participant's sector were supposed to fly at an altitude of 4000 feet. However, airplanes in the sectors surrounding the hand-off region sometimes committed an altitude deviation by dropping to an altitude of 3000, 2000 or 1000 feet, prompting an interruption of the main ongoing task. These interruptions were randomized in a controlled manner and occurred once every 36, 42 or 48 seconds so that their occurrence was somewhat unpredictable. Operators were instructed to acknowledge the interruptions immediately upon noticing them. In the absence of peripheral visual or tactile notifications, interruptions could only be spotted by scanning the eight sectors occasionally and looking for altitude deviations in the airplane data blocks. Altitude levels were displayed at the bottom of the data block in white. An altitude deviation was considered gentle if the altitude dropped to 3000ft, moderate when it was at 2000ft and severe when the aircraft was flying at an altitude of 1000ft. When an interruption was detected, the participant had to respond using the series of three steps listed below:

Step 1: Hit the space bar x times in quick succession to indicate that the deviation was noticed and what its severity is (if the deviation was gentle, x=1, if moderate, x=2 and if severe, x=3).

Step 2: Immediately after step 1, left-click on the sector in which the interruption is occurring.
If this step is not performed within 1000ms after completing the previous step, the response will time out and the airplane will resume normal operations.

Step 3: Left click on the airplane which is experiencing a deviation.
This step can be performed at any time before the interruption times out. This step was omitted in case of gentle deviations.

In a few cases, two altitude deviations occurred simultaneously. In those cases, the participant first had to acknowledge the concurrent onset by hitting the Ctrl button on the spacebar once and then respond to the more severe altitude deviation first, followed by the less severe deviation, using the same three steps for each case.

Once all the required steps were completed correctly, the deviation was corrected and the airplane resumed normal operations. Figure 5-8 details the program responses for all the combinations of potential responses and the associated points awarded to participants when a minor deviation occurred.

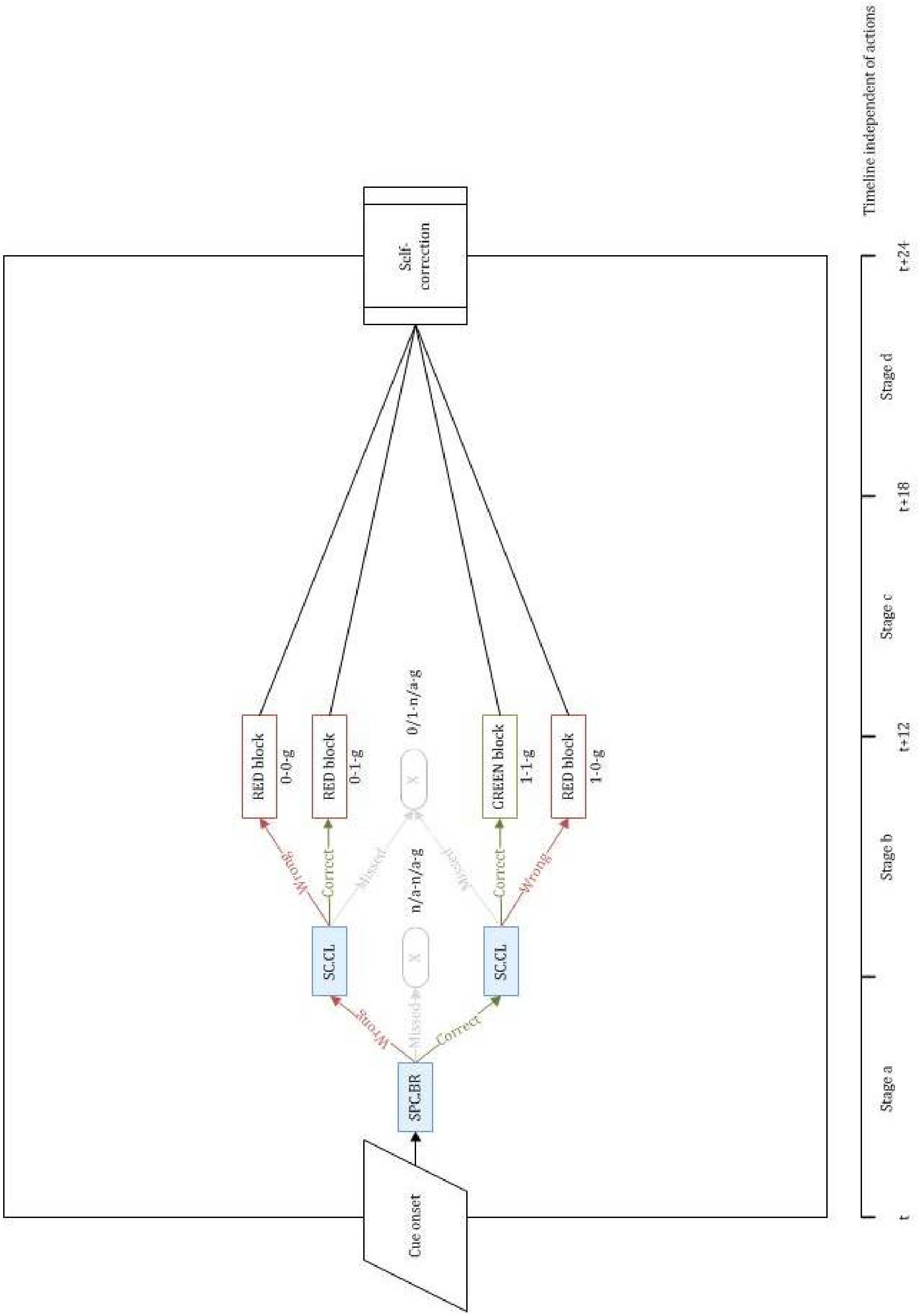


Figure 5-8: Potential participant responses and award scheme for minor altitude deviations

In case of moderate or severe altitude deviations, participants were required to complete a secondary task. This was done to simulate the nature of air traffic controllers tasks. Controllers occasionally need to attend to lengthy interruptions that are situated away from the primary task location. The secondary task was presented at the bottom of the screen (referred to here as the 'task bar'), outside the sector. The task bar would get populated with a row of 35 'X' symbols with three '+' symbols hidden among them in randomized locations. Participants were instructed to find and click on all the three '+' symbols to complete the task. When a '+' was correctly clicked on, it turned green (see Figure 5-9).

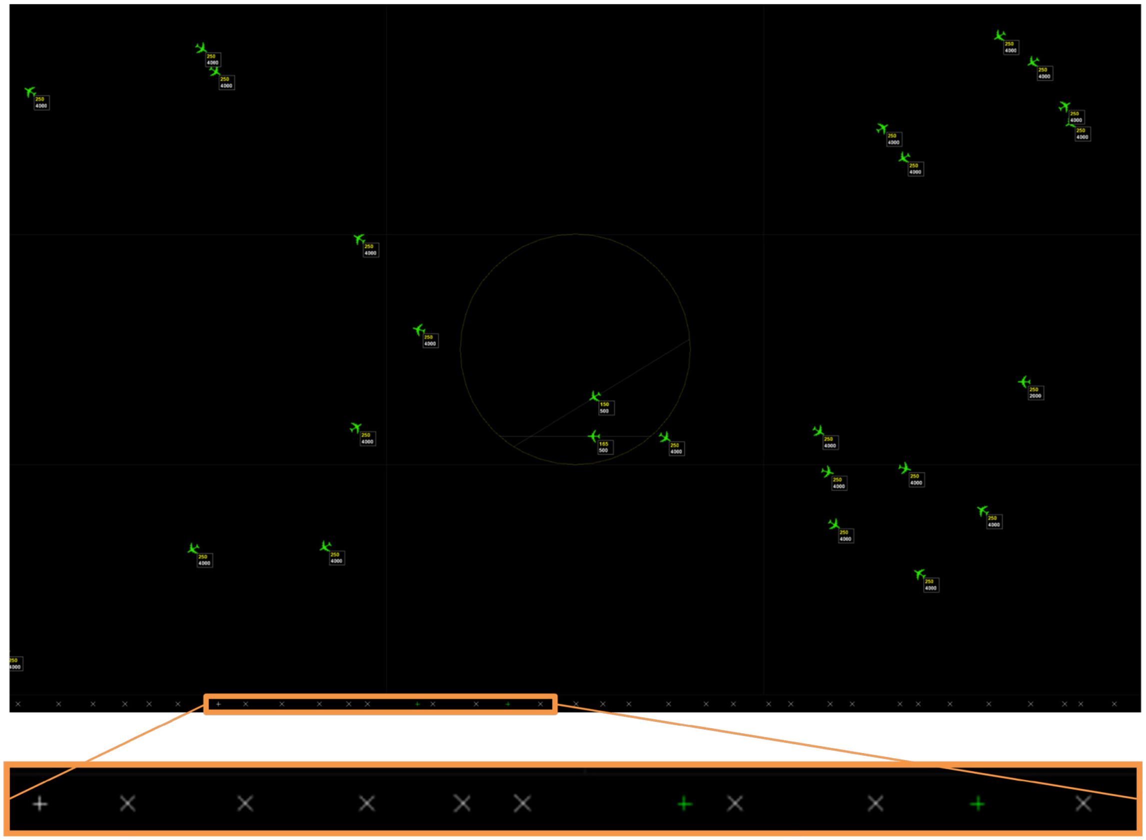


Figure 5-9: Additional task (detecting 3 '+' symbols hidden among 'x' symbols) in case of moderate and severe altitude deviations

The required response to the above task differed depending on the severity of the altitude deviation. If the task bar appeared after a moderate deviation, participants could postpone the task if necessary until their workload was relatively low. However, if it appeared after a severe deviation, participants were required to complete the task before they could resume their main task, since the rest of the screen was locked until the last of the three '+'s was clicked on. The task bar routine was designed to mimic the time constraints on an ATC operator that would vary depending on the severity of a problem situation being handled. Figure 5-10 and Figure 5-11 detail the program responses for all combinations of potential responses, including treatment of the secondary task, and the associated points awarded to participants when moderate or severe incidents occurred.

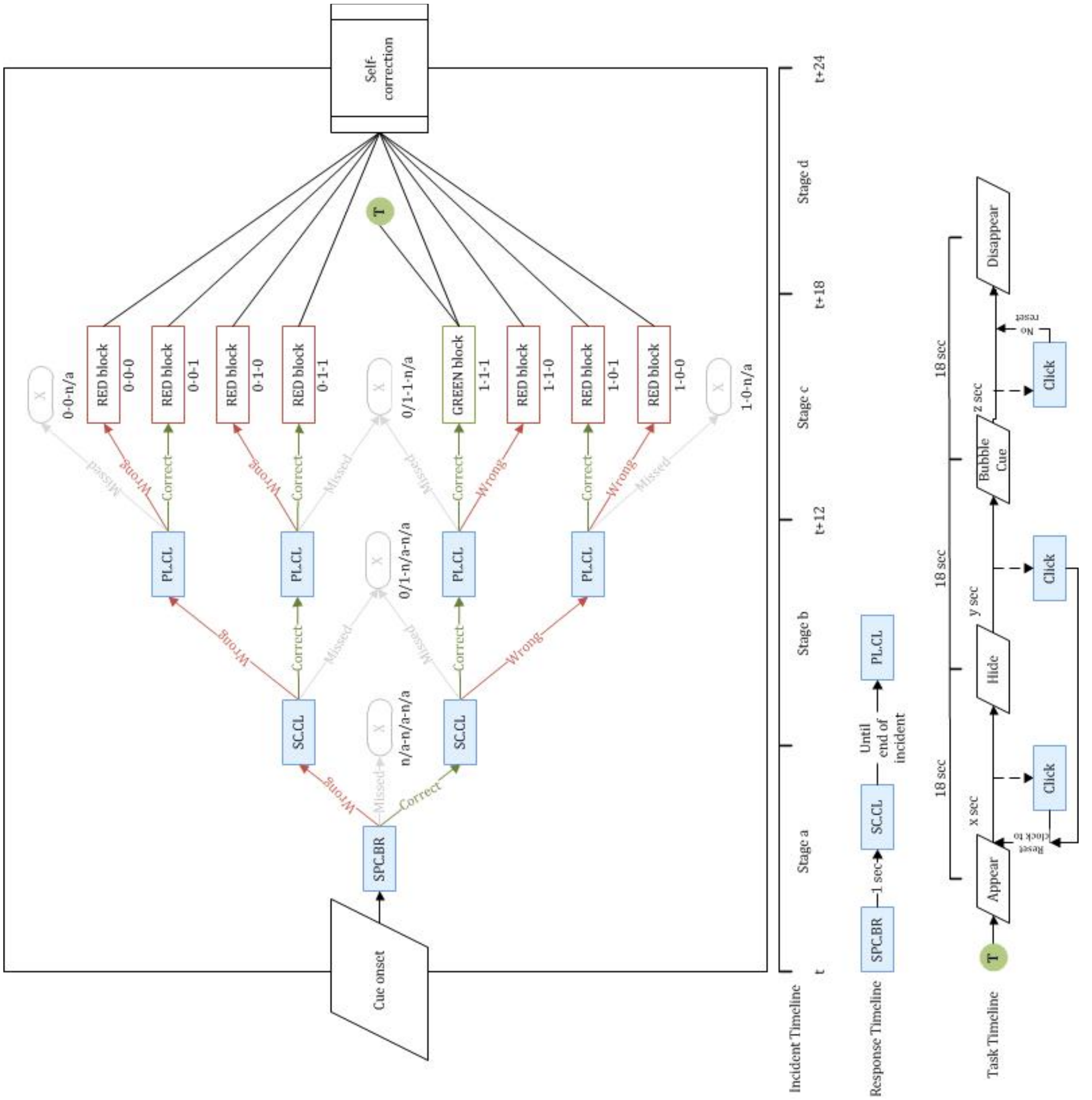


Figure 5-10: Potential participant responses to Moderate incidents and points awarded

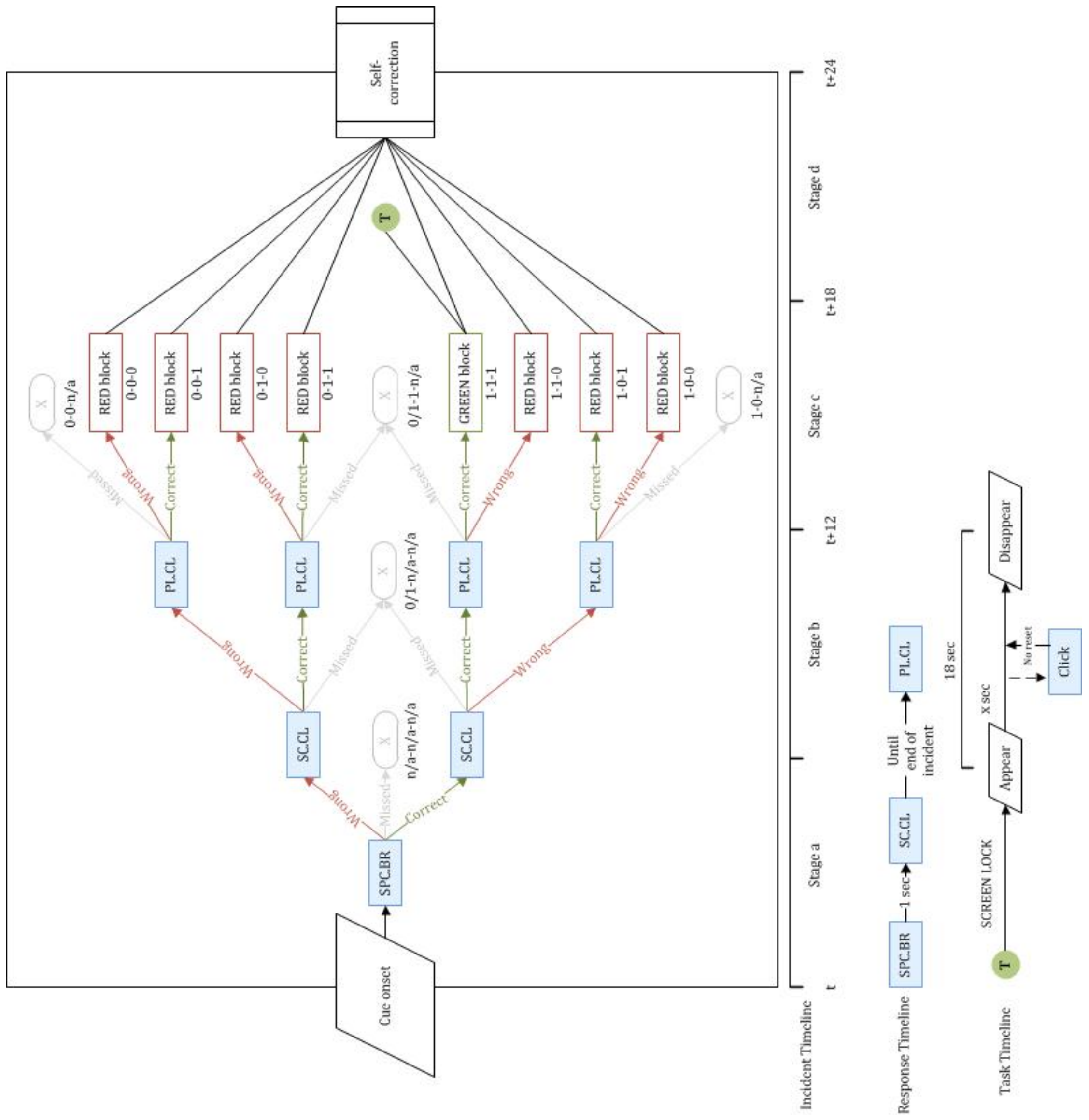


Figure 5-11: Potential participant responses to Severe incidents and points awarded

5.6 Procedure

The experiment was conducted over a two-day period. Participants were trained to use the simulator on the first day and performed the actual experiment on the second. On the first day, participants answered a background questionnaire (see Appendix I) and then familiarized themselves with the ATC simulation. Instructions were read from a script (see Appendix II) to ensure that all participants received the same information. As part of the familiarization, participants completed the following 8 tutorial sessions over a period of approximately 52 minutes, with an optional break in-between:

1. performing main tasks (two 2 1/2-minute sessions, one with assistance, one without),
2. understanding incidents (2 minutes),
3. responding to single incidents with the help of four-stage and graded peripheral visual cues (two 4:40 minute sessions each),
4. responding to single incidents with the help of four-stage and graded tactile cues (two 4:40 minute sessions each),
5. responding to single incidents with the help of graded multimodal cues (6 minutes)
6. responding to concurrent incidents (6 minutes),
7. responding to single but dynamic incidents (6 minutes), and
8. responding to concurrent dynamic incidents (8 minutes).

After completing the tutorials, participants were shown a point-system sheet (see Table 5-2) that detailed the rewards and penalties for possible responses to the various scenario events. There were a total of 1320 primary tasks, 192 incidents, out of which 48 were concurrent 162 secondary tasks, but participants were not informed of this beforehand. Participants were only informed that a balanced performance on main tasks and interruptions was rewarded more than preferential performance in any one of the tasks. This was done to ensure that participants tried to timeshare as much as possible and did not prioritize one task over another. Participants were informed that there was a $10 bonus for the top 50th percentile scores. This was added incentive for participants to show balanced performance since they could not predict the performance capabilities of other participants and hence had to do their best to earn the bonus. Performance statistics for the main tasks and the interrupting events, as well as total points earned were displayed on the screen at the end of each session and were reviewed with the experimenter. The introduction and tutorials lasted approximately an hour and 45 minutes. Participants then completed four 12-minute comprehensive training sessions which included all elements of the simulated ATC operations. Each session featured all five types of notifications in random order, similar to the final experiment. This was done to ensure that participant performance did not exhibit a learning curve in the final experiment. A performance debriefing was conducted at the conclusion of the training sessions (see Appendix III). Completion of the above activities on the first day took between 120 to 150 minutes, depending upon the number and duration of breaks and the length of debriefing.

	Reward	**Penalty**
Main task (handoff)		
Correct	**1**	
Miss		**-0.5**
Incorrect		**-0.25**
Notification acknowledgment (space bar + sector click)		
Timely correct acknowledgment	**1**	
Total miss		**-1**
Wrong acknowledgment between gentle/moderate incidents		**-0.25**
Wrong acknowledgment involving severe incident		**-0.5**
Correct identification of dual cues (Ctrl click)	**0.25**	
Wrong prioritization involving severe incident		**-0.5**

Wrong prioritization between gentle/moderate incidents		-0.25
Postponed acknowledgment	*No action*	
Cue response (block click)		
Timely response (correct)	0.5	
Miss (correct acknowledgment but no click)		-0.5
Wrong (click without associated acknowledgment)		-0.25
Completely forgotten postponement		-0.5
Secondary task		
Completed task (identified all 3 '+' marks)	2	
Incomplete task		-2

Table 5-2: Point system for participant performance

On the second day, participants first recorded their fatigue levels on a Likert scale (these levels were later compared with fatigue levels measured at the end of the experiment) and then completed six review sessions of three minutes each. There was one session for each of the five different cued conditions and one without any notifications of interrupting events. These sessions gave participants a chance to refresh their memory of the various tasks and cue types before the experiment. Next, participants completed one uncued and four cued experimental sessions with optional breaks in between. The cued sessions were 28 minutes long each, and the un-cued session took 22 minutes to complete. The order of the unced session was varied between participants such that it was presented first or third or fifth in the in session order. After completion of the experimental sessions, the experimenter conducted a 20-minute debriefing to record participants' fatigue levels and to gain insight into difficulties they experienced with the tasks/interface, their stress levels, and the strategies that they adopted to perform the experimental tasks. Completion of the activities on the second day took anywhere between 150 - 180 minutes, depending upon the number and duration of breaks and the length of the debriefing. Participants were compensated an average of $15 per hour for both days.

The order of presentation of the 5 experimental sessions was randomized with the un-cued session positioned either at the beginning, middle or the end of the session sequence, as described in the previous section. Randomizing the order of experimental sessions served to minimize potential learning effects (in addition to the extensive training on day 1) and potential effects of fatigue. Workload varied within each experimental session. The duration of the level of workload is schematically represented in Figure 6-11.

Figure 5-12: Workload changes within each experimental session

Each participant was presented with 32 interruption tasks in the uncued and 40 events in the cued session, for a total of 192 interruptions over the entire experiment. There was one uncued session and four cued sessions featuring a randomly ordered mixed of the five notifications types: four-stage visual, four-stage tactile, graded visual, graded tactile, and crossmodal graded in each session. This mix of notification types ensured that participants could not orient their attention to a particular modality in anticipation of a notification in that modality. Therefore, every cued session had 40 incidents and the uncued session had 32. The ratio of single incidents to two events occurring concurrently was 3:1. The ratio of gentle, moderate, and severe deviations was 1:2:1. The complete experimental designs for all five sessions are shown in Appendix V.

5.7 Experimental design and analysis

The experiment consisted of five timed experimental sessions, four cued and one uncued. Notification type (6 levels: uncued, four-stage visual, four-stage tactile, graded visual, graded tactile and crossmodal graded), workload (2 levels: low and high), and intensity (3 levels: gentle, moderate and severe) were the primary independent measures and were varied within participants (see Appendix V for details). The dependent measures in this experiment were response times and accuracy of responses to interrupting events. The data was analyzed using repeated measures ANOVA, using PASW (formerly known as SPSS) Statistics 17. Significant effects were further analyzed using post-hoc Fisher's Least Significant Difference (LSD) tests to identify differences between factor levels. Quantitative data from debriefing was analyzed using non-parametric Friedman's ANOVA; qualitative data was interpreted using an inductive analysis protocol.

Chapter 6 Evaluation of Multimodal Graded Notifications

6.1 Results and discussion

The following sections show the results from the statistical analyses of the data from the final experiment. The abbreviations V, T, GV, GT and GVT will be used henceforth to denote 4-stage peripheral visual, 4-stage tactile, graded peripheral visual, graded tactile, and graded multimodal notifications respectively.

6.1.1 Goal 1: Improve performance in detection of, and response to, interruptions

Prediction 1: Detection rates and accuracy for interrupting tasks are higher, and response times are faster, in all cued conditions

Prediction 2: The highest accuracy and detection rates are observed with graded cues

All five types of notifications were expected to positively affect the detection rate and accuracy for interrupting tasks (i.e., detecting and correcting altitude deviations), compared to the uncued condition. The highest level of performance was expected for graded notifications because (a) they present information in two under-utilized modalities of peripheral vision and touch to enable cues to be processed in parallel with ongoing tasks, and (b) they do so in increasing levels of salience to ensure that cues are reliably detected. Here, detection rate is defined as the percentage of interruptions that were detected and acknowledged by the participants. Accuracy refers to the percentage of interruptions that were *correctly* responded to (i.e., the complete sequence of required steps was performed - including hitting the space bar the correct number of times, clicking on the appropriate sector, and clicking on the deviating airplane).

Average detection rates were first compared between cued and uncued conditions (see Figure 6-1), using repeated measures ANOVA. Detection rates were found to be significantly higher in the cued condition (77.12%) than in the uncued condition (43.87%) (F(1,30) = 267.075, p < 0.001).

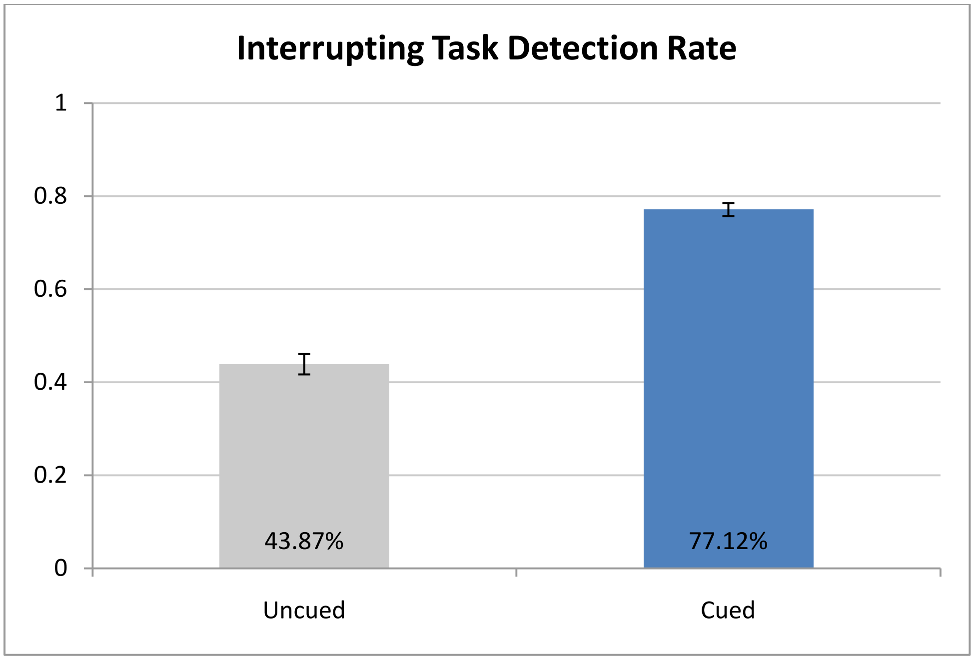


Figure 6-1: Interrupting task detection rates (mean and standard error) during cued and un-cued conditions

Detection rate was computed and compared across all 6 experimental conditions and was found to differ significantly (F(5,150) = 114.316, $p < 0.001$). Detection rates were significantly higher in each of the cued conditions than in the un-cued conditions ($p < 0.001$ for all pair-wise comparisons with cued conditions) (see Figure 6-2).

As expected, performance in GVT was significantly higher than in all other cued conditions, with a detection rate of 81.6% ($p < 0.005$ for all pair-wise comparisons with other conditions). Among the single-modality cues, V and GV resulted in the highest detection rates (V: 78.6%, GV: 78.3%) which weren't significantly different from GT (76%). However, T resulted in the lowest detection rates (71.2%) among all cued conditions ($p <= 0.005$ for all pairwise comparisons).

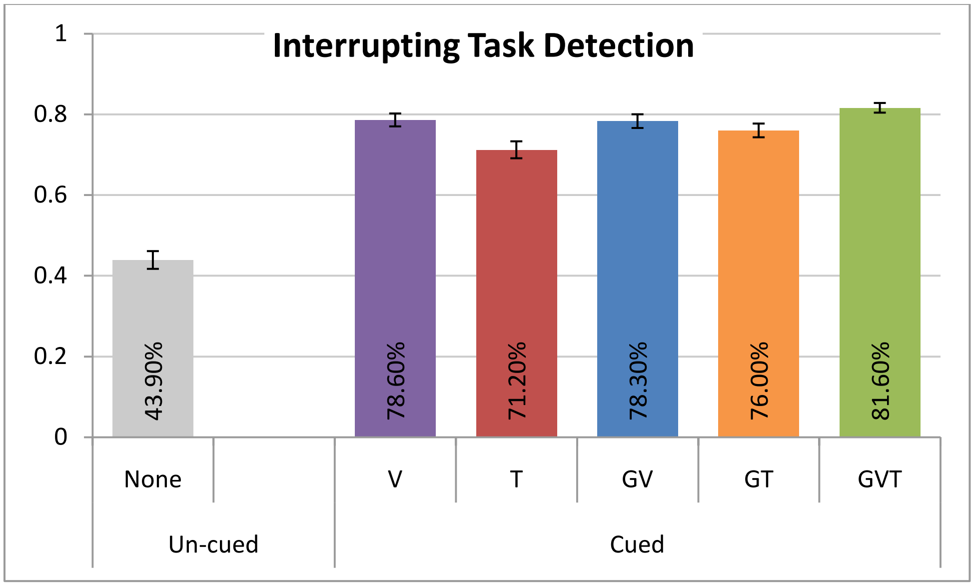


Figure 6-2: Interrupting task detection rates (mean and standard error) for all cue conditions

It was assumed that one of the reasons for the high detection rate for GVT resulted from the switch in modality from peripheral vision to touch. To verify this assumption, the stages at which participants responded to notifications were scrutinized (see Figure 6-3). The detection rates for each notification type are depicted on a centum scale. The response stages in which detection occurred are depicted as percentages of total detection. As predicted, detection percentage was highest in the first response stage. Initial detection (detection in stage a) was highest for V (77%) and lowest for GT (47%). Accordingly, the spread of detection across the stages was lowest for V and highest for GT.

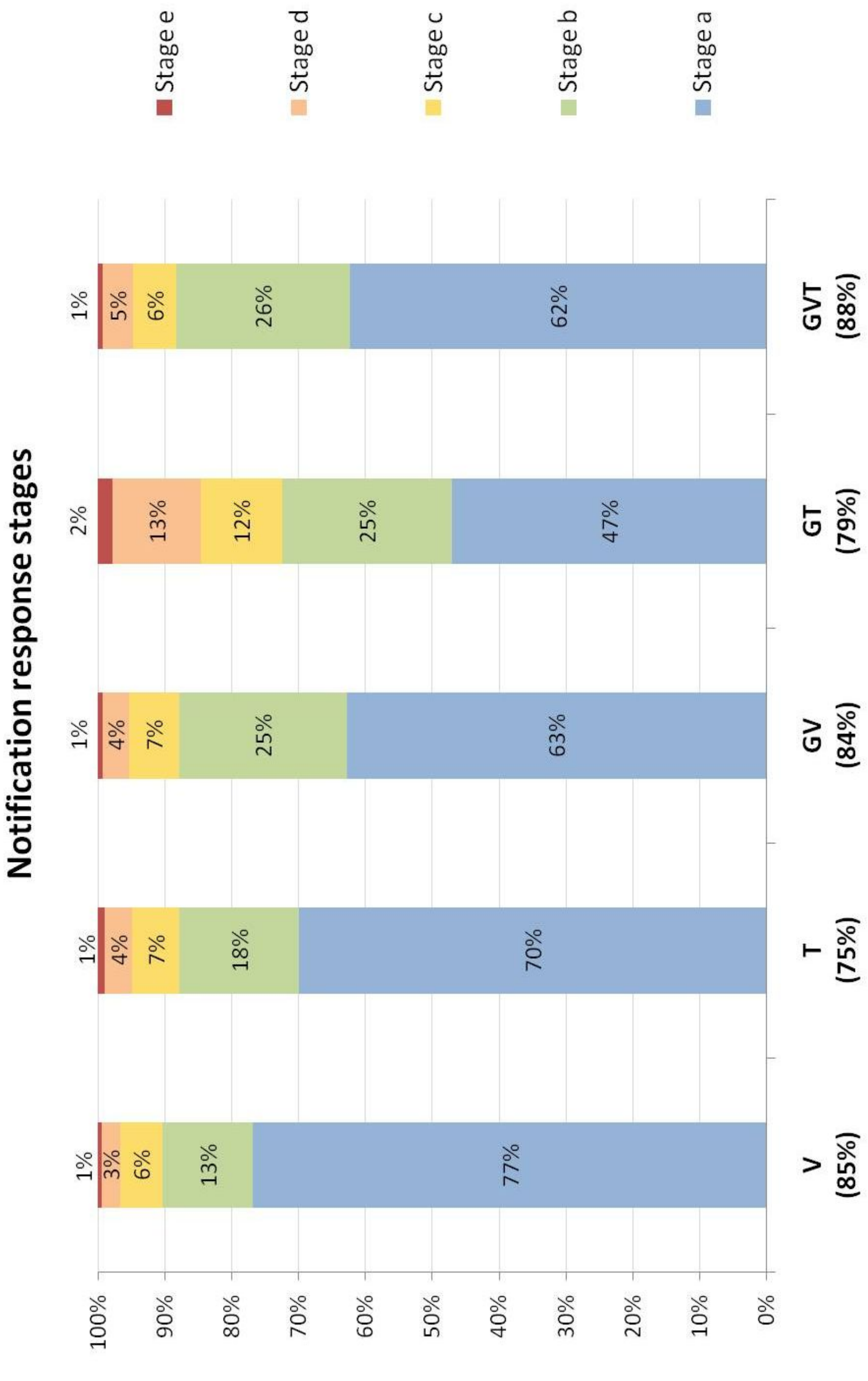


Figure 6-3: Response stages in which notifications were detected. Values on the X-axis denote notification type and mean detection rate. Data labels on histogram show distribution of detection across 5 stages a, b, c, d, and e.

Accuracy refers to the percentage of interruptions that were *correctly* responded to, i.e., the complete sequence of required steps was performed. This includes hitting the space bar the correct number of times, clicking on the appropriate sector, and clicking on the deviating airplane. Accurate detection of notifications was first compared between cued and uncued conditions (Figure 6-4). Accuracy in the uncued condition was significantly higher (94.80%) than in the cued conditions (86.6%) (F(1,30) = 21.745, p<0.001) which is not surprising. In the cued conditions, participants had to interpret interruption signal parameters in parallel with performing ongoing tasks in order to respond to them accurately. However, in the uncued condition, participants looked directly at the interrupting airplane to report the details of the interruption.

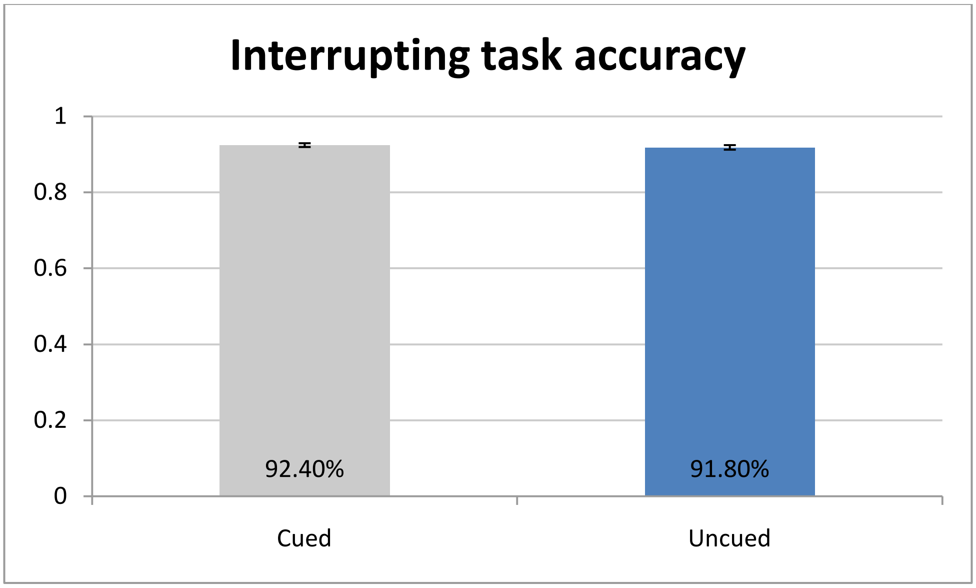


Figure 6-4: Interrupting task accuracy (mean and standard error) during cued and un-cued conditions

Among the cued conditions (see Figure 6-5), notifications featuring visual cues had better accuracy (GVT = 95.7%, V = 96.1%, GV = 96.2%) than those with only tactile cues (GT = 90.6%, T = 90.8%). Since accuracy is determined by the identification of the correct location of interruptions as well as the correct interpretation of the intensities, the means for these two measures were calculated and then compared across the different cue conditions (Figure 6-6). Both measures were significant (F(5,150) = 76.065, p < =0.001; and F(5,150) = 65.434, p < 0.001, respectively). The location accuracy followed the same trend as overall accuracy: visual notifications (V=96.1%, GV=96.2%, GVT=95.7%) fared significantly better than tactile notifications (T=90.8%, GT=90.6%). This reconfirms the theory that visual spatial discrimination is superior to tactile spatial discrimination (Intriligator & Cavanagh, 2001; Johnson & Phillips, 1981; Moy, Singh, Tan & Fearing, 2000; Sarter, 2002). No such significance was seen in the intensity accuracy, although there was a trend that showed that accuracy was higher in graded conditions than in the four-stage conditions. Highest intensity accuracy was observed in GVT, which shows that the combination of modalities were instrumental in providing a clear indication of intensity rather than the use of any single modality.

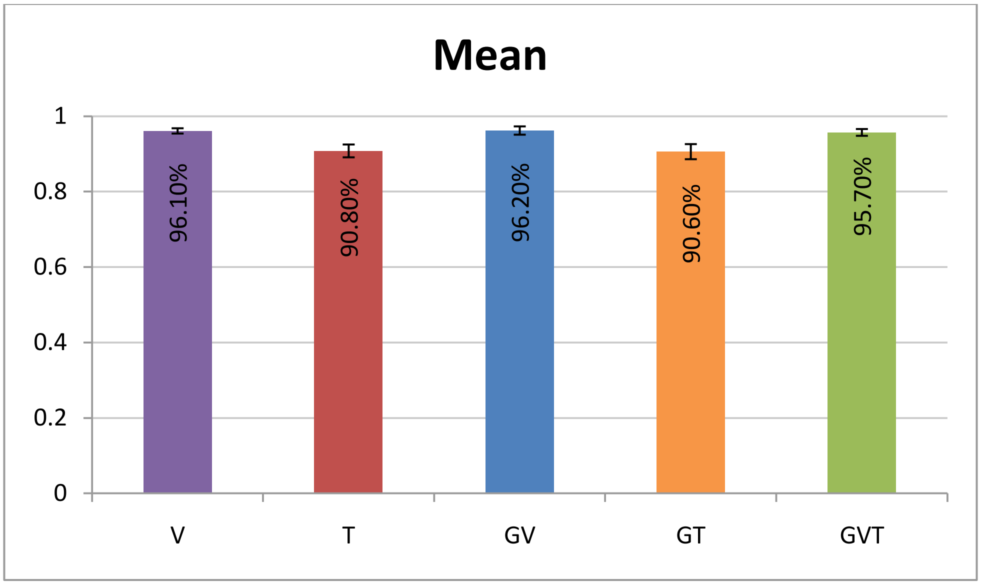


Figure 6-5: Interrupting task accuracy (mean and standard error) for all cue conditions

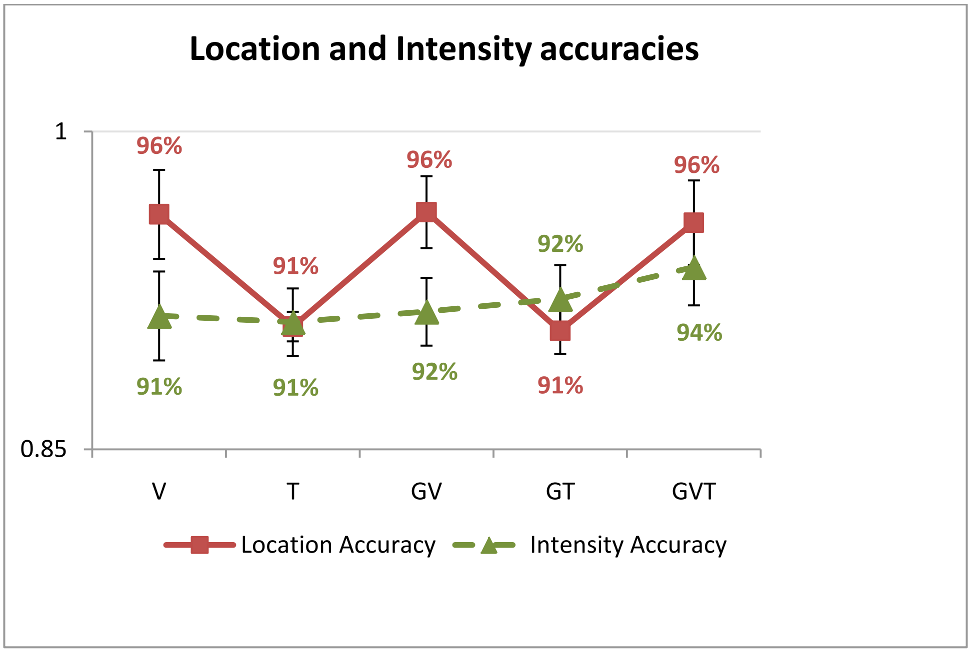


Figure 6-6: Location and intensity accuracies (mean and standard error) for interrupting tasks in all cued conditions

Discussion

Predictions 1 and 2 were confirmed. The high detection rates seen in cued conditions can be attributed to the addition of effective exogenous attention orientation mechanisms that reduce the need to rely on purely endogenously driven orientation. Because the interruption signals employ underutilized modalities and can be processed with minimal interference with the primary task, the cues support more efficient switching between primary and interrupting tasks, and thus better performance on both.

Among the single-modality notifications, V, GV and GT were all detected more often than T. The significant difference between the T and GT conditions can be explained by the fact that the detection of tactile notifications is not entirely dependent on the salience of the notification alone, but also on the gradation. The discrete steps in salience from low to medium to high frequency in the GT seem to have had a favorable effect on the detection of interruptions. As predicted, the highest detection rates and accuracy measures were found with GVT notifications. However, the modality switch itself (peripheral vision to touch) does not explain this high detection rate.

Initial detection (detection in stage-a) was expected to be higher for notifications that featured peripheral visual cues in stage-a (V, GV and GVT) than notifications that did not (T and GT). Except in cases of extreme cognitive load, the powerful and involuntary orienting power of peripheral vision is strong enough to capture attention. Among V, GV and GVT, highest detection was seen in V since stage-a salience for V is higher (16%) than stage-a salience for GV or GVT (13% each). Between T and GT, initial detection was lowest for GT since cue salience at stage-a was lower (70Hz) than T (105Hz).

Among the graded notifications, it was expected that GV and GVT would have similar detection rates during the first two stages since the cues employed in these notifications were identical. This can be confirmed from Figure 6-3 which shows that 88% of the cues were detected within the first two stages. It was also expected that the detection rates in stage-c would be higher in GVT than in GV since GVT features a modality shift from peripheral vision to touch between stages b and c. However, the results do not support this expectation. One possible explanation is that there was a ceiling effect affecting detection at the later stages (c, d and e). Since a very large percentage of notifications were detected early on, there were few opportunities for observing the effect of later cues on the detection rate.

Despite the apparent lack of the effect of modality shift, GVT still featured the highest overall rate of detection. There are two possible explanations for this finding. First, it is known that peripheral vision is susceptible to the phenomenon of attentional narrowing, especially during high cognitive loads. Since GVT employs tactile cues in the latter half of the notification, it may be unaffected by this phenomenon at the later stages and hence is able to capture the attention of the operator eventually. Although this deduction is based on attention literature (Mackworth, 1965; Easterbrook, 1959), it cannot be confirmed in the context of this study because the use of eye-tracking was not employed in this study. Second, since the tactile modality is a proximal sense and is applied directly to the surface of the skin, tactile cues of reasonably high salience can capture attention even during high cognitive load. Accuracy of detection was far higher for visual notifications than for tactile notifications. Although visual spatial discrimination is superior to tactile spatial discrimination, there are other potential reasons for the observed differences in accuracy.

During the experiment debriefing sessions, some participants expressed difficulty in distinguishing the spatial location of tactile cues presented from adjacent tactors. This could be the result of several possibilities. (a) Some participants experienced trouble in perceiving the tactors that were placed below the navel. The arrangement of the tactors was based on the findings of Cholewiak, Brill, and Schwab (2004) which tested locations above the navel for sensitivity and accurate localization of vibrotactile stimuli, but not below. Perhaps the thickness of lower belly tissues affected these two measures. However, the same study also found that there was no effect of differences between the skin over upper belly tissue, muscle tissue, and the ribs, therefore it is unlikely that the skin over the lower belly tissue alone would be an exception. Nevertheless, it will be interesting to explore this possibility in future research. (b) The clothing worn by participants differed in thickness. Although all participants were requested to wear "a thin upper garment", there were differences not only in the thickness of the clothing, but also in the material itself. This potential effect could have been avoided by providing participants with upper garments of the same material and thickness. (c) The tactor harness used in this study was flexible, however, it did not tightly and evenly wrap around the participants abdomen. The gaps between the different contours of the abdomen and the relatively straight surface of the tactile harness, particularly exacerbated if participant was slouching, might have caused a difference in the force applied by each tactor on the skin, therefore causing a difference in the sensitivity and localization of the tactile stimuli.

To address the spatial discrimination problems with tactile presentations, two further courses of action can be examined in addition to the possibilities mentioned above: (1) spatially arrange tactors on participants based on their individual tactile sensitivity threshold and localization instead of assuming common values for all participants, and (2) explore alternative methods to secure tactors on the skin of participants for long periods of time without the use of cumbersome devices so that other body sites, such as the back and/or other locations on the limbs, can be used effectively.

Prediction 3: Response times to interrupting tasks are faster in cued conditions

Response time was measured as the time from the onset of an interruption to the time of acknowledgment (when participant hits the space bar). Response times were compared between cued and uncued conditions using repeated measures ANOVA. Response times to interrupting tasks were significantly shorter for the cued conditions (mean = 6.598s), almost half the time taken compared to uncued conditions (mean = 11.284s), as seen in Figure 6-7 ($F(1,30) = 121.912$, $p < 0.001$).

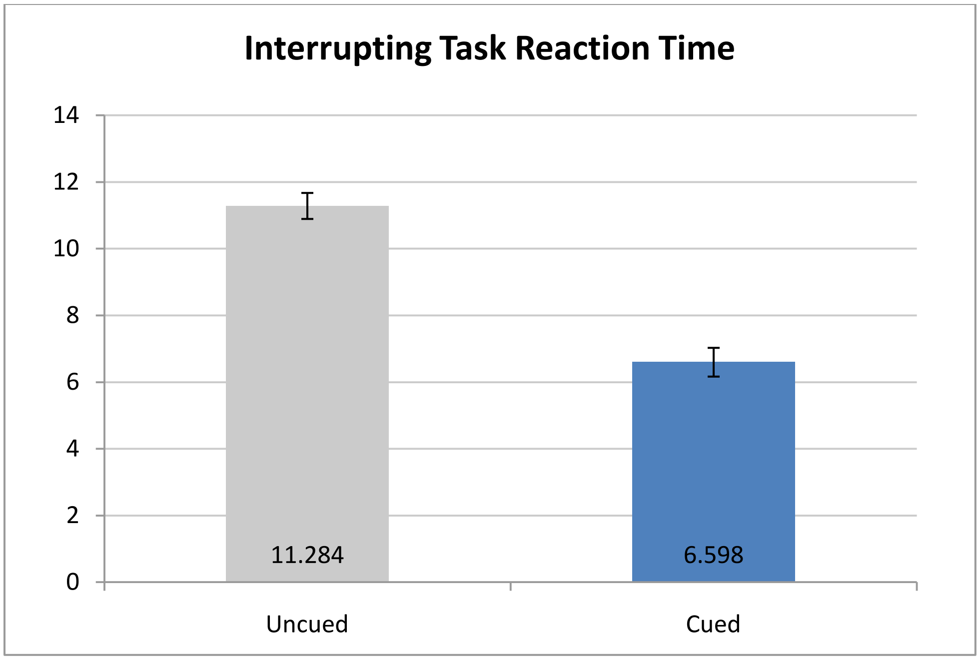


Figure 6-7: Interrupting task reaction times (mean and standard error) for all cued and uncued conditions

Average response times were also compared across the uncued conditions and cued conditions with each notification type, and were found to differ significantly ($F(5,150) = 44.794$, $p < 0.001$) (see Figure 6-7). As expected, mean response time in the uncued condition was the longest (11.28s) and was significantly different from all cued conditions ($p < 0.001$ for all pairwise comparisons). Among the cued conditions, mean response time was shortest for V (5.20s) and longest for GT (9.02s), both of which were significantly different from all other conditions ($p < 0.012$ for all pairwise comparisons for V; $p < 0.06$ for all pairwise comparisons for GT). Response times did not differ significantly between T, GV and GVT.

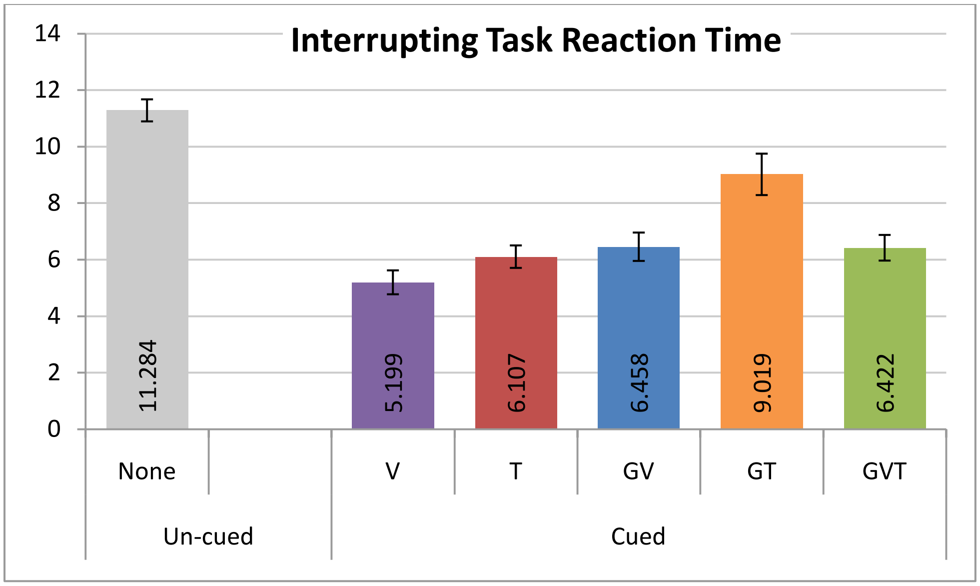


Figure 6-8: Interrupting task reaction times (mean and standard error) for all cue conditions

Discussion

Prediction 3 was confirmed. The time taken for a participant to respond to interruptions in the uncued condition directly depends upon the likelihood of the interruption coinciding with the endogenous shift in the participant's attention. In other words, for response times in uncued conditions to be comparable with cued conditions, participants would have to shift their attention from their primary task to scan the sector for interruptions every five or six seconds. Since participants were instructed to prioritize their primary monitoring task, they did not engage in such rapid shifts in attention that could have been costly to their primary task performance (Trafton & Monk, 2008). Therefore the endogenous shifts of attention should have occurred at random, resulting in the observed reaction time of 11.284 seconds is close to this expected value.

Comparisons of reaction times within cue modalities show that the non-graded cues (V and T) elicited shorter reaction times than their graded counterparts. This is possibly due to the design of the cue saliences. The first 6 seconds of the non-graded notifications are presented with higher salience levels (16% brightness for V and 105Hz for T) than those in graded notifications (13% brightness for GV and GVT and 70Hz for GT). Therefore, considering all other things to be equal, the expectation that non-graded notifications would be detected earlier than graded notifications is observed in this experiment. Finally, the response time to V was significantly shorter than the response time for all other cued conditions ($p < 0.01$ for all pair-wise comparisons).

6.1.2 Goal 3 : Improve performance of interruption tasks under cognitive load

The need for, and the benefit of, data-driven attention guidance (in this case, in the form of interruption notifications) is greater during high workload conditions. In the present study, an increase in workload was achieved by either increasing the number of airplanes in the sector or by presenting multiple concurrent interrupting tasks. The following predictions were made concerning the effectiveness of notifications under these greater cognitive loads.

Prediction 5: Graded interruption cues result in better performance (detection rates and accuracy) than non-graded ones, especially during high workload

Average detection rates, accuracies and reaction times were calculated and compared for the different cue conditions under low and high workload conditions using repeated measures ANOVA.

Detection rate

Overall, detection rates were significantly higher in low workload conditions (75.7%) than in high workload conditions (67.9%) ($F(1,30) = 56.92$, $p < 0.001$). A significant interaction effect on detection rate was observed between notification modality and workload ($F(5,150) = 52.214$, $p < 0.001$), indicating that the notification modalities were differently affected by changes workload. Specifically detection rates for V and GVT were significantly different from each other and from all other effects (all $p < 0.05$). The lowest detection rate was found in the uncued condition under high workload (mean=24.67%) while the highest detection rate was observed for V under low workload (84.4%). A close second were detection rates with GVT under high workload (mean=83.87%) (see **Error! Reference source not found.**).

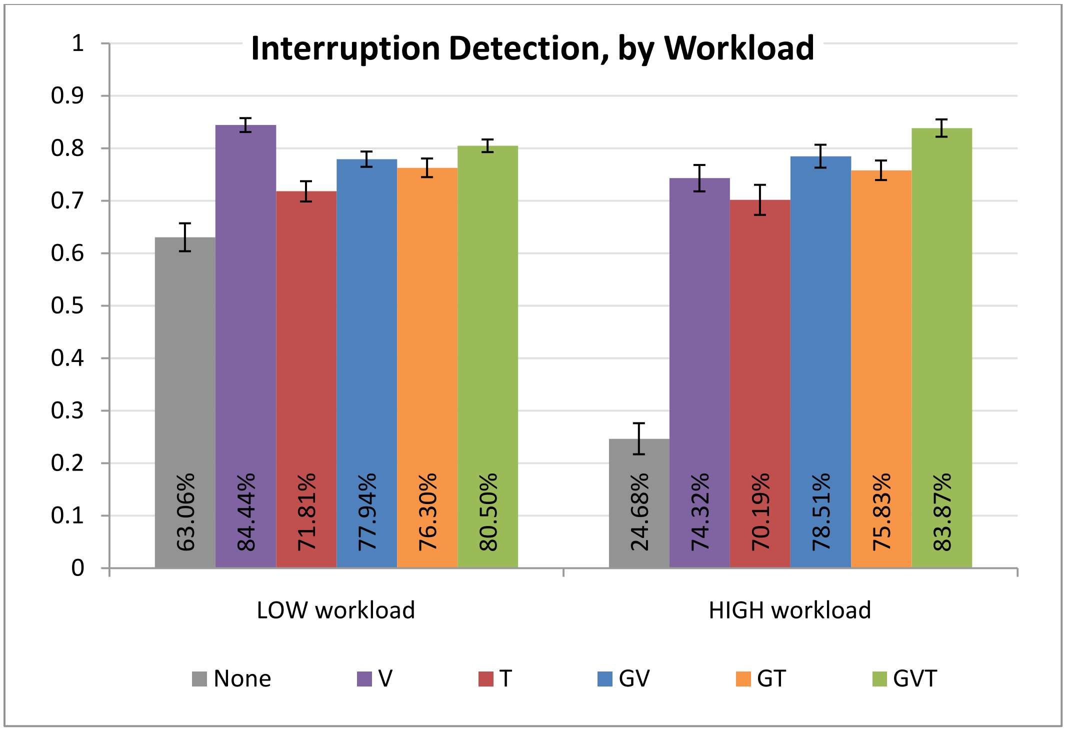


Figure 6-9: Interrupting task detection rates (mean and standard error) for all cue conditions in low and high workload conditions

Tukey's LSD post-hoc tests were used to determine differences between means for significant effects. Pairwise comparisons between the factor combinations revealed significant differences between performance in low and high workload for three out of the six cue conditions (see Figure 6-10). A significant drop in detection rate between low and high workload situations was observed in the no-cue condition (drop from 63.06% to 24.68%; $p < 0.001$) and in V (drop from 84.44% to 74.32%; $p < 0.001$). However, there was no significant difference between performances in low and high workload situations in GV, T and GT.

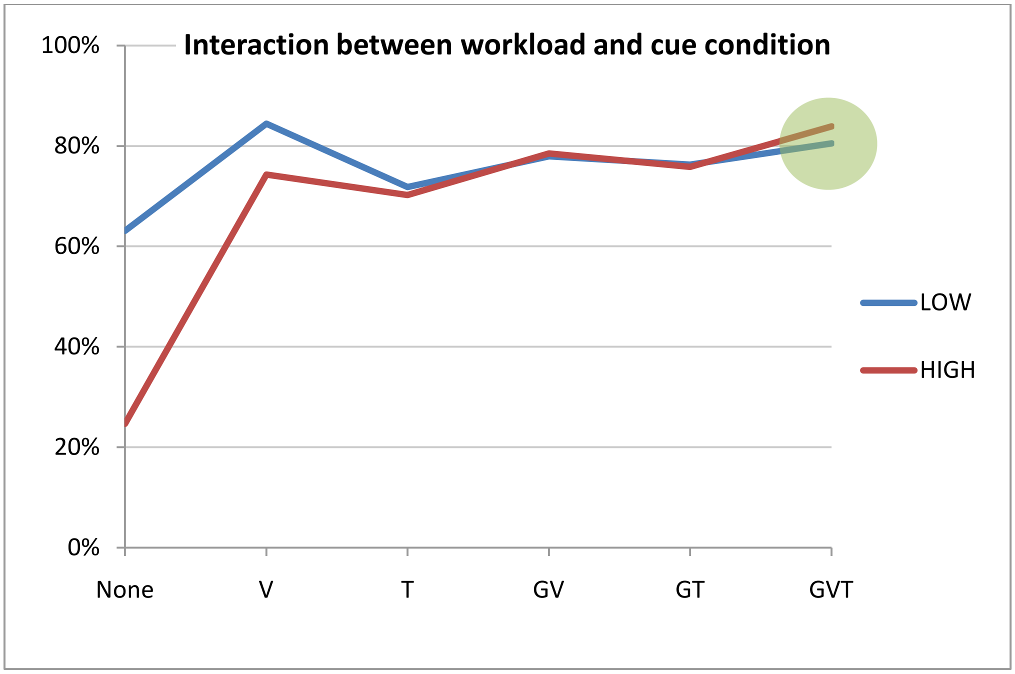


Figure 6-10: Interaction effect of workload and cue condition on detection rate

Finally, a significant *increase* was observed in the GVT condition between low and high workload situations (from 80.5% to 83.8%; p = 0.034) which further strengthens the case for using graded multimodal cues.

Accuracy

In addition to ensuring fast and reliable detection of notifications, it is important that participants can accurately interpret the information that is conveyed about the interrupting task. The comparison of accuracy across cue conditions yielded similar results to those found for detection rates (see Figure 6-11).

Overall, accuracy was significantly higher in low workload conditions (82.44%) than in high workload conditions (90.3%) (F(1,30) = 68.525, $p < 0.001$). A significant interaction effect on accuracy was observed between notification modality and workload (F(4,27) = 3.862, $p < 0.031$), indicating that the effect of modality on accuracy was different for different levels of workload. An increase in workload significantly reduced the performance in four of the cue conditions (V, GT: $p < 0.001$; T: $p = 0.004$; GV: $p = 0.024$) but not in GVT, in which the decrease in performance was an insignificant 1.4%. This shows that GVT notifications are robust in the face of increased workload.

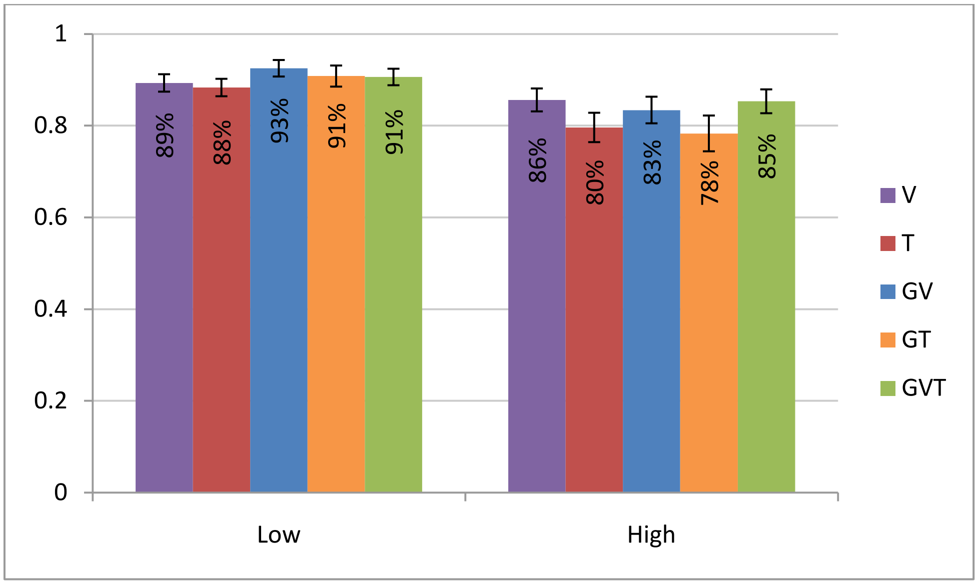


Figure 6-11: Interrupting task accuracy (mean and standard error) for all cue conditions in low and high workload conditions

Among the cued conditions, accuracy was lowest in T for both low (88.26%) and high (79.58%) workload conditions.

Discussion

Prediction 5 is confirmed. Graded and multimodal interruption cues do result in better performance than all other interruption cues during high workload. Performance in interrupting tasks is typically expected to decline during high workload due a fall in the number of endogenous attention switches (e.g., Underwood, 2003) due to preoccupation with the primary task. The drop in performance from low to high workload was clearly observed and reached significance in all the notification cases except GVT. Potential reasons for performance decline include the effects of perceptual and attentional narrowing. Increased complexity of foveal tasks (in this case, monitoring an airspace with a four-fold increased number of airplanes while multi-tasking to identify and attend to interruptions) affects performance in perception of peripheral cues (Mackworth, 1965) and is made worse by stress and fatigue (Easterbrook, 1959). Since the narrowing effect is not only perceptual but attentional as well (Hancock & Dirkin, 1983), it is natural that there is a general decline in detection of all types of notifications, whether they feature peripheral visual cues or not.

However, the decline in performance during high workload is affected by the modality or modalities used in the notifications. For example, the largest drop in performance was observed in the uncued condition. This is perhaps because participants shift attention endogenously only when it is possible to do so, which happens rarely during high workload. Such situations demand the participants' visual attention in the hand-off region to identify main tasks, and retain it for longer periods of time due to the increase in the number of airplanes on the screen. Therefore, participants can scan their sector fewer number of times and for shorter periods of time during high workload situations resulting in lower detection rates.

Another case of a significant drop in performance is the 10% decline in detection in V. It is possible that the increase in the amount of visual clutter on the radar screen during high workload competed with the visual resources allocated to processing peripheral visual cues simultaneously. This drop in performance could also be explained by perceptual narrowing that affects performance in perception of peripheral visual cues (Mackworth, 1965). Unfortunately, this cannot be verified due to the lack of eye-tracking data. Furthermore, this phenomenon was not seen in the case of GV. Interestingly, there was no significant difference between performances in low and high workload situations in GV.

Similar to the case of GV cues, no significant difference was seen in T and GT either. This result could be due to the prospect that the types of tactile cues used in this research do not get affected by perceptual or attentional narrowing. Further research needs to be conducted in this area to explore this possibility.

Finally, the most surprising result was observed in GVT where a significant *increase* was observed between low and high workload situations. GVT notifications definitely have the advantage of both the peripheral visual and tactile cues, as well as the advantage of being detected at any of the four gradually increasing levels of salience. It would not be surprising to see that they mitigate the detrimental effects of increased workload. However, a significant increase in performance under high workload is a startling find. It will be interesting to replicate the results of this experiment under varying conditions to explore this idea further.

The results for the accuracy were very similar to those of the detection rates, with the exception of the performance in GT. The potential reasons for the low performance in this condition have been analyzed in depth in the discussion section for predictions 1 and 2.

Prediction 4: Graded notifications result in better performance (detection rates and accuracy) than other cued or uncued conditions when multiple interruptions coincide

When two notifications (always in the same modality) were presented concurrently, participants were required to hit the Ctrl button on the keyboard if they perceived the second cue before responding to the first. Since the onset of simultaneous cues was staggered by 500ms, participants often responded to the first cue before they realized that another cue was co-occurring. Some participants also did not remember to hit the Ctrl button although they perceived concurrent interrupting cues. Due to these two reasons, there is a large difference between the number of times when participants hit the Ctrl button and the number of times they attended to both interruptions of a concurrent interruption onset. We will first examine performance in acknowledgment of concurrent interruptions by analyzing the Ctrl button responses and then examine the performance in attending to concurrent interruptions (detection rate and accuracy).

Ctrl button response

The means for correct Ctrl button response were calculated for each cue condition and compared (Figure 6-12). There was a significant difference between the means among the cue conditions ($F(5,150) = 14.365$, $p < 0.001$), primarily because the performance under the no-cue condition was an extremely low 0.4%. Pairwise comparisons of the cue means showed that mean Ctrl button response for GT (17.4%) was not significantly different from response to T (20.6%) but was lower than and significantly different from responses for V (29.4%), GV (25.6%), and GVT (27.8%) (p values for all comparisons <=0.036).

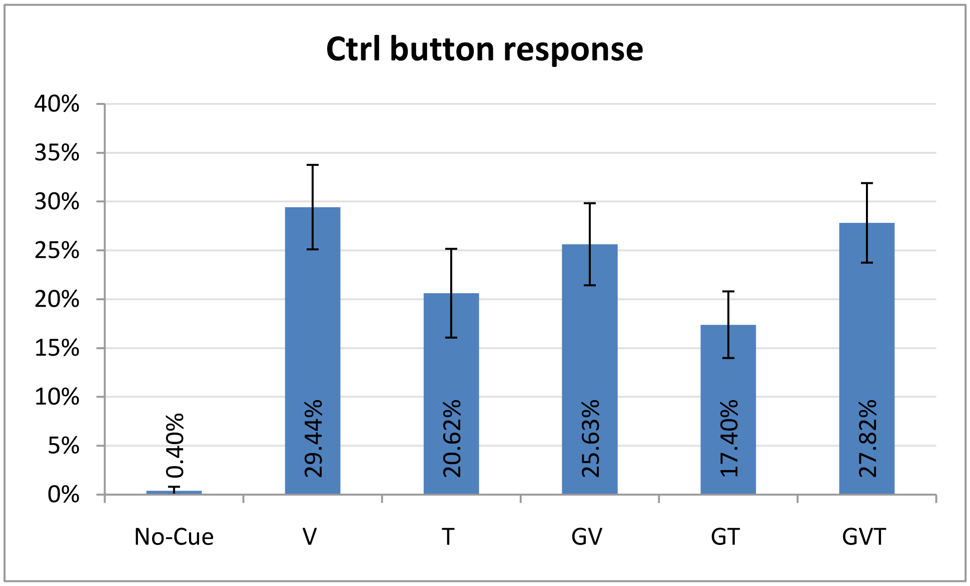


Figure 6-12: Ctrl button responses (mean and standard error) for every cue type

Detection

The effect of cue condition on detection of dual cues was significant (F(5,150) =26.319, p < 0.001) (see Figure 6-13). The absence of notifications to indicate the onset of concurrent cues in the un-cued condition resulted in a very low mean detection of concurrent cues (29.64%). This mean was significantly different from mean detection rates in all other cue conditions (p < 0.001 for all pairwise comparisons). Mean detection rate for GVT was the highest among all cue conditions (56.53%). It was significantly different from T (p = 0.02) and GT (p=0.14).

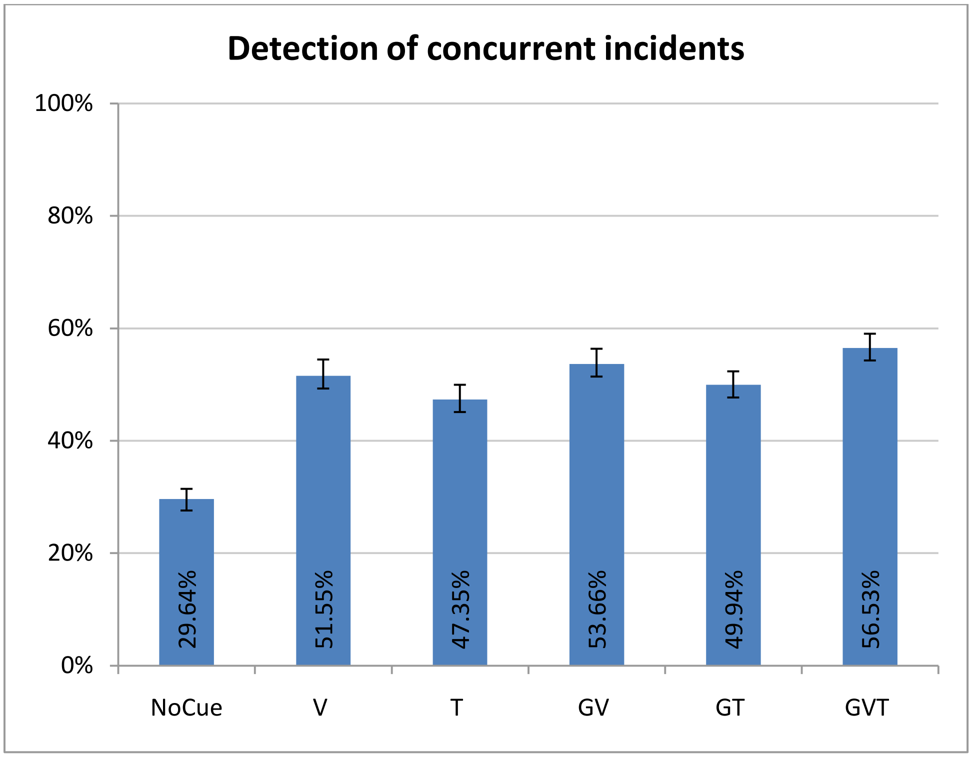


Figure 6-13: Detection of concurrent incidents (mean and standard error) for every cue condition

Accuracy

The effect of cue condition on accuracy was not significant. Accuracy was between 92%-93.5% for all conditions.

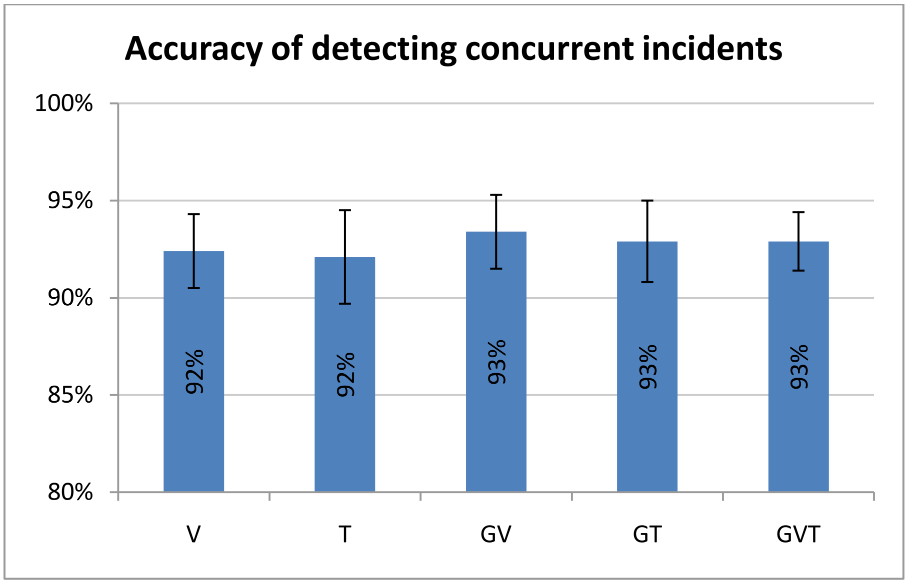


Figure 6-14: Accuracy of concurrent incidents (mean and standard error) for every cue condition

Discussion

Prediction 4 was confirmed, however, not entirely. Graded and multimodal notifications undoubtedly resulted in better performance than other cued or uncued conditions when interruptions co-occur. However, the participants did not seem to know in advance that they are faced with concurrent interruptions, and therefore could not manage the interruptions as well as they were expected to. The results show that participants attended to more concurrent interruptions during the graded notification conditions than in the non-graded conditions, but they did not always know a priori they that were about to respond to two interruptions.

As explained earlier, the onset of concurrent cues was staggered by 500ms. It is possible that the response sequence to the first interruption notification was triggered in the participant before the information for the second interruption was processed (one can argue that the nature of the processing time required for these notifications makes the reaction time akin to a recognition reaction time (Donders, 1868) which takes an average of 384 msec (Laming, 1968)). Dialogues with the participants during experiment debriefing sessions also revealed during some of the trials, many participants simply did not remember to hit the Ctrl button although they perceived concurrent interrupting cues. Therefore a clear commentary on the effectiveness of notifications on informing participants *that* notifications are in fact occurring concurrently cannot be made.

6.1.3 Goal 2: Improve decision making about attending to interruptions

Prediction 5: Response times to altitude deviations will be inversely proportionate to the severity of the deviations with graded notifications

Altitude deviations were gentle, moderate or severe. The extent of the altitude deviation was communicated to the participant through notifications and should determine how fast participants respond to an interruption.

Response times were calculated for all three intensities of altitude deviation across all participants, and means were compared using repeated measures ANOVA. Mean response times for gentle, moderate and severe interruptions were 8.238s, 6.115s and 5.228s respectively (see Figure 6-15) and the effect of intensity on the means was found to be significant (F(2,60)115.563, $p < 0.001$). Tukey's LSD post-hoc tests showed that the means were significantly different from each other as well. Mean reaction time for gentle interruptions was significantly longer than for moderate and severe interruptions, and reaction time to moderate interruptions was longer than for severe ones ($p < 0.001$ for all pairwise comparisons).

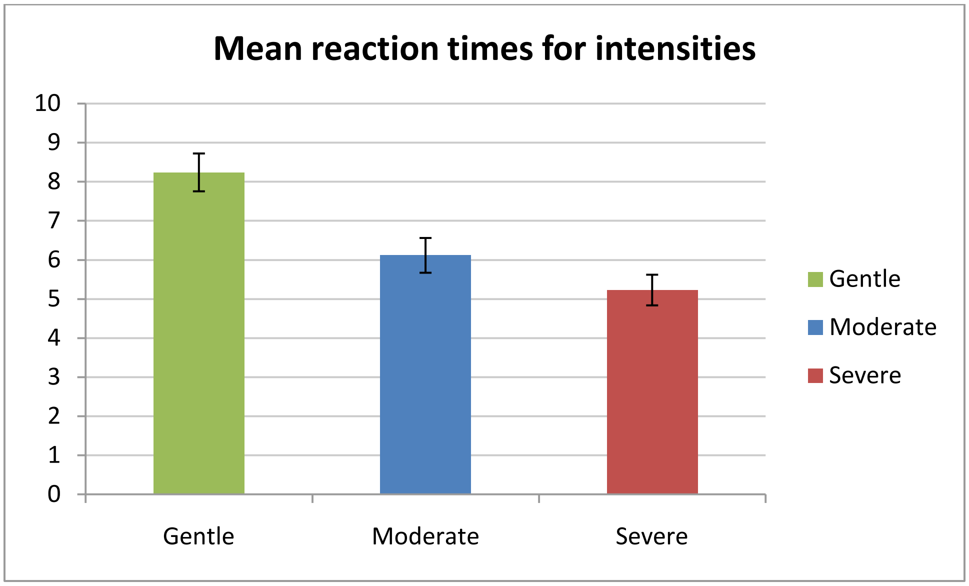


Figure 6-15: Reaction times (mean and SE) for gentle, moderate and severe interruptions

There was a significant interaction between notification type and the severity of the altitude deviation (F(10,290) = 2.061, p = 0.28), such that reaction times were not affected by severity in the uncued condition but decreased in all cued conditions as a function of increasing severity.

Pairwise comparisons showed that, for V and GT, reaction times for gentle deviations were significantly longer than for moderate deviations (V: $p < 0.001$; GT: $p = 0.001$) and for severe deviations ($p < 0.001$ for both). Also, reaction times to moderate deviations were longer than for severe deviations (V: $p = 0.016$; GT: $p < 0.001$). For T, GV and GVT, reaction times differed significantly between gentle and severe ($p < 0.001$), but not between moderate and severe, altitude deviations.

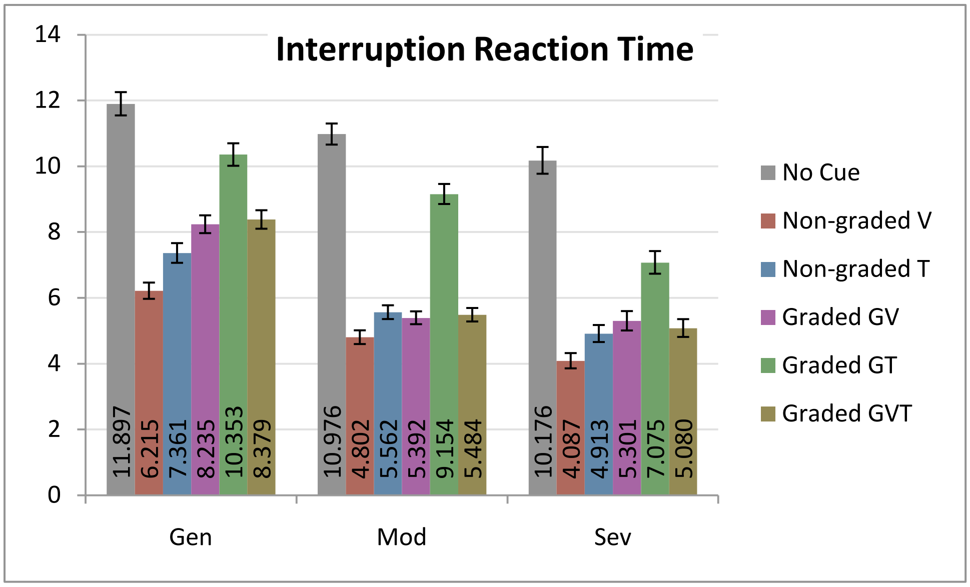


Figure 6-16: Reaction times (mean and standard error) for gentle, moderate and severe interruptions, by cue type

Discussion

Prediction 5 is confirmed. Significant differences in reaction times between gentle, moderate and severe notifications were observed in all three graded notifications. However, a closer look reveals that while there is a significant decrease in reaction time between gentle and moderate (also between gentle and severe) interruptions in all three notifications, only graded tactile notifications showed a significant drop in reaction time between moderate and severe notifications, while graded peripheral visual and graded multimodal notifications did not.

The results show that participants exhibited a clear difference in interruption attendance strategy in the graded peripheral visual notification condition. In the other two graded conditions, participants did not sufficiently vary their strategy between attending to moderate and severe interruptions.

6.1.4 Goal 4: Maintain or improve performance on primary tasks

Prediction 6: Peripheral visual and tactile interruption notifications do not adversely affect primary (focal visual) task performance (both detection rates and accuracy)

The interruption notifications in this study were presented via the under-utilized modalities of peripheral vision and touch in order to minimize interference with the primary task which required the processing of information in foveal vision and to avoid unnecessary reorientation of attention. Performance measure for the main tasks was detection rate, the percentage of main task events (i.e., responding to incorrect heading and airspeed) that were detected and responded to by the participants. First, average detection rate for main tasks was computed for each participant and compared between cued and uncued trials. Results of repeated-measures ANOVAs showed that there was no significant effect of the presence of interruption notifications on detection rate (58% in both cases, see Figure 6-17).

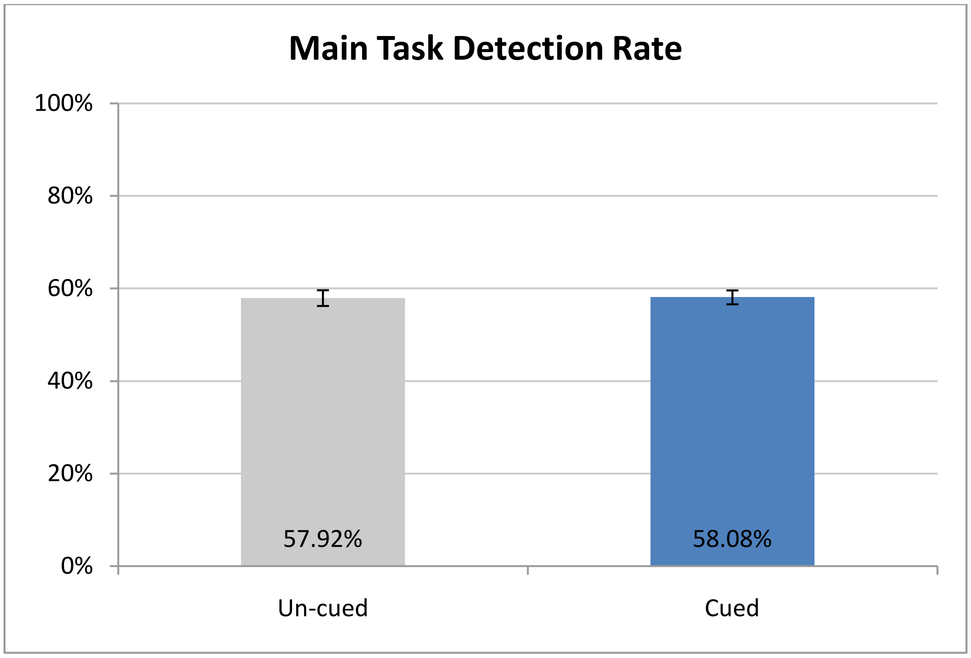


Figure 6-17: Main task detection rate (mean and standard error) during cued and un-cued trials

Next, a more in-depth analysis was conducted to compare detection rate for main tasks between all 6 conditions to determine if any of the cued conditions compete with the ongoing tasks. Results of repeated measures ANOVAs of these averages showed that there was a significant effect of notification type on detection rate ($F(5,150) = 29.53$, $p <$

0.001). Tukey's LSD post-hoc tests were used to determine differences between means for significant effects. Pair-wise comparisons show that T and GVT resulted in higher main task detection rates (61.9% and 63.1%, respectively) than any of the other cued and un-cued conditions (p <= 0.001 for all comparisons). In contrast, GV led to a lower main task detection rate (52.7%) than all other conditions, including the un-cued condition (p <= 0.001 for all comparisons). Figure 6-18 shows the main task detection rates for each of the notification types.

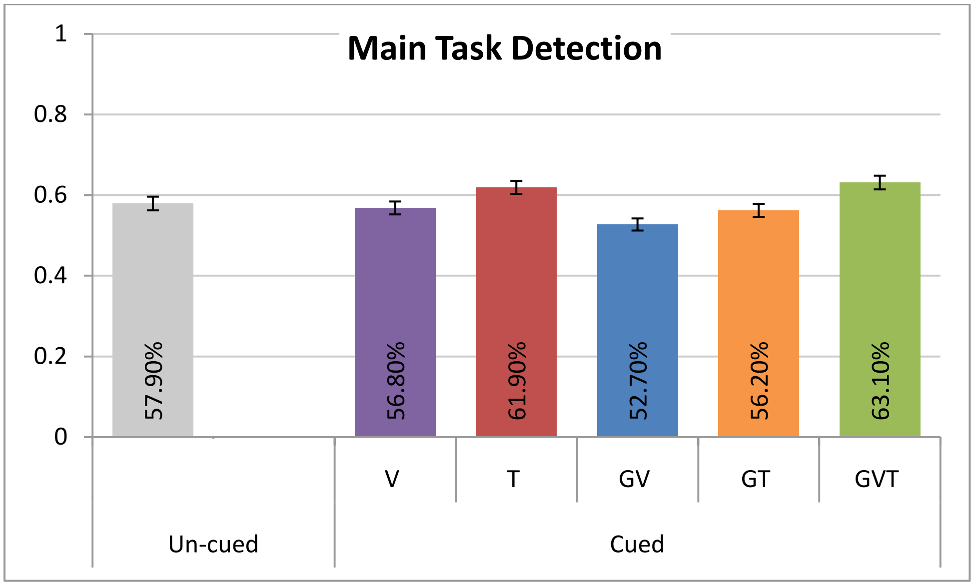


Figure 6-18: Main task detection rate (mean and standard error) for uncued and all cued trials

Discussion

Prediction 5 is confirmed when cued and uncued trials are compared. This means that the cued trials on the whole do not take away from the performance on the primary task. This confirms the assertion of the Multiple Resource Theory (Wickens, 2002; 2008) that time sharing between multiple tasks results in minimal interference to the extent that the tasks require separate information processing resources, especially when perceptual modalities are different. However, the effects of the individual notifications reveal that GV does affect performance in primary tasks negatively. The observed drop in performance was a surprising finding, since peripheral cues can be processed automatically (Muller and Rabbitt, 1989), and therefore do not reduce performance on concurrent tasks (Nikolic and Sarter, 2001). A closer look into the data reveals that the other notification conditions that featured peripheral visual cues, namely V and GVT, did not negatively affect performance on the primary visual tasks.

Certain factors set GV apart from V. GV begins at a lower salience than V and then becomes brighter than V towards the end. The steady and discrete increase in brightness might have competed with the participants' foveal visual attention. GV is also very different from GVT, in that GV is entirely made up of peripheral visual cues, unlike GVT that begins at low peripheral visual cue salience and increase to high tactile cues. Low peripheral visual cues do not seem to compete with the primary tasks, as seen in the results for V. The higher salience tactile cues also do not interfere, as evidenced by the T and GT, since the primary and interrupting tasks engage entirely different processing modalities (vision and touch). Therefore, although GVT includes peripheral visual cues, it does not adversely affect primary task performance.

Therefore, we can assume that the cause of the main task performance decline during the GV condition is possibly the result of (a) the discrete increments in salience of cues, (b) the employment of higher salience cues, or (c) a combination of the two.

Prediction 6: Response times to primary task events are faster in all cued conditions

In the absence of interruption notifications, operators have to divide their attention between the main task (watch for and correct airspeed deviations or heading deviations) and monitoring for altitude deviations in surrounding sectors. This requirement can be expected to lead to slower response times for both tasks, compared to when interruption notifications provide exogenous attention guidance to support detection of altitude deviations.

Average reaction times for main task events were computed for each participant and compared between cued and un-cued trials. Repeated measures ANOVAs showed a significant effect of the presence of interruption cues on reaction time ($F(1,30) = 24.673$, $p < 0.001$), such that participants responded to main task events faster in cued trials (RT = 3.23 sec) compared to when no notifications were presented (RT = 3.455 sec) (see Figure 6-19).

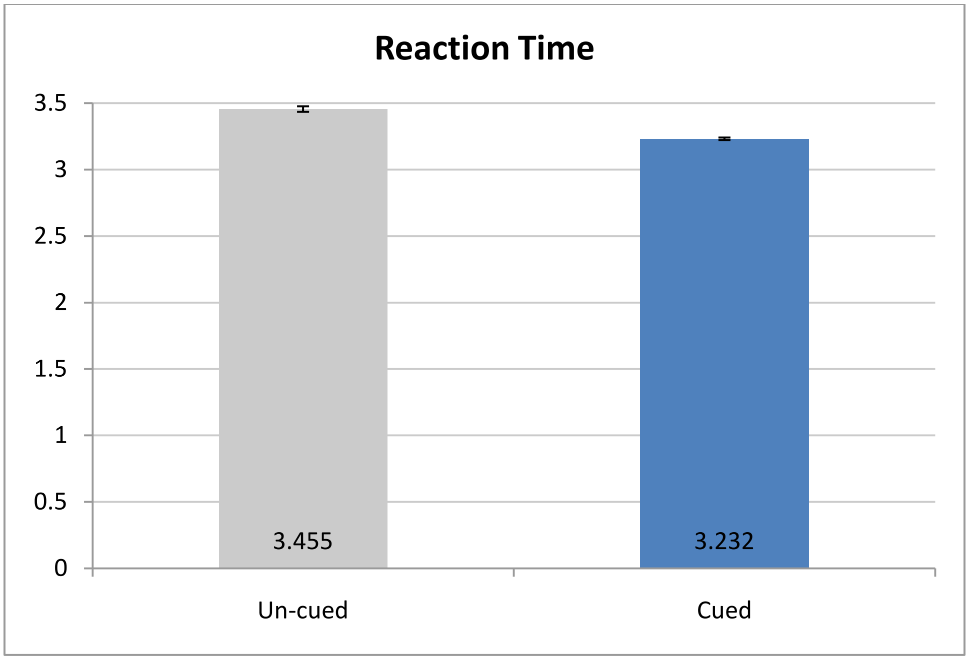


Figure 6-19: Main task reaction times (mean and standard error) during cued and un-cued interruptions

A significant interaction effect between workload and cue presence was also found (F(1,30) = 9.207, p = 0.005). The main effect of workload was significant resulting in higher mean response time during high workload (3.655s) than during low workload (3.224s) (F(1,30) = 101.825, p < 0.001). The interaction effect, however, shows that participant performance did not improve significantly when workload was high, although it improved significantly from 3.090s in uncued condition to 3.357s in the cued condition in the case of low workload (p < 0.001) (see Figure 6-20). Finally, no significant response time differences were found between the 5 cued conditions.

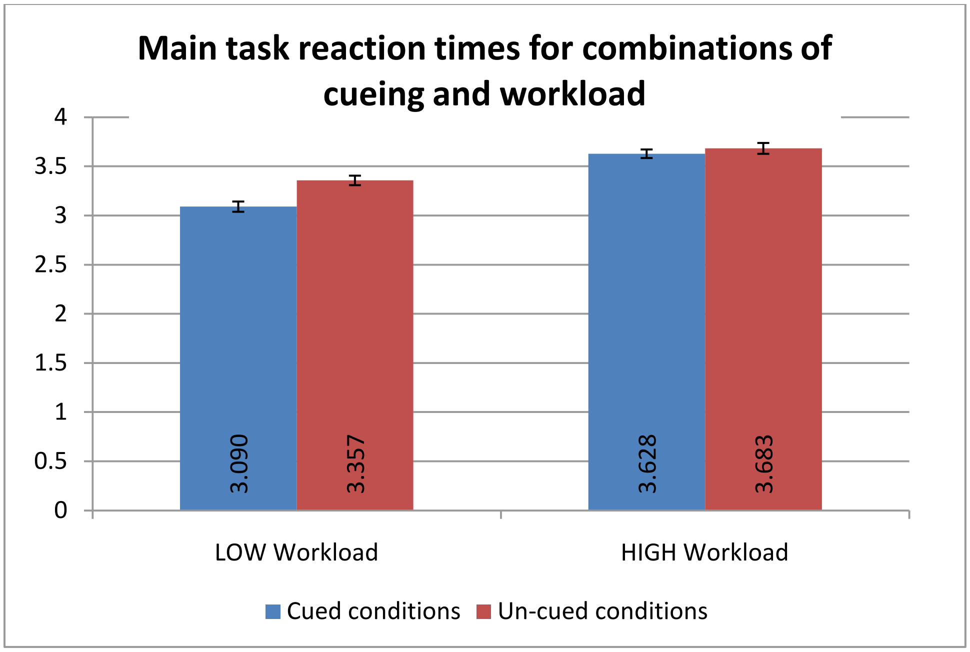


Figure 6-20: Reaction times (mean and standard error) for main tasks for combinations of cued or uncued conditions and low or high workload

Discussion

Prediction 6 is confirmed. Interruption notifications provide important information in parallel, thereby freeing up some of the resource required to perform multiple monitoring tasks. Although the presence of interruption cues reduces the need to solely rely on endogenous mechanisms (Posner, 1980), the benefit of interruption cues seem to be less under higher workload. The lack of significance seen in the case of high workload can be explained by the effect of the number of distracters on search time. In high workload conditions, the number of airplanes that need to be monitored increased almost four-fold. This has a dominant effect on search time (Drury and Clement, 1978; Egeth & Dagenbach, 1991; Wolfe, Cave & Franzel, 1989). This is especially so when the difference between the target and the distracters are small (Nagy and Sanchez, 1992), which was the case in the current study, where the target airplanes differ from the distracters only by the digits in the airplane data block or by the subtle change in the direction of travel. Therefore, it is not surprising that the increase in reaction times to main tasks due to the presence of a large number of airplanes on the screen was not significantly mitigated by interruption cues.

6.1.5 Goal 5: Improve time sharing performance and management of primary and interrupting tasks

Prediction 7: Multitasking performance is better in cued than un-cued conditions

Participants were informed that they would be rewarded for every primary task, interrupting task and secondary task they performed correctly, and penalized for every task that was missed or responded to incorrectly. This tradeoff forced participants to manage their tasks and interruptions effectively in order to maximize their rewards (for the detailed reward scheme, see appendix x) which can be considered a multitasking metric in this experiment.

Figure 6-21 shows the reward space in the form of a parallelogram. The lowest possible reward (-1608 points) is earned when all the experimental tasks are completely missed. The highest reward (1944 points) is earned if each primary, interrupting and secondary task is performed on time and correctly, following all the prescribed steps in each sequence. Rewards earned by participants in this study were in the range depicted by the light blue band (57.55 to 406.7 points).

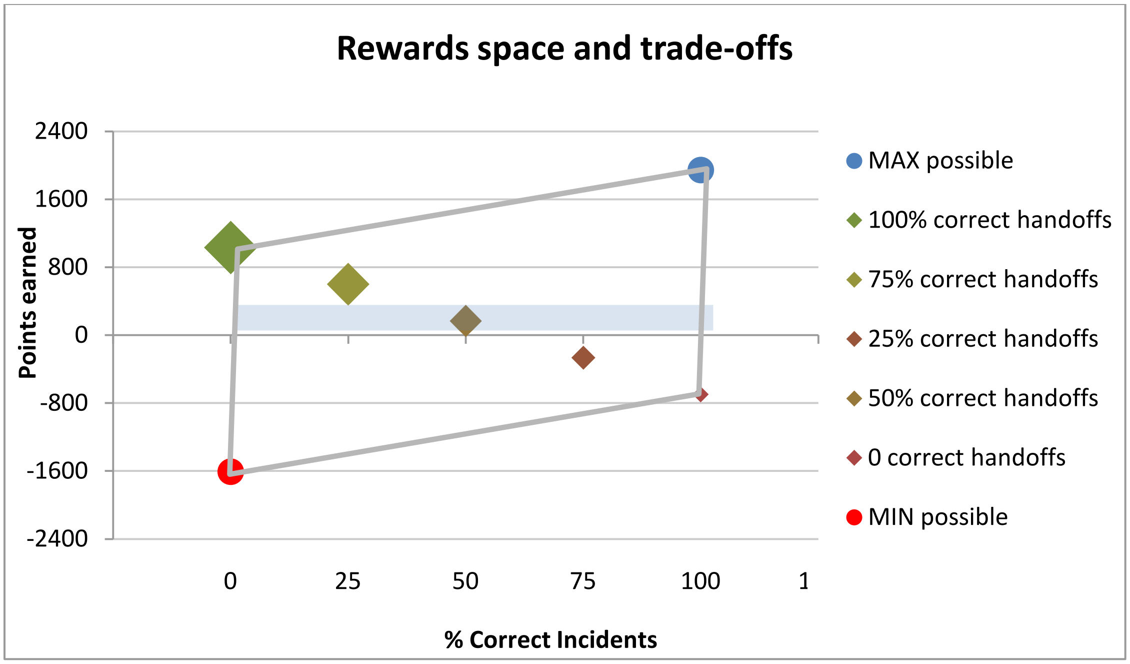


Figure 6-21: Limits of the reward space (grey parallelogram) with a sample trade-off between performing main tasks and interruption tasks. Light blue band shows range of points earned by participants in this study

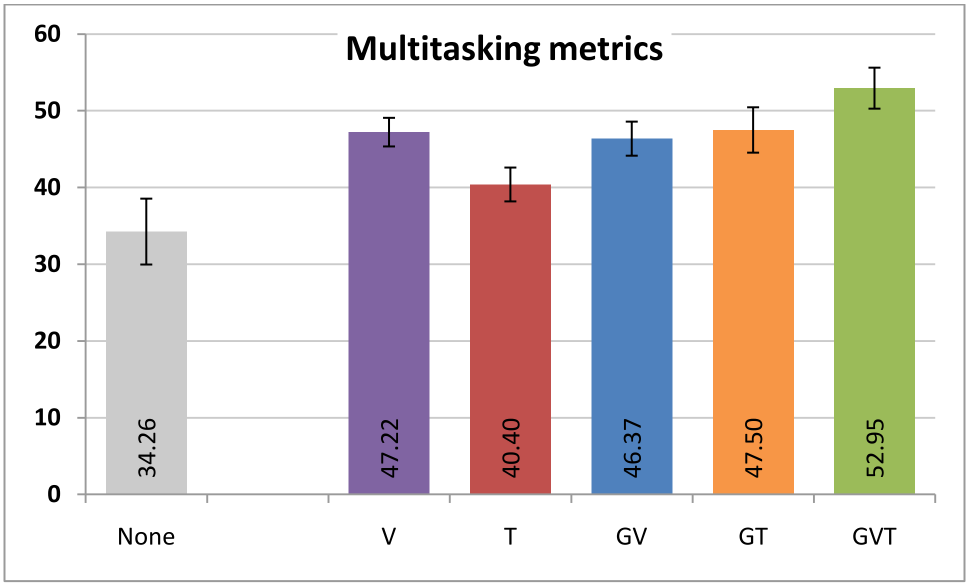


Figure 6-22: Multitasking performance metrics (mean and SE) for each type of notification

A repeated measures ANOVA on the number of points earned showed a main effect of notification type on multitasking performance ($F(5,150) = 8.494$, $p < 0.001$). Overall performance was significantly better with GVT notifications (52.9 pts) than with all other notification types ($P <= 0.032$ on all pairwise comparisons). The worst performance was seen in the no-cue condition (34.3 pts). Except for tactile notifications, all other conditions resulted in significantly better performance than the uncued case ($p <= 0.007$ for all pairwise comparisons). Performance in the tactile cueing condition (40.4 pts) was significantly lower than in all other cued conditions ($p <=0.043$ for all pairwise comparisons). Thus, in summary, multitasking performance was indeed significantly better in all cued conditions, except for tactile notifications. And crossmodally graded interruption signals resulted in the best overall performance.

Discussion

Hypothesis 8 is confirmed. The multitasking metric was not only higher for cued than un-cued conditions, but it was also the highest for the graded multimodal notification condition. The metric, composed of scores for main task performance, interrupting task performance as well as secondary task performance, rewarded a balanced approach towards the controller's tasks the most. As seen in Figure 6-21, extreme prioritization of the main task over the interrupting tasks, was penalized, and prioritization of the interrupting tasks over the main tasks was penalized severely. Although the results do not clearly show the exact prioritizing techniques adopted by the participants, they do show that graded multimodal notifications assisted participants in making the most gainful choices in managing their task demands.

6.1.6 Participant experiences

Both quantitative ratings/rankings as well as qualitative feedback on the experiment and designs were collected. The background questionnaire and debriefing questionnaire can be found in Appendices I and III.

Quantitative ratings and rankings

Participants provided subjective ratings and rankings on different aspects of the experiment as well as the notification types. Subjective ratings were collected for the following:

1. Adequacy of training, on a scale of 1 (inadequate) to 5 (excellent)
2. Fatigue caused by the experiment, on a scale of 1 (not at all fatigued) to 5 (extremely fatigued)
3. Stress levels experienced during cued and un-cued conditions, on a scale of 1 (no stress) to 10 (most stressed I have ever been)

The median value for adequacy of training was 5 (excellent). Fatigue caused by the experiment was determined by comparing the ranking of fatigue levels before the experiment to the ones after using non-parametric Kruskal-Wallis test. Reported fatigue levels increased significantly over the course of the experiment ($H(1) = 27.097$, $p < 0.001$). Stress levels for cued and uncued conditions were also compared using non-parametric Kruskal-Wallis test. Stress levels were not significantly affected by the presence or absence of cues ($H(1) = 3.227$, $p = 0.072$).

Participants were asked to rank order the notifications based on the following criteria:

1. Perceived difficulty of:
    a. Detecting notifications
    b. Identifying altitude deviation severity
    c. Determining incident location
2. Annoyance

Participant rankings were analyzed using non-parametric Friedman's ANOVA.

With respect to the difficulty of detecting notifications, participants' rankings differed significantly between notification types ($X^2(4) = 12.469$, $p = 0.014$). Figure 6-23 shows the errors and quartiles.

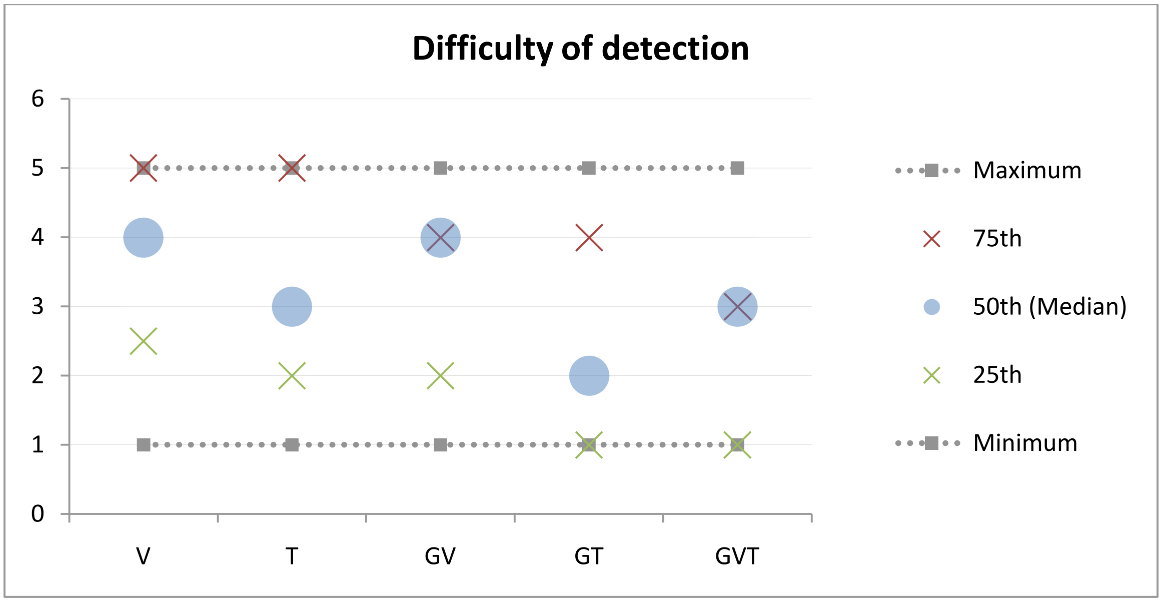


Figure 6-23: Subjective ranking and quartiles of difficulty of detecting incident notifications

Participant ranking for the difficulty of identifying interruption severity based on notifications was also significant ($X^2(4) = 13.497$, $p = 0.009$). Figure 6-24 shows the means, standard errors and quartiles.

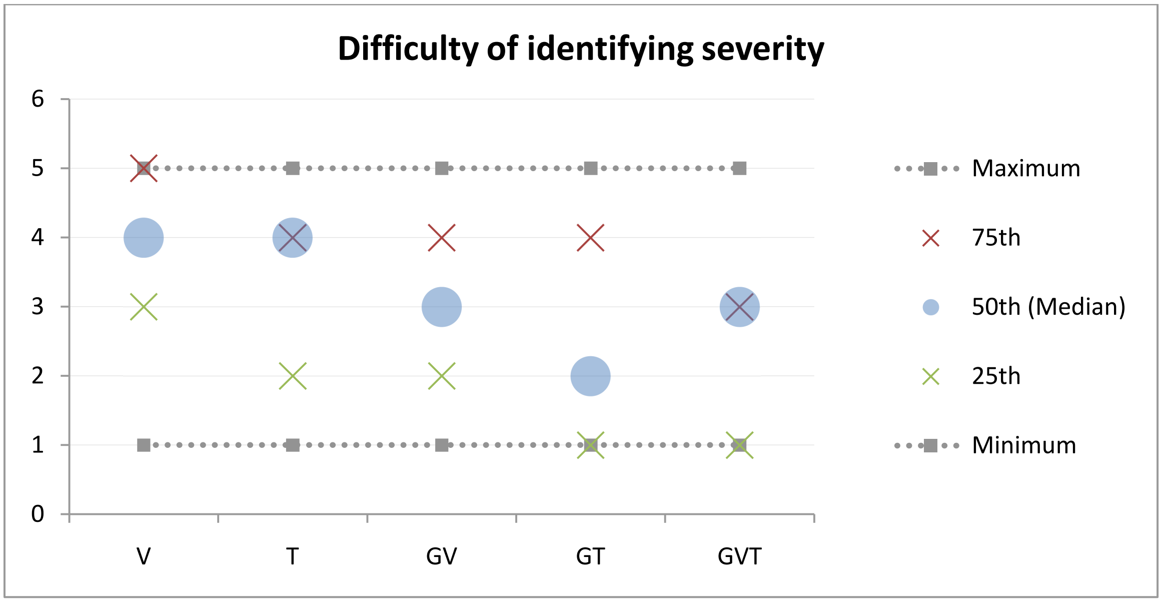


Figure 6-24: Subjective ranking and quartiles of difficulty of severity of incident based on notifications

Participant ranking were significant for the difficulty of determining the location of the interruption based on notifications as well ($X^2(4) = 29.032$, $p < 0.001$). Figure 6-25 shows the means, standard errors and quartiles.

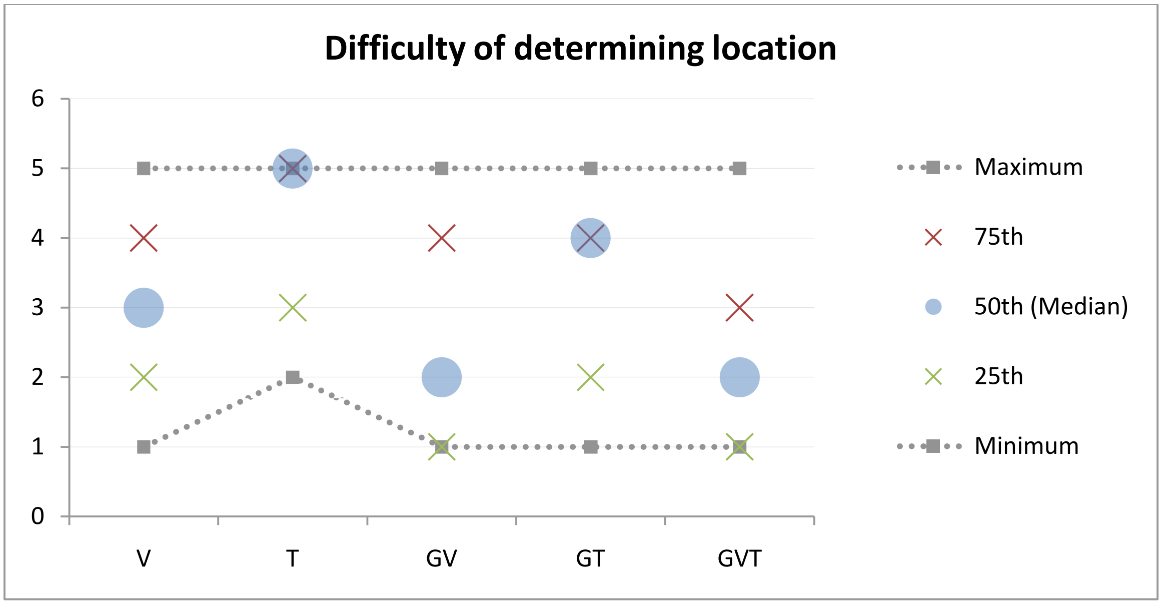


Figure 6-25: Subjective ranking and quartiles of difficulty of determining location of incident based on notifications

Qualitative feedback

Participants provided feedback about their experiences during the experiment by answering a series of open-ended questions (listed below). The responses were analyzed and interpreted using an inductive analysis protocol. The total number of participant responses to each open-ended question is listed, as are the most common responses to each question. Each of these responses is further characterized by the number of participants who indicated the issue.

Un-cued condition:

1. Parts of the experiment that seemed particularly difficult (Total no. of responses = 31)
    - *Absence of notifications for interrupting tasks* (8 out of 31 responses)

  - *Attempt to correct an airplane just as it times out* (7 out of 31 responses)
2. Strategies adopted to perform task and attention management (Total no. of responses = 31)
   - *Search for interrupting incidents immediately after a hand-off* (10 out of 31 responses)
   - *Search surrounding sectors according to a routine pattern* (9 out of 31 responses)
   - *Search for interrupting incidents when hand-off region was sparse* (8 out of 31 responses)
3. Most enjoyable aspect of having no cues (Total no. of responses = 30)
   - *Less pressure to detect and interpret notifications, more freedom* (10 out of 30 responses)
   - *"Ignorance is bliss"* (8 out of 30 responses)
4. Stressful aspects of not receiving interruption notifications (Total no. of responses = 31)
   - *Increased difficulty in high workload situations* (12 out of 31 responses)
   - *Necessity to scan both hand-off region and surrounding sectors* (11 out of 31 responses)

Cued conditions:

1. Parts of the experiment that seemed particularly (Total no. of responses = 31 responses)
   - *High workload situations* (15 out of 31 responses)
   - *Feeling and interpreting tactile cues* (8 out of 31 responses)
   - *Simultaneous onset of interruption notifications* (8 out of 31 responses)
2. Strategies adopted to perform task and attention management (Total no. of responses = 31)
   - *Time shared between secondary task –bar and hand-offs* (13 out of 31 responses)
   - *Hand-offs (main tasks) were prioritized* (12 out of 31 responses)

3. Most enjoyable aspect of receiving interruption notifications (Total no. of responses = 29)
    - *Variety of different notifications* (9 out of 29 responses)
4. Stressful aspect of receiving interruption notifications (Total no. of responses = 29)
    - *High workload situations* (13 out of 29 responses)
    - *Trouble detecting/interpreting cues* (7 out of 29 responses)
    - *Detecting two types of hand-offs* (6 out of 29 responses)

Overall experiment:

1. Methods adopted to overrule experiment instructions (Total no. of responses = 25)
    - *Deviant airplanes were identified before interruption notifications were acknowledged* (6 out of 25 responses)
2. Most frustrating aspect of experiment (Total no. of responses = 31)
    - *Tasks time out just when they are about to be responded to* (12 out of 31 responses)
    - *Incorrectly responding to a hand-off or incident* (7 out of 31 responses)
3. Issues with simulator set-up (Total no. of responses = 18)
    - *Weak tactile cues* (5 out of 18 responses)
    - *Interrupting tasks that were barely missed were recorded as wrong* (5 out of 18 responses)
4. Overall comments (Total no. of responses = 19)
    - *Simulator/experiment experience challenging/enjoyable* (13 out of 19 responses)
    - *Notification salience levels need to be increased* (9 out of 19 responses)

Discussion

The feedback provided by the participants is an important component of this experiment. While the experiment results provide an objective analysis of the effectiveness of notifications, the subjective feedback gives insight into how easy or difficult it was for participants to deal with interruption notifications or the lack of them.

The relationship between the perceived difficulty of performing an activity and the success of performing the same activity was not a linear one. For example, subjective rankings for difficulty of detecting notifications show that graded peripheral visual notifications were easiest to detect, while in fact, it was graded multimodal notifications that were detected the highest number of times. Another way to look at the subjective rankings is that the top two notification types for all the three measures always happen to be graded notifications (graded tactile and graded multimodal for ease of detection and severity, and graded multimodal and graded visual for location). However, participant performances show that the top two modalities for the same measures were consistently graded multimodal and 4-stage peripheral visual notifications. Interestingly, participants favored the 4-stage peripheral visual notifications in two of the three measures (detection and severity).

Qualitative feedback provided by the participants was insightful. Participants generally found it difficult to identify and respond to interruptions in the absence of notifications, particularly during high workload. It was frustrating for some participants to have to deal with the additional burden of having to scan for interruptions intermittently alongside main tasks. On the other hand, some participants also found that not having interruption cues was in a way less stressful and gave them more freedom to perform the tasks at their own pace. Especially during high workload when it was almost impossible to scan and find every hand-off and interruption, some participants felt that they were better off: “ignorance is bliss”.

It was frustrating for some participants to have to deal with the additional burden of having to scan for interruptions intermittently alongside main tasks. On the other hand, some participants also found that not having interruption cues was in a way less stressful and gave them more freedom to perform the tasks at their own pace. Especially during high workload when it was almost impossible to scan and find every hand-off and interruption, some participants felt that they were better off: “ignorance is bliss”. Participants overcame the difficulties of the absence of interruption assistance by adopting techniques such as “sweeping” the screen for interruptions right after completing a hand-off, or when the hand-off region was sparsely populated with airplanes, or simply following a routine pattern to scan the screen.

High workload was considered as a problem even when notifications were presented. Participants found that it was both difficult and stressful to perform their tasks during high workload. They also found that detecting and interpreting cues, especially tactile cues, was difficult and stressful. Interestingly, no special mention was made about peripheral visual cues, although they were ranked high in difficulty in the ranking scores. Participants seem to enjoy the “variety” of receiving different types of notifications within sessions, as opposed to the monotony of having to perform the same tasks repeatedly. Attention and task management techniques used include time-sharing between primary tasks (hand-offs) and secondary tasks (task bar operations) and prioritizing hand-offs over interruptions.

Almost all participants expressed frustration over missing several tasks (both primary and interrupting) at the very last moment, indicating that their performances might have increased if the response windows were lengthened. Overall, participants conveyed that the experiment was challenging and enjoyable, but that it would have been a much better experience had the notification saliences, particularly in the tactile modality, been higher. Given that the superiority of cross-modal graded cues over the regular four-stage notifications has been attributed to the graded and cross-modal nature of the notifications rather than an increase in salience, the constraint on the equality of peripheral visual cues and tactile cues can be lifted. It might be worth repeating the experiment with increased tactile cue saliences to see if responses to notifications can be increased even further without affecting performance on concurrent tasks.

6.2 Summary of findings

Three experiments were conducted with the goal to identify notifications that can most effectively support attention and interruption management. To this end, experiment 1 examined the possibility of exploiting crossmodal links between vision and touch to guide an operator's spatial attention. Experiment 2 was conducted to explore the use of informative peripheral visual and tactile cues to present partial information about pending interruptions without requiring a reorientation of attention away from an ongoing task. Finally, experiment 3 tested the effectiveness of graded and multimodal notifications for supporting multitasking, monitoring, and interruption management in a rather demanding task environment.

The results from these experiments confirm that target identification in complex environments benefits greatly from simple forms of cueing. A further increase in target identification performance is observed when location-specific cueing is used. Both increased accuracy and reduced search time for the target (and thus less time away from ongoing tasks) were observed. Spatial information is best communicated via the peripheral visual channel while the tactile channel, given its high temporal resolution, can be used effectively to communicate categorical information (e.g., the urgency of attending to an interruption) by modulating the pulse rate of the signal.

All forms of cueing significantly improved performance on interrupting tasks, both in terms of detection time and accuracy, without compromising performance on ongoing tasks. However, under high cognitive load, performance with single-stage peripheral visual cues was significantly affected, both in terms of detecting and interpreting these cues. Multimodal notifications were more robust due to the fact that they exploit and combine the individual strengths of different sensory channels while offsetting their weaknesses.

Graded multimodal notifications had the most pronounced positive effect on ongoing and interrupting task performance. This benefit was observed even during high cognitive load conditions and with concurrent tasks and/or events. It appears, however, that the benefits of graded multimodal notifications may not result primarily from switching between modalities. Gradation and the different characteristics of sensory channels may play a more important role. In particular, the relatively low salience of peripheral visual cues helped reduce interference with primary tasks, especially during low workload. And the proximal nature of later stage tactile cues ensured detection of notifications that were missed during the early stages.

Chapter 7 General Discussion

The previous chapters described the comparative evaluation of different kinds of interruption notifications in the context of a simplified air traffic control simulation. These notifications were designed to support attention and interruption management in data-rich event-driven domains. Four of the five notifications, namely, the four-stage peripheral visual and tactile notifications as well as the graded peripheral visual and tactile ones, were adaptations and refinements of alarm designs that were tested in a very limited number of earlier research efforts (e.g., Hopp-Levine, Smith, Clegg, & Heggestad, 2006; Nikolic & Sarter, 2001; Lee, 1992; Hess and Detweiler, 1994). The "cross-modal graded" notifications were novel in that they employ two-stage gradation within a modality (peripheral vision and touch) as well as gradation across modalities (from peripheral vision to touch). This approach was chosen as it allowed us to exploit the orienting power of peripheral vision even at low levels of salience while also addressing the risk of attentional narrowing under high workload in this channel. The findings from this final experiment show that cross-modal graded notifications provide the most significant performance benefits in both the primary tasks as well as the interrupting tasks, thereby assisting in attention and interruption management.

The present research was motivated by the goal to address the so-called 'alarm problem' (Woods, 1995) by improving a) the informativeness, b) the attention directing capabilities and c) the perceptual functions of notifications. The following sections elaborate on how each of the three proposed solutions have been addressed.

Improve the informativeness of notifications

An increased level of informativeness is what sets the notifications developed in this research apart from traditional alarms. The effectiveness of this approach confirms the assertion of Multiple Resource Theory (Wickens et al., 1980, 1984) that parallel information processing (in this case, processing of partial information about an interrupting task/event) can be supported by distributing information across multiple sensory channels. There are two important aspects to informativeness: 1) what information *can* be conveyed through secondary input channels given their characteristic bandwidth, and 2) what information *ought* to be conveyed to help operators manage their attention more effectively – the aspect of proper preattentive reference.

Extensive research has been done on the information carrying capabilities of the modalities of peripheral vision and touch (e.g., Hopp-Levine, Smith, Clegg, & Heggestad, 2006; Nikolic & Sarter, 2001; Sarter, 2002; Brewster and Brown, 2004). The issue was also addressed in the second study (see chapter 3) of this line of research where peripheral visual and tactile notifications were designed to convey partial information about interruptions. The cues were presented in one of three locations to indicate the subsystem affected by a problem. Different pulse rates were used to indicate the importance of attending to the problem. And two different cue durations were employed to signal how long it would take to address the problem. Although domain, duration, and importance were all encoded in the same notification, participants could reliably interpret all three cue attributes.

Based on this finding and similar results from previous research, multiple parameters were varied in the peripheral visual and tactile modalities for the final study also although the particular information content differed from that in the previous experiment. In addition to using pulse rates of the cues to convey the importance of attending to a notification, three of the five notification designs tested in this study employed gradation (Lee, 1992; Hess and Detweiler, 1994), both intra- and crossmodally, to remind participants of pending interruptions and indicate changes in the urgency of attending to the event. Adding this dynamic component to the design proved to be highly beneficial. Even at high workload levels, crossmodally graded notifications showed the highest detection rates and, importantly, the most appropriate response times for various levels of urgency, thus avoiding unnecessary interruptions of the ongoing task.

Improve attention direction capabilities of notifications

The capability of notifications to not only inform operators about the presence of an interruption but also direct their attention to the location or source of the interruption is of great significance to operator performance. An effective way to achieve this goal is to exploit crossmodal links in spatial attention, or the increase in readiness of one modality to perceive a signal due to the appearance of a signal in another modality in the same or similar location (Spence and Driver, 1997; Driver and Spence, 1998; and Spence et al, 2004). In the first study described in chapter 2, the effectiveness of crossmodal spatial links between vision and was examined. Performance on locating unexpected events or interruptions was tested and compared between a baseline and two cued conditions. One of the cued conditions featured generic uninformative tactile cueing and the other had location-specific tactile cueing that guided visual attention to the location of the interruption. The comparison between these conditions showed that the attention directing capabilities of location-specific cueing resulted in a dramatic 44% increase in the detection and identification of interruptions.

This approach was used also in the final study to direct operators' attention to one of eight sectors on an ATC radar screen (the ninth sector being the focus of the operators at the center). In this case, not only tactile but also peripheral visual cues were used to direct spatial attention. The accuracy with which interruptions were located in the four-stage and graded peripheral visual conditions, as well as in the crossmodal graded conditions, was significantly higher than in other conditions. However, for tactile notifications, no comparable benefit was observed. This may be explained by the difference in body sites used for presenting tactile cues in the two experiments. Extensive literature review of the effectiveness of various body sites suggested that the front of the abdomen was as effective as the back of the abdomen (used in the previous experiments) in both sensitivity to, and localization of, vibrotactile cues. However, this was not confirmed by the findings from the final experiment. One potential explanation for this unexpected result is the uneven application of pressure on the various tactors due to the gaps between the different contours of the abdomen and the relatively straight surface of the tactile harness. In addition to the high accuracy in locating interruptions, a significant reduction in response times was found also across all experiments when attention-directing notifications were used.

Improve perceptual functions of notifications

The goal of improving the perceptual functions of notifications was addressed in all three experiments. The first study, described in chapter 2, showed that even a basic, rather uninformative, single-stage tactile cue can significantly improve the detection of unexpected events without impacting the primary task. This supports the assertion of Multiple Resource Theory (MRT; Wickens, 2002) that concurrent information processing and timesharing incurs minimal interference if the information or tasks involve different perceptual modalities. The second study, described in chapter 3, went beyond the first study in showing that not only tactile cues, but also peripheral visual cues, can be processed, to some extent, in parallel with ongoing foveal visual tasks, thus supporting the more recent version of MRT (Wickens, 2008) that distinguishes between these two visual channels. Note, however, that while peripheral visual cues were detected reliable, they did affect performance on the primary tasks (unlike tactile cues). This suggested that even less salient peripheral visual notifications may be needed.

In the final study, intra- and crossmodal gradation as well as a combination of modalities were employed to improve performance. The gradation and the multimodality had significant positive effects on performance; however, the switch from one modality (peripheral vision) to another (touch) did not lead to an improvement in performance.

While the newly designed notifications led to significant improvements in performance, there may be ways to further enhance their effectiveness. For example, careful adjustments of the salience levels of individual cues could possibly lead to even better detection and overall task performance. Care was taken to choose the lowest detectable salience levels for peripheral visual cues and the highest acceptable salience levels for tactile cues based on related literature, previous experiment results and pilot studies. While two out of the three notifications that featured peripheral visual cues were effectively detected and processed in parallel with ongoing tasks, the third graded notification type which employed a high level of salience which was not used in the other two notifications showed a significant reduction in main task performance.

The opposite issue arises with respect to the tactile notifications. Although tactile cues did not interfere with the performance of the main tasks, they were not as reliably detected at very low levels of salience. Cues used in the first two experiments employed a 150Hz vibration that was reliably detected. In the final experiment, the need to lower the salience of peripheral visual cues and ensuring crossmodal matching, i.e., equivalent perceived salience of peripheral visual and tactile cues, resulted in the use of tactile cues at 105Hz and even 70Hz in the graded cases. Some participants struggled to notice the signals at these frequencies, and hence there was a drop in detection performance for some tactile notifications but performance on the primary tasks was unaffected.

In summary, the results of all three studies in this line of research make significant contributions to a better understanding of benefits and limitations of multimodal and graded information processing and presentation. They demonstrate how the peripheral visual and tactile channel can be used to support attention and interruption management in complex data-rich domains without necessarily incurring performance costs on ongoing primary focal visual tasks. In the following chapter, the contributions of this research and future directions for investigation will be discussed in more detail.

Chapter 8 Conclusion

The responsibilities of operators in complex environments are exemplified by competition for attentional resources. Operators typically struggle to timeshare numerous tasks while being burdened with overloaded sensory modalities. Aviation, Air Traffic Control in particular, is a domain that is experiencing these growing problems at an increasing rate. ATC operators struggle to schedule and manage an increasing number of tasks and interruptions. One of the most critical tasks they perform is traffic separation in four-dimensional space (Nolan, 1994). Currently, about a million air carrier departures are performed from US airports per month, that is, twice the number of departures performed three decades ago. The rate of increase in air traffic is only going to intensify in the next decade or two, resulting in a 2-3-fold increase in air traffic by the year 2025 (JPDO, 2010) (see Figure 8-1). The transformation designed to accommodate this projected increase in air carriers includes a shift of authority from air traffic controllers to pilots under a new system called NextGen (Next Generation Aviation System). In NextGen, tasks that were traditionally performed by air traffic controllers, like changing flight paths and maintaining traffic separation, will now be performed by pilots, leaving it to the air traffic controller to maintain awareness of the traffic situation. This new system is intended to improve automation to assist in monitoring; however, it translates to more monitoring responsibilities and more unexpected interruptions than ever before, calling for effective support of attention management.

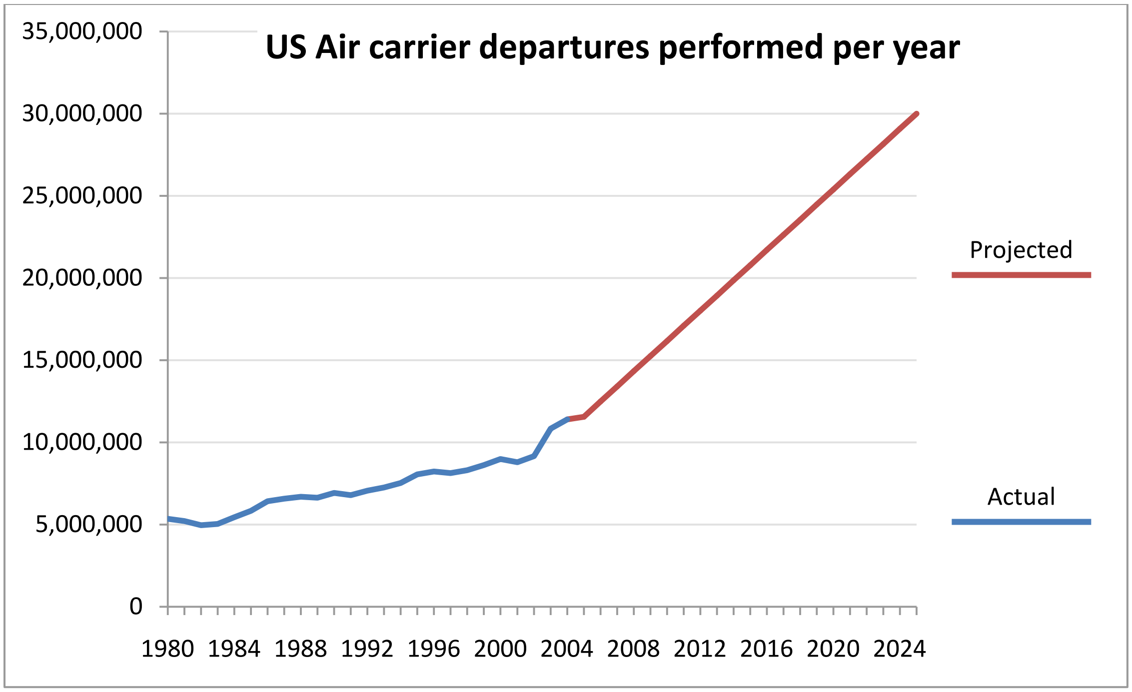


Figure 8-1: Actual air carrier departures between the years 1980 and 2004 (blue) (source: RITA, BTS) and projected increase until the year 2025 (red) (source: JPDO, 2010)

Several efforts have been made to support attention management, particularly by addressing the alarm problem (Woods, 1995) through the use of notifications that focus on three aspects of information processing (a) improving the perceptual functions, (b) improving the informativeness, and (c) improving the attention directing capabilities of notifications. Based on the Multiple Resource Theory (Wickens, 1984), it has been shown that multimodal notifications, or signals presented via alternative modalities like peripheral vision and touch (Sarter, 2002, Oviatt, 2002), can help alleviate cognitive overload. Studies have also shown that graded notifications that convey the degree of importance of attending to interruptions through variations of signal intensity (Lee et al., Sarter, 2005) assist with attention management and workload distribution. This thesis combines the two approaches into the creation of a novel interruption notification system that can alleviate cognitive load and assist with effective attention allocation by employing notifications that are both graded and multimodal.

The present research compared the benefits of using five different types of notification systems: a four-stage peripheral visual notification, a four-stage tactile notification, a graded peripheral visual notification, and a graded multimodal notification which combined peripheral visual and tactile cues. These notifications were compared amongst each other and to a baseline condition which mimicked a typical air traffic control radar screen with no interruption notifications. The design of this study was informed by two previous studies that were conducted to explore the three aspects of information processing that could address the alarm problem. The first study (chapter 2) addressed the first two aspects, namely, improving perceptual functions and informativeness of notifications. This study explored whether informative peripheral visual and tactile notifications can support task switching and interruption management in a complex environment. The second study (chapter 3) addressed the third aspect as well, namely, the attention directing capabilities of notifications. This study investigated the extent to which cross-modal spatial links can be exploited in the use of tactile cues to mitigate data overload and support attention allocation. Based on these two studies, a new system of notifications that incorporated all the three aspects of information processing was developed. As described in chapter 4, four smaller studies were conducted to determine the specific attributes of the notifications including cue arrangement and cue salience. The notifications that were developed from these studies were then tested in the concluding experiment described in chapter 5.

Graded multimodal notifications, among all other notification designs, were found to be highly effective in supporting monitoring and interruption management in a complex environment. The use of these notifications for interruption cueing did not affect performance on main tasks; furthermore, it improved reaction time to main task events. The graded multimodal notifications were found to be the most effective among all the notifications in significantly improving detection rates and accuracy of responding to interruptions, as well as shortening the reaction times to respond to interruptions. Since operators require most assistance during periods of high cognitive load, these notifications were tested in two such conditions that are typically seen in complex environments: conditions of increased workload, and conditions during which interruptions occur simultaneously. When tested under these conditions, graded multimodal notifications provided superior performance benefits, compared to all other notifications. They also seemed to assist in improving decision making about task switching. Finally, in a comparison of scores awarded for optimal multi-tasking behavior, graded multimodal notifications emerged most successful.

The findings of this research can directly inform the design of displays and interfaces in complex environments. The methods used in this research to maintain the equilibrium between providing information that can be reliably detected while being minimally intrusive can directly be applied to most complex systems that operate under such constraints. The study also goes beyond earlier studies by examining the robustness of peripheral visual and tactile graded cueing under conditions of high workload and in the presence of concurrent tasks. It examines the effectiveness of crossmodal gradation of signals in addressing the problem of data overload by facilitating attention capture and providing attention guidance. Acknowledging the need for context-sensitive information presentation, it uses graded feedback to account for differences in the significance and urgency of tasks and events, avoiding untimely and/or unnecessary attention switching and associated performance costs, supporting workload planning and distribution, minimizing errors of omission, and ultimately leading to improved interruption management.

These research efforts can be carried forward in several ways. The first and most natural extension to the current research is the enhancement of the design of graded multimodal notifications. Pushing the envelope of the capabilities of the operators in the right direction and to the right extent could potentially result in much higher performance benefits than those seen in this research. For example, encoding more information into the notifications to enable better interruption management or increasing the salience levels of individual cues to improve detectability of interruptions will be interesting avenues to explore.

Secondly, adding redundancy to interruption notifications, especially to critical ones, might be very meaningful and beneficial. A comparison of the current graded multimodal notifications with a set of similar notifications that feature a blend of concurrent peripheral visual and tactile cues, each blend differing in cue characteristics to support various needs of the operator, would be a useful exercise in further understanding human information processing needs and capabilities.

Third, in an effort to overcome some of the limitations in this study, adaptable mechanisms can be incorporated into the design of notifications to enable the operator to set the notification saliences based on his/her own thresholds of sensitivity, as opposed to having universal values for the same. Such an effort will also make the notification design more applicable in the real world in which operators demographics would differ widely from those of experiment participants.

Finally, developing a model based on the findings of this research, such as Latorella's Interruption Management Stage Model (1998, 1999) and Wickens's Multiple Resource Theory model (2002, 2008) would be most beneficial to future researchers who endeavor to further explore this line of research.

The current research belongs to a long lineage of efforts made in the understanding and improvement of human attention and interruption management, particularly in domains such as aviation where ineffective information presentation contributes to human error. This study, and continued research efforts in this direction, will contribute to safer and more efficient operations in complex domains.

Appendix I. Participant background questionnaire

Name:

Day 1
Please fill out this question when you arrive on Day 1.

1. Age

2. Are you:
* Right handed
* Left handed
* Ambidextrous

3. Do any of the following apply to you?
* My vision is fine - I do not require vision correction
* My vision is fine - I am wearing corrective contact lenses
* My vision is fine - I am wearing corrective glasses
* My vision is fine – but I have color blindness (please elaborate under "other")
* Other:

4. Please rate your current level of fatigue on Day 1 (circle one)
1 2 3 4 5
Not fatigued at all Extremely fatigued

5. Please rate your level of experience with computer/video games (circle one):
1 2 3 4 5
None Play more than 5 hours per day

6. Are you a pilot?
* No
* Yes (please explain below under "other")
* Other:

7. Do you have any prior knowledge about air traffic control operations?
* No
* Yes (please explain below under "other")
* Other:

Day 2
Please fill out this question when you arrive on Day 2.

Please rate your current level of fatigue on Day 2 (circle one)

	1	2	3	4	5
Not fatigued at all					Extremely fatigued

Appendix II. Participant instructions

Participant Instructions - DAY 1

Notes to researcher:

0. Exit Dropbox from system tray (bottom right) and turn off back-up software.
1. Have the participants fill out the Background Day 1 Questionnaire
2. Attach tactor belt
3. Ask participants to adjust their chairs to a comfortable height
4. Turn on flashlight
5. Seat the participant, shut door, turn light off
6. Ensure second monitor is turned on
7. Turn on speakers
8. Ensure desktop background is ATC screenshot
9. Start experiment loader, Tutorial box is *unchecked*
10. Ask participant to type their name in the box
11. Start *MAINTASK* <<< 2:27 >>> on each of the 2 computers

Please hit Enter. I request you to please sit back in your seats and watch the screen as I speak. Do not perform any actions using the mouse yet. You will be allowed to try it in a couple of minutes.

Welcome to our AIR TRAFFIC CONTROL simulation! You are sitting in an approach control tower and looking at the radar screen that shows you the airplanes that are flying over your assigned air space. The little blocks you see are simplified airplane data blocks. They display the speed and the altitude of the airplane they belong to. The number on top indicates the speed (yellow) and the number below is the altitude (white). The area that you see at the center is the hand-off region. This is the region in which planes have to fly at a particular speed and altitude so that you can safely hand them off to Local Control who will help them land. The rest of the screen is divided into 8 regions called “sectors”. At this point, I will be showing you your tasks in this experiment. It is natural that you have questions, but it might be more meaningful if they can wait until the end of the introduction section. Thank you for your patience!

Your primary responsibility is to monitor the center of the screen, a.k.a. the hand-off region. You must ensure that:
a. airplanes that arrive inside the circle fly at exactly 150 knots (point to or describe)

b. airplanes inside the circle strictly follow their prescribed paths (point to or describe the path on the screen)
Sometimes, it is possible that either of these are not followed. I will now show you what these look like, and how often they occur.

Quit MAINTASK
Start *MAINTASK* in Tutorial mode <<< 2:27 >>>
<wait for a minute>

While monitoring the screen, if you see either of these events occurring, you should rectify them immediately to prevent any danger from occurring to that airplane and those around it. If an airplane exceeds 150 knots speed, you must instruct it to decrease its speed by the correct amount (immediately click on it, select the difference). If an airplane swerves away from its path, you must click on it and instruct it to rectify its heading so that it follows the correct path. All actions in this entire experiment are performed using left-click only. You will never have to use the right click button! And you always want to click on the green plane itself and not on its data block.

Note that if you do not respond to these two events within the set period of time, the airplanes will correct themselves. Please let me know if you have any questions, or you can now try this out yourself. Remember, you will not see the training labels anymore - you have to look for the events yourself.

Quit MAINTASK

I would like to urge you, for the entire duration of this training, to lean back in your chairs and stretch out your torso. This will serve 2 purposes: 1. when we start using the tactors on your belly, it will help you to be able to feel them better and 2. it will help you maintain the same distance between you and the screen, which is important to keep as constant as possible throughout the experiment.

Start MAINTASK, un-check Tutorial mode <<< 2:27 >>>
Let them practice their maintasks. Watch closely to see if they are okay.
MAINTASK will quit

Thank you - you did great! Now we will look at your other responsibilities.

While you ensure that planes are safe in the hand-off region, airplanes in other sectors might require your attention. Although any plane can fly at the wrong altitude or speed at any time or in any space, for the purposes of this experiment, we will focus on speed changes only, as well as path changes, inside the circle and altitude changes only outside the circle. Now we will look at incidents that can occur outside the circle. All airplanes outside the circle should fly only at 4000 ft. If any airplane is flying at a wrong altitude, it will be reflected in its data block (that little white number) and you may be informed via a notification. I would like to show you how the notifications look/feel.

There are five types of notifications: visual, graded visual, tactile, graded tactile, and graded multimodal. Though they are of different modalities, they indicate the same thing. For example, a severe incident occurring at sector 6 can be denoted in one of five ways. You could get a visual notification in the form of blinking lights around that sector [show octagon pic]. Or you could get a tactile notification in the form of a buzz on your stomach corresponding to the sector on the screen [explain how tactors are set up]. "Graded" simply means that the cue will increase in strength over time, instead of remaining at the same level of salience. Graded multimodal notifications are special in that they are a combination of visual and tactile cues. This notification starts off as a visual cue and gradually becomes a tactile cue. You will only receive ONE of these five notification types in any experiment. You will also be told which notification type to expect before each experiments. Since the information they convey is the same, you can treat them all the same way.

In the following experiment, you will see what visual notifications look like.

Start 02-INCIDENT_SCRIPT_1-V <<< 4:40 >>> (count down to hit enter simultaneously)
Look at the command window to spot the first incident sector <severe>

The blinking corners indicate two things: a) the sector that the incident is taking place in, and b) the severity of the altitude drop. So when you perceive a notification, you can automatically tell where its coming from, and how bad the situation is. Note that when an incident occurs, you will receive four sets of cues in each notification. So even if you miss one set at first, you have a couple other chances to catch it before it disappears. Let us try this again, but this time you will experience *graded* visual cues, meaning the cues will increase in salience from one "cue-set" to the next. Note that this increase occurs *within each notification*, so each new incident will start at the lowest salience level.

Quit 02-INCIDENT_SCRIPT_1-V **(count down)**
Start 02a-INCIDENT_SCRIPT_1-GV <<< 4:40 >>>
Cues will be presented in this order: S, S, M, M, G, G, S, M, G
Look at the command window to spot the first incident sector <severe>

Do you see that cue? This means that an incident is occurring in the sector that is blinking. The pulse rate at which these lights are blinking indicates the severity of the incident. This high pulse means that the deviation is "severe", 3000 ft from normal.

In case you didn't catch that, there will be another severe incident soon after.

The next incident will be a moderate one.

Wait until the next incident occurs:

Look at this deviation now. The pulse rate is a little lower than before, meaning that the deviation is 2000 ft from normal. This is of "moderate" intensity.

You can see this again in the next moderate notification now.

Wait until the gentle incident occurs:

This deviation is a gentle one, only 1000 ft from normal, and so it is indicated by a low pulse rate.

There will be 4 more incidents. While you are watching for those, I'd like to point out that these peripheral visual cues will be easier to see when you are not looking at them directly. Light changes are easier to detect in your periphery.

02a-INCIDENT_SCRIPT_1-GV will quit

Now you saw the three types of incidents - gentle, moderate and severe. You were notified of them using visual notifications. I will now show you what the tactile cues feel like, and also tell you how to respond to notifications. There are two parts to responding to a notification: the acknowledgment and the task. The rule of thumb is that you always always acknowledge a cue immediately upon noticing it. This is very important. The second part, the "task", has to be completed only for moderate or severe notifications. So once again, you acknowledge ANY cue that you receive, and you perform a task ONLY for moderate and severe cues.

Minimize all windows and point to desktop image

Let us assume that you are getting a notification about a severe incident in sector 7 (bottom left). So immediately, you should first ACKNOWLEDGE the task by indicating what you just saw. (1) Hit the spacebar either once or twice or thrice in quick succession to denote the level of severity. In this case, since its severe, you hit the space bar three times quickly but distinctly. (2) Sector click to denote the sector in which the incident occurred and (3) if the incident is moderate or severe, plane click on the plane that is having trouble. [Demonstrate this yourself, then let them try - on the desktop screen.]

When you perceive a cue, the first thing you should do is hit the space bar corresponding to the intensity of the cue. Please make sure that you DO NOT look for the plane before hitting the space bar! If the cue is gentle, hit the spcbr once. If its moderate, hit it twice in quick succession. If its severe, hit it three times in quick succession. *As soon* as you hit the space bar, you should click on the sector (one of the 8 large squares) from which you received the cue. Only then, for moderate and severe incidents, should you identify the particular airplane that needs your attention. Remember that if you click on the plane before identifying the sector, the program will understand it to be a sector click and it will still be waiting for you to click on the plane.

So once again: spacebar, sector-click and then plane-click for moderate and severe incidents, spacebar, sector-click only **(no plane-click)** for gentle incidents. These steps have to be done in very quick succession, particularly the spacebar and sector-click, otherwise your actions will time out. If you acknowledge an incident correctly (either all 3, or all 2, steps must be completely correct), the outline of the airplane data block will turn green. If you make a mistake in any of the steps the outline will turn red and the plane will correct itself.

Now you should try acknowledging some incidents:

Start 03-INCIDENT_SCRIPT_2-T <<< 3:20 >>> (count down)

Cues will appear in the order of M, M, G, M, G G, M, G, M, M

[Wait for first notification.] Can you feel that cue on your stomach? Note how the pulse rate is the same as that in the visual notifications.

Sometimes you will see a bar appear at the bottom with a row of Xs. You might also feel all the tactors go off at the same time. Please ignore these for now. I will explain their significance in a few minutes.

Wait until they respond to all the notifications.

In this next script, you will again be experiencing tactile cues, but this time they will be *graded*, meaning the cues within a notification will gradually get stronger just like you saw with the graded visual notifications.

Start 03a-INCIDENT_SCRIPT_2-GT <<< 3:20 >>> (count down)

[Wait until they respond to a moderate incident so they can see they task bar.] Now I will tell you about the X-bar at the bottom. Notice how this appears when you got a moderate notification but not a gentle notification? As soon as you perform the acknowledgment (spacebar - sectorclick - planeclick) to identify a moderate or severe incident, you will be presented with this task bar. Recall that you will have to complete a task for every moderate or severe incident. Here is how you will handle the task. There are three '+' hidden among these 'X's. When you click on a '+' it will turn green. Once you find and click on all three, the task bar will disappear.

You have to complete this task as quickly as possible and get back to your main task. However, since this incident is only moderate, you can decide if you want to complete the task immediately or postpone it to a later time when your workload lightens up. Remember that while you perform these tasks, airplanes are still experiencing events at the center of the screen and that is your primary responsibility.

Now please try performing this task of identifying the '+'s.

Allow participant to try until finish.

Well done! You were shown how gentle and moderate notifications should be responded to. And you know that whenever you identify a moderate or severe cue, you will be presented with a task which you have to complete. As a note of caution, you should know that since severe incidents are of high importance, immediately after you respond to them, the rest of the screen will get disabled until you complete the task immediately.

Again, remember that only moderate and severe incidents will require you to complete a task. I will tell you more about moderate tasks. You can postpone a moderate task or perform your other duties while you complete it. Once a moderate task shows up, it will disappear from the screen after a few seconds, but it can be recalled by clicking on the empty task region. Note that if you start a task but do not finish it, when you recall that task, the pluses you already found will still be green. Moderate tasks stack up and will not go away until you complete them. You will get one reminder for each task you postpone. The reminder cue looks like 8 simultaneous cues (what we call a bubble cue). Depending on whether you are getting visual peripheral notifications or tactile notifications, you will either see all visual cues light up at once or feel all tactors buzz at once, respectively. Did you both feel all the tactors buzz at once during that last script? That was the tactile bubble cue. Do you feel comfortable in your ability to recognize that the next time you postpone a task? If not, we can start this script again and make sure you are familiar with the bubble cue.

Start 03-INCIDENT_SCRIPT_2-T <<< 3:20 >>>
Respond to moderate cue
Do not complete task
Show bubble-cue

Severe tasks, however, have to be completed immediately. The rest of the screen will get locked out, forcing you to complete the task before performing other operations. Now let us try a session to assimilate what you have learned so far. This time, lets switch to using graded notifications. Remember that the information you will receive is the same. However, instead of either a visual cue or a tactile cue, you will receive a mix of both. The notification will start off with a mild visual cue, then a stronger one, then a mild tactile cue and then a strong tactile cue. When I say "mild" and "strong," I am not referring to the severity of the incident. I am referring to the strength, or salience, of the cue. Severity depends only on the pulse-rate of the cue, not on its strength.

Start 04-STATIC_SINGLE-GVT <<< 6:08 >>> (computer on left requires mouse click before pressing enter)

04-STATIC_SINGLE-GVT will quit

That was a good attempt. In the next session, you will get notifications that sometimes change in their intensity. For example, a notification that starts off as gentle might become severe over the course of time. It is also possible that a severe notification becomes gentle. Note that if the severity changes when you are responding to a cue, do not worry - the program can handle it. Such dynamic changes are rare in the final experiment, but you will get plenty of practice in this one. This experiment will also feature graded multimodal notifications. Don't confuse them with dynamic notifications! :)

Start 05-DYNAMIC_SINGLE-GVT <<< 6:08 >>>

05-DYNAMIC_SINGLE-GVT will quit

Well done! Please let me know if you have any questions so far.

Now we will see another phenomenon - the dual onset. There are times when two incidents occur at once and you will have to attend to either one of them first. When that happens, you will receive cues from two different sectors simultaneously. In this case, your first reaction should be to hit the Ctrl button (you might find the one on your left easier to operate). This is to show that you did perceive two cues at once indeed. Once you do it, you have to respond to the notification with the higher intensity first. For example, if you received a gentle and a moderate cue at once, you should hit Ctrl, then hit spcbr twice for the moderate cue, click on that sector and the plane, and then hit spcbr once for the gentle cue, and click on that sector and then the plane. You can then proceed to monitor the planes in the center and complete tasks if they arise. Remember, you only have to hit the Ctrl button once when you perceive the simultaneous notifications. Note that if one of the cues is severe, you will still have to complete the task bar *immediately* before you are able to respond to the other cue (you will not be able to acknowledge them in quick succession like in the above example; you will have to do the task bar in between). It's possible that the other cue will time out before you have time to acknowledge it, depending on the time it takes to complete the task bar. Do you have an questions?

Now you can try and put that in context. This time again we'll have graded notifications. You will also experience an increase in the number of planes on the screen in the middle of the experiment. You will still have to keep working as hard as ever to save those planes!

Start 06-DYNAMIC_DUAL-GVT <<< 8:14 >>>

06-DYNAMIC_DUAL-GVT will quit

Excellent! A final most important thing to remember is that you will be asked to perform these same main tasks and incident-related tasks *without any notifications whatsoever!* During those sessions, you will be required to be alert and catch any plan that has trouble both inside and outside the hand-off region.

[Good time to offer bathroom/beverage/snack break.]

Now you are ready to begin your complete training sessions. By this time you must have become quite comfortable with the ATC interface. Before we begin, I would like to show you our point system.

Show point system sheet

Please take some time to review it. You will find that your best strategy is to follow all instructions, be as alert as you can, and respond to as many situations as possible. It is really easy to earn the bonus - you only have to border on the higher side of average, and I'm sure you can manage that! Now, onto the training. There are 4 training modules and you will be able to take quick 2-3 minute breaks in between. You can choose to skip the break if you wish, depending on how you feel. Each of your training sessions will be recorded and we will review your performance at the end of each session. I would like to stress to you the importance of taking every scenario very seriously and and doing the best you can in each one. The purpose of this research is to explore how people would perform in real control stations, given the experimental conditions. For this reason, you will be doing scientific research a great service if you give this your best shot! Ready? Good luck!

Take point system sheet away, start training

Please don't hit escape at the end of each session - I would like to see the scores.

FULL_TRAINING_1

Look at scores on screen, ask questions about what the participants felt. Take notes.

FULL_TRAINING_2

Look at scores on screen, ask questions about what the participants felt. Take notes.

FULL_TRAINING_3

Look at scores on screen, ask questions about what the participants felt. Take notes.

FULL_TRAINING_4

Look at scores on screen, ask questions about what the participants felt. Take notes.

1. **Conduct debriefing for each participant (together is fine, but ask both to give their own answers)**
2. **Have participants fill out cash payment form or write down their address to be added to payment-by-check spreaksheet**
3. **Bye bye :)**
4. **Turn back-up software back on**

Script # cheat sheet for Full Training scripts:

1 = **07-Training1-V-1**
2 = **08-Training2-T-2**
3 = **09-Training3-NC-3**
4 = **10-Training4-G-4**

Participant Instructions - DAY 2

Notes to researcher:
0. Exit Dropbox from system tray (bottom right) and turn off back-up software
1. Have participants fill out Background Day 2 Questionnaire
2. Attach tactor belt
3. Ask participants to adjust their chairs to a comfortable height
4. Seat the participant, shut door, turn light off
5. Ensure second monitor is turned on
6. Turn on flashlight
7. Turn on speakers
8. Ensure desktop background is ATC screenshot
9. Start experiment loader, Tutorial box is *unchecked*
10. Ask participant to type their name in the box
11. Show Visual Cue and Point Sheets to participants again

Welcome back! Hope you are all prepared to show us what great Air Traffic Controllers you are going to be today! Before we start, let us review the experiment real quick.

I would like to remind you that, for the entire duration of this training, to please lean back in your chairs and stretch out your torso. Again, this will serve 2 purposes: 1. it will help you to be able to feel the tactors better and 2. it will help you maintain the same distance between you and the screen, which is important to keep as constant as possible throughout the experiment.

I would also like to remind you that is it very important for you to take every scenario seriously and do the best you can. We realize that this experiment is long and repetitive, but it is essential to our research that you stay focused and motivated through the very end. To ensure this, please utilize the breaks available to you between scenarios so that you can get refreshed and continue to try your hardest to save those planes!

Start training sessions for each modality before the final experiment

Conduct all 4 experiment sessions

1. Conduct debriefing for each participant (together is fine, but ask both to give their own answers)
2. Fill out, sign, and hand back cash payment form, or add amount to payment-by-check spreadsheet
3. Bye bye :)
4. Turn back-up software back on

Script # cheat sheet for Day 2 Training scripts:

1 = 11-Day2-Practice-V-1
2 = 12-Day2-Practice-T-2
3 = 13-Day2-Practice-NC-3
4 = 14-Day2-Practice-G-4

Script # cheat sheet for Final Experiment scripts:

1 = 15-CompleteExperiment-V
2 = 16-CompleteExperiment-T
3 = 17-CompleteExperiment-NC
4 = 18-CompleteExperiment-G

Appendix III. Participant debriefing documents – Day 1

Debriefing Day 1

Name:

1. Please rate your level of fatigue after training.

not at all fatigued	1	2	3	4	5	extremely fatigued

2. Were there parts of the experiment that seemed particularly difficult? If so, please explain.

3. Please describe the strategies you adopted to perform your various responsibilities. For example, did you prioritize one thing over another? How did you handle situations where more than one thing required your attention? How did you utilize time(attention)-management?

4. Did you adopt any methods to trick/cheat the program to make your job easier or more enjoyable?

5. Please order the cue-types, from easiest to hardest, on the difficulty of purely detecting incidents outside of the handoff region.

1 (Easiest)
2
3
4
5 (Hardest)

6. Please order the cue-types, from easiest to hardest, on determining the correct level of severity of the incident by cue alone (before locating the plane or responding to the incident).

1 (Easiest)
2
3
4
5 (Hardest)

7. Please order the cue-types, from easiest to hardest, on determining the correct exact location (sector) of the incident by cue alone.

1 (Easiest)
2
3
4
5 (Hardest)

8. For which cue-type was it easiest to detect the "bubble" cue reminding you of a pending task?

* Visual Peripheral
* Tactile

9. Were there any issues with the current simulator setup? (anything out of place, awkward timing, etc.)

Additional Comments / Notes:

Appendix IV. Participant debriefing documents – Day 2

Debriefing Day 2

Name:

1. In your opinion, was the training adequate for performing your tasks in the experiment?

	1	2	3	4	5	
Inadequate						Excellent

If you felt that the training was less than adequate, please explain what task(s) you did not feel sufficiently prepared for.

2. Please rate your level of fatigue after the experiment.

	1	2	3	4	5	
not at all fatigued						extremely fatigued

3. Were there parts of the experiment that seemed particularly difficult? If so, please explain.

4. Please describe the strategies you adopted to perform your various responsibilities.

5. Did you adopt any methods to trick/cheat the program to make your job easier or more enjoyable?

6. Please order the cue-types, from easiest to hardest, on the difficulty of purely detecting incidents outside of the handoff region.

1 (Easiest)
2
3
4
5 (Hardest)

7. Please order the cue-types, from easiest to hardest, on determining the correct level of severity of the incident by cue alone (before locating the plane or responding to the incident).

1 (Easiest)
2
3
4
5 (Hardest)

8. Please order the cue-types, from easiest to hardest, on determining the exact, correct location (sector) of the incident by cue alone.

1 (Easiest)
2
3
4
5 (Hardest)

9. For which cue-type was it easiest to detect the "bubble" cue reminding you of a pending task?

* Visual Peripheral
* Tactile

10. What effect did having many different cue-types in the same session have?

11. Please rate your overall stress level during the combined-cues sessions on a scale from 1-10 (no stress-most stressed you have ever been).
1 2 3 4 5 6 7 8 9 10
no stress
most stressed I have ever been

12. Please describe your stress level during the combined-cues sessions in words.

13. Did you find that your stress level changed dramatically each time the workload changed during this session? Details can be added to question 12 above.

* Yes
* No
* Maybe

14. What was the most stressful aspect of the combined-cues sessions?

15. What was the most frustrating aspect of the experiment?

16. What did you like least about these sessions?

17. What did you like most about these sessions?

18. If there was an "annoyance factor" on a scale from 1-10 (least-most), how would you rate the single-level visual cues?

1 2 3 4 5 6 7 8 9 10
least most

19. If there was an "annoyance factor" on a scale from 1-10 (least-most), how would you rate the graded visual cues?

1 2 3 4 5 6 7 8 9 10
least most

20. If there was an "annoyance factor" on a scale from 1-10 (least-most), how would you rate the single-level tactile cues?

1 2 3 4 5 6 7 8 9 10
least most

21. If there was an "annoyance factor" on a scale from 1-10 (least-most), how would you rate the graded tactile cues?

1 2 3 4 5 6 7 8 9 10
least most

22. If there was an "annoyance factor" on a scale from 1-10 (least-most), how would you rate the graded multimodal cues?

1 2 3 4 5 6 7 8 9 10
least most

23. Do you have any suggestions for how we might improve upon our current notification design?

24. Were there any issues with the current simulator setup? (anything out of place, awkward timing, etc.)

Any other observations or general comments are appreciated!

Appendix V. Experiment Design and Incident Scheme

			Incident I				Incident II				
Incident #	**Single/ Concurrent**	**Static/ Dynamic**	**Stage a**	**Stage b**	**Stage c**	**Stage d**	**Stage a**	**Stage b**	**Stage c**	**Stage d**	**Workload**
1	1	S	S	S	S	S					L
2	1	S	M	M	M	M					L
3	1	S	G	G	G	G					L
4	1	S	S	S	S	S					L
5	1	D	M	M	S	S					H
6	1	S	M	M	M	M					H
7	1	D	G	G	M	M					H
8	1	D	G	G	S	S					H
9	2	S	M	M	M	M	G	G	G	G	L
10	1	S	M	M	M	M					L
11	2	S	G	G	G	G	S	S	S	S	H
12	1	S	M	M	M	M					H
13	2	S	M	M	M	M	G	G	G	G	L
14	1	S	M	M	M	M					H
15	1	S	S	S	S	S					L
16	1	S	G	G	G	G					H
17	2	S	M	M	M	M	S	S	S	S	H
18	1	S	S	S	S	S					H
19	1	D	G	G	M	M					H
20	2	S	M	M	M	M	S	S	S	S	H
21	1	S	M	M	M	M					H
22	1	S	G	G	G	G					H
23	2	S	S	S	S	S	M	M	M	M	H
24	1	D	G	G	S	S					H
25	1	D	M	M	S	S					L
26	1	S	M	M	M	M					L
27	1	S	G	G	G	G					L
28	1	S	M	M	M	M					L
29	1	D	G	G	M	M					L
30	2	S	S	S	S	S	G	G	G	G	L
31	1	D	M	M	S	S					L
32	2	S	G	G	G	G	M	M	M	M	L

Table 8-1: Un-cued session

			Incident I				Incident II					
Incident #	Single/ Concurrent	Static/ Dynamic	Stage a	Stage b	Stage c	Stage d	Stage a	Stage b	Stage c	Stage d	Workload	Cue type
1	1	D	M	M	S	S					L	T
2	1	S	M	M	M	M					L	GVT
3	2	S	G	G	G	G	M	M	M	M	L	V
4	1	S	M	M	M	M					L	GT
5	1	S	G	G	G	G					L	GVT
6	1	S	M	M	M	M					L	V
7	1	D	G	G	S	S					H	GVT
8	1	S	M	M	M	M					H	GV
9	1	S	S	S	S	S					H	T
10	1	S	M	M	M	M					H	GV
11	1	S	M	M	M	M					H	GT
12	2	S	S	S	S	S	G	G	G	G	H	GT
13	1	S	S	S	S	S					L	V
14	2	S	M	M	M	M	S	S	S	S	L	GVT
15	1	S	S	S	S	S					L	GT
16	1	S	S	S	S	S					H	V
17	2	S	G	G	G	G	M	M	M	M	H	GV
18	1	S	M	M	M	M					H	T
19	1	D	M	M	S	S					L	V
20	1	D	G	G	M	M					H	GT
21	2	S	M	M	M	M	S	S	S	S	L	V
22	1	D	G	G	M	M					H	GV
23	2	S	S	S	S	S	M	M	M	M	H	GV
24	1	S	M	M	M	M					H	T
25	2	S	M	M	M	M	S	S	S	S	H	GV
26	1	D	M	M	S	S					H	GVT
27	2	S	M	M	M	M	G	G	G	G	H	T
28	1	D	M	M	S	S					H	T
29	1	D	G	G	M	M					H	GV
30	1	S	M	M	M	M					H	GV
31	1	S	G	G	G	G					H	GV
32	2	S	G	G	G	G	M	M	M	M	L	T
33	1	D	G	G	S	S					L	GT
34	2	S	M	M	M	M	S	S	S	S	L	GT
35	1	D	G	G	M	M					L	T
36	1	S	M	M	M	M					L	V
37	1	S	G	G	G	G					L	T
38	1	S	M	M	M	M					L	T
39	2	S	M	M	M	M	G	G	G	G	L	GVT
40	1	S	G	G	G	G					L	GV

Table 8-2: Cued session A

			Incident I				Incident II					
Incident #	Single/ Concurrent	Static/ Dynamic	Stage a	Stage b	Stage c	Stage d	Stage a	Stage b	Stage c	Stage d	Workload	Cue type
1	1	D	G	G	S	S					L	GV
2	1	S	S	S	S	S					L	GVT
3	1	S	S	S	S	S					L	V
4	2	S	M	M	M	M	G	G	G	G	L	GV
5	1	S	M	M	M	M					L	T
6	1	D	G	G	M	M					L	T
7	1	S	G	G	G	G					H	GT
8	1	S	G	G	G	G					H	GT
9	1	S	M	M	M	M					H	V
10	2	S	M	M	M	M	G	G	G	G	H	GT
11	1	S	G	G	G	G					H	V
12	1	S	S	S	S	S					H	GVT
13	2	S	S	S	S	S	G	G	G	G	L	T
14	1	D	M	M	S	S					L	V
15	2	S	S	S	S	S	G	G	G	G	L	GV
16	1	S	S	S	S	S					H	GT
17	1	S	M	M	M	M					H	GVT
18	1	S	G	G	G	G					H	T
19	1	S	S	S	S	S					L	GV
20	1	D	G	G	M	M					H	GV
21	2	S	M	M	M	M	G	G	G	G	L	GVT
22	1	S	G	G	G	G					H	GV
23	1	S	G	G	G	G					H	V
24	1	S	M	M	M	M					H	V
25	1	S	S	S	S	S					H	GT
26	2	S	S	S	S	S	M	M	M	M	H	T
27	1	D	M	M	S	S					H	T
28	1	D	G	G	M	M					H	GT
29	1	S	S	S	S	S					H	GV
30	2	S	S	S	S	S	M	M	M	M	H	GT
31	1	D	G	G	S	S					H	GV
32	1	D	G	G	M	M					L	V
33	1	S	S	S	S	S					L	GT
34	1	D	G	G	M	M					L	GVT
35	2	S	S	S	S	S	G	G	G	G	L	V
36	1	D	G	G	M	M					L	V
37	1	S	G	G	G	G					L	GT
38	2	S	M	M	M	M	S	S	S	S	L	T
39	1	S	M	M	M	M					L	GVT
40	1	D	M	M	S	S					L	GVT

Table 8-3: Cued session B

			Incident I				Incident II					
Incident #	Single/ Concurrent	Static/ Dynamic	Stage a	Stage b	Stage c	Stage d	Stage a	Stage b	Stage c	Stage d	Workload	Cue type
1	1	S	M	M	M	M					L	GV
2	1	S	M	M	M	M					L	GT
3	1	S	M	M	M	M					L	V
4	2	S	G	G	G	G	S	S	S	S	L	GVT
5	1	D	G	G	S	S					L	T
6	2	S	M	M	M	M	S	S	S	S	L	GT
7	1	D	G	G	M	M					H	GVT
8	1	S	S	S	S	S					H	T
9	2	S	M	M	M	M	S	S	S	S	H	GV
10	1	S	G	G	G	G					H	GVT
11	2	S	M	M	M	M	G	G	G	G	H	V
12	1	D	M	M	S	S					H	GVT
13	2	S	G	G	G	G	S	S	S	S	L	T
14	1	D	G	G	M	M					L	V
15	1	D	G	G	M	M					L	GVT
16	1	D	G	G	S	S					H	T
17	2	S	M	M	M	M	G	G	G	G	H	T
18	1	S	G	G	G	G					H	V
19	2	S	G	G	G	G	S	S	S	S	L	GT
20	1	D	M	M	S	S					H	GT
21	2	S	G	G	G	G	M	M	M	M	L	GVT
22	1	S	M	M	M	M					H	GV
23	2	S	G	G	G	G	M	M	M	M	H	GT
24	1	S	S	S	S	S					H	GV
25	1	D	G	G	S	S					H	V
26	1	S	M	M	M	M					H	GT
27	2	S	M	M	M	M	S	S	S	S	H	V
28	1	S	S	S	S	S					H	GV
29	2	S	S	S	S	S	M	M	M	M	H	V
30	1	D	M	M	S	S					H	GV
31	1	S	G	G	G	G					H	GVT
32	1	S	S	S	S	S					L	T
33	2	S	M	M	M	M	G	G	G	G	L	GV
34	1	S	M	M	M	M					L	T
35	1	S	M	M	M	M					L	GT
36	1	S	S	S	S	S					L	GVT
37	1	D	G	G	S	S					L	V
38	1	S	M	M	M	M					L	GT
39	1	D	G	G	M	M					L	GT
40	1	S	M	M	M	M					L	GVT

Table 8-4: Cued session C

			Incident I				Incident II					
Incident #	Single/ Concurrent	Static/ Dynamic	Stage a	Stage b	Stage c	Stage d	Stage a	Stage b	Stage c	Stage d	Workload	Cue type
1	1	S	M	M	M	M					L	GVT
2	1	D	G	G	S	S					L	GVT
3	1	S	M	M	M	M					L	T
4	1	S	M	M	M	M					L	GV
5	2	S	S	S	S	S	M	M	M	M	L	GVT
6	1	S	M	M	M	M					L	T
7	1	S	M	M	M	M					H	GV
8	1	S	M	M	M	M					H	V
9	2	S	G	G	G	G	S	S	S	S	H	V
10	1	S	S	S	S	S					H	GVT
11	1	D	M	M	S	S					H	GT
12	1	D	M	M	S	S					H	V
13	1	S	M	M	M	M					L	T
14	1	S	M	M	M	M					L	GT
15	2	S	G	G	G	G	S	S	S	S	L	GV
16	1	S	M	M	M	M					H	GVT
17	1	S	G	G	G	G					H	T
18	1	D	M	M	S	S					H	GV
19	1	S	M	M	M	M					L	GV
20	1	S	G	G	G	G					H	V
21	2	S	M	M	M	M	G	G	G	G	L	GT
22	1	D	M	M	S	S					H	GT
23	1	S	G	G	G	G					H	GT
24	1	S	G	G	G	G					H	T
25	2	S	M	M	M	M	G	G	G	G	H	V
26	1	S	M	M	M	M					H	GT
27	2	S	M	M	M	M	S	S	S	S	H	GVT
28	1	S	M	M	M	M					H	V
29	1	S	S	S	S	S					H	V
30	2	S	S	S	S	S	G	G	G	G	H	GVT
31	1	S	M	M	M	M					H	V
32	1	S	M	M	M	M					L	GVT
33	1	S	G	G	G	G					L	GVT
34	2	S	M	M	M	M	S	S	S	S	L	T
35	1	D	G	G	S	S					L	GT
36	1	S	M	M	M	M					L	GVT
37	1	D	M	M	S	S					L	GV
38	1	S	S	S	S	S					L	T
39	1	S	G	G	G	G					L	GV
40	1	D	G	G	M	M					L	T

Table 8-5: Cued session D

Bibliography

Anderson, J. R. (2004). *Cognitive psychology and its implications*.: Worth Publishers.

Bolanowski, S. J., Gescheider, G. A., and Verrillo, R. T. (1994). Hairy skin: Psychophysical channels and their physiological substrates. *Somatosensory and Motor Research, 11,* 279-290.

Bond, B., and Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception and Psychophysics, 6*(337), 339

Bonferroni, C. E. (1936). Theory statistics of the classes and calculation of the probabilities (in Italian). *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8,* 3-62.

Brewster, S. A., and Brown, L. M. (2004). *Tactons: Structured tactile messages for non-visual information display.* Dunedin, New Zealand: Australian Computer Society, Inc..

Brewster, S. A., and Brown, L. M. (2004). *Non-visual information display using tactons.* Vienna, Austria: ACM Press.

Broadbent, D. E. (1977). The hidden preattentive processes. *American Psychologist, 32*(2), 109-118.

Cholewiak, R. W., Brill, J. C., and Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Perception and Psychophysics, 66* (970), 987

Cook, R. I., Potter, S., Woods, D. D., and McDonald, J. S (1991). Evaluating the human engineering of microprocessor controlled operating room devices. *Journal of Clinical Monitoring, 7,* 217-226.

Czerwinski, M. P., Chrisman, S. E., and Rudisill, M. (1991). *Interruptions in multitasking situations: The effects of similarity and warning.* Houston, Texas: Aeronautics and Space Administration, Lyndon B. Johnson Space Center.

Czerwinski, M.P., Chrisman, S.E., and Schumacher, B. (1991). The effects of warnings and display similarities on interruptions in multitasking environments. *SIGCHI Bulletin, 23*(4), 38-39.

Deouell, L. Y., Bentin, S., and Soroker, N. (2000). Electrophysiological evidence for an early (pre-attentive) information processing deficit in patients with right hemisphere damage and unilateral neglect. *Brain, 123,* 353-365.

Dobbins, T., and Samways, S. (2002). *The use of tactile navigation cues in high-speed craft operations.* London, UK: The Royal Institution of Naval Architects.

Donders, F. C. (Translated by Koster, W. G.) (1868). On the speed of mental processes. *Acta Psychologica, 30,* 412-431.

Driver, J (1996). *What can visual neglect and extinction reveal about the extent of "preattentive" processing?.* Washington, DC: APA Press.

Driver, J., and Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences, 2,* 254-262.

Driver, J., and Spence, C. (1998). Crossmodal attention. *Current Opinion in Neurobiology, 8,* 245-253.

Driver, J., and Spence, C. (2004). *Crossmodal spatial attention: Evidence from human performance.* Oxford, UK: Oxford University Press.

Drury, C. G., and Clement, M. R. (1978). The effect of area, density, and number of background characters on visual search. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 20*(5), 597-602.

Easterbrook, J. A. (1959). The effects of emotion on cue utilization and the organization of behavior. *Psychological Review, 66* (May), 183-201.

Egeth, H. E., and Dagenbach, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 551-560.

Ferris, T. K., and Sarter, N. B (2010). When content matters: The role of processing code in tactile display design. *IEEE Transactions on Haptics, 3*(3), 199-210.

Folk, C. L., and Remington, R. W. (2008). Bottom-up priming of top-down attentional control settings. *Visual Cognition, 16*(2), 215-231.

Folk, C. L., and Remington, R. W. (2006). Top-down modulation of preattentive processing: Testing the recovery account of contingent capture. *Visual Cognition, 14,* 465-465.

Franke, E. K., von Gierke, H. E., Oestreicher, H. L., and von Wittern, W. W. (1951). *The propagation of surface waves over the human body.* Dayton, OH: United States Air Force, Aeromedical Laboratory.

Funk, K., and Braune, R. (1999, October 19). *Agenda Manager: A knowledge-based system to facilitate the management of flight deck activities.* San Francisco, CA: 1999 World Aviation Congress.

Gillie, T., and Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research, 50,* 243-250.

Greenspan, J. D., and Bolanowski, S. J. (1996). *The psychophysics of tactile perception and its peripheral physiological basis.* San Diego: Academic Press.

Hameed, S., Ferris, T. K., Jayaraman, S., and Sarter, N. B. (2006). *Supporting interruption management through informative tactile and peripheral visual cues.* Santa Monica, CA: Human Factors and Ergonomics Society.

Hancock, P. A., and Dirkin, G. R. (1983). *Stressor induced attentional narrowing: Implications for design and operation of person-machine systems.* Toronto, ON, Canada: Human Factors Association of Canada.

Hancock, P. A., and Hart, S. G. (2002). Defeating terrorism: What can human factors/ergonomics offer?. *Ergonomics and Design, 10,* 6-16.

Henson, D. B. (1993). *Visual fields.* Oxford, UK: Butterworth-Heinemann Ltd.

Hess, S. M., and Detweiler, M. (1994). *Training to reduce the disruptive effects of interruptions.* Nashville, TN: Human Factors and Ergonomics Society.

Ho, C-Y, and Sarter, N. B. (2004). *Supporting synchronous distributed communication and coordination through multimodal information exchange.* New Orleans, LS: Human Factors and Ergonomics Society.

Ho, C-Y, Nikolic, M. I., Waters, M. J., and Sarter, N. B. (2004). Not now!: Supporting interruption management by indicating the modality and urgency of pending tasks. *Human Factors, 46*(3), 399-409.

Ho, C-Y, Tan, H. Z., Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F, 8,* 397-412.

Hopp-Levine, P. J., Smith, C. A. P., Clegg, B. A., and Heggestad, E. D. (2006). Tactile interruption management: Tactile cues as task-switching reminders. *Cognition, Technology and Work, 8,* 137-145.

Intriligator, J., and Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171-216.

James, W. (1890). *The principles of psychology.* New York, NY: Henry Holt.

Johnson K. O., and Phillips J. R. (1981). Tactile spatial resolution. I: Two-point discrimination, gap detection, grating resolution, and letter recognition. *Journal of Neurophysiology, 46,* 1177-1191.

Joint Planning and Development Office (JPDO) (2007). *Concept of operations for the next generation air transportation system. Version 1.2.* Washington, DC: JPDO.

Jones, L. A., and Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors, 50*(1), 90-111.

Jones, L. A., Lockyer, B., and Piateski, E. (2006). Tactile display and vibrotactile pattern recognition on the torso. *Advanced Robotics, 20*(12), 1359-1374.

Jonides, J. (1981). *Voluntary versus automatic control over the mind's eye's movement.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Joseph, J. S., Chun, M. M., and Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature, 387,* 805-808.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kemeny, J. C., Babbitt, B., Haggerty, P. E., Lewis, C., Marks, P. A., Marrett, C. B., McBride, L., McPherson, H. C., Peterson, R. W., Pigford, T. H., Taylor, T. B., and Trunk, A. D. (1979). *The President's commission of the accident at Three Mile Island.* Washington, DC: U.S. Government Printing Office.

Laming, D. R. (1968). *Information theory of choice-reaction times.* London, UK: Academic Press.

Latorella, K. A. (1998). *Effects of modality on interrupted flight deck performance.* Chicago, IL: Human Factors and Ergonomics Society.

Latorella, K. A. (1996). *Investigating interruptions: Implications for flight deck performance.* Buffalo, NY: (Also published in 1999 as NASA Technical Memorandum 209707).

Lee, J. D., Hoffman, J., and Hayes, E. M. (2004). *Collision warning design to mitigate driver distraction.* Vienna, Austria: ACM Press.

Lee, J. D., McGehee, D. V., Brown, T. L., and Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44,* 314-334.

Lee, W. O. (1992). *The effects of skill development and feedback on action slips.* University of York, UK: British Computer Society Human Computer Interaction Specialist Group.

Mackworth (1965). Visual noise causes tunnel vision. *Psychonomic Science, 3*(67), 68

Martin, C., Schreckenghost, D., Bonasso, P., Kortenkamp, D., Milam, T., and Thronesbery, C. (2003). *An environment for distributed collaboration among humans and software agents.* Melbourne, Australia: The International Foundation for Autonomous Agents and Multiagent Systems.

Martin, C., Schreckenghost, D., Bonasso, P., Kortenkamp, D., Milam, T., and Thronesbery, C. (2003). *Helping humans: Agents for distributed space operations.* Nara, Japan: Japanese Aerospace Exploration Agency.

Mattingley, J. B., Davis, G., and Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science, 275,* 671-674.

McCarley, J. S., Wickens, C. D., Goh, J., and Horrey, W. J. (2003). *A computational model of attention/situation awareness.* Baltimore, MD: Human Factors and Ergonomics Society.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423-1442.

Miyata, Y., and Norman, D. A. (1986). *Psychological issues in support of multiple activities.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Moll van Charante, E., Cook, R. I., Woods, D. D., Yue, L., and Howie, M. B. (1992). *Human-computer interaction in context, physician interaction with automated intravenous controllers in the heart room.* New York, NY: Pergamon Press.

Moy, G., Singh, U., Tan, E., and Fearing, R. S. (2000). Human psychophysics for teletaction system design. *Haptics-e, The Electronic Journal of Haptics Research, 1,*

Murray, C., and Cox, C. B. (1989). *Apollo: The race to the moon.* New York, NY: Simon and Schuster.

Müller, H. J., and Rabbitt, M. A (1989). Spatial cuing and the relation between the accuracy of "where" and "what" decisions in visual search. *Quarterly Journal of Experimental Psychology, 41,* 747-773.

Müller, H. J., and Rabbitt, M. A. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315-330.

Nagy, A. L., and Sanchez, R. R. (1992). Chromaticity and luminance as coding dimensions in visual search. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 34*(5), 601-614.

Neisser, U. (1967). *Cognitive psychology.* New York, NY: Appleton-Century -Crofts.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco, CA: W. H. Freeman.

Nikolic, M. I., and Sarter, N. B. (2001). Peripheral visual feedback: A powerful means of supporting effective attention allocation in event-driven, data-rich environments. *Human Factors, 43*(1), 30-38.

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

Nolan, F. (1994). *Phonetic correlates of syllable affiliation.* New York, NY: Cambridge University Press.

Obermayer, R. W., and Nugent, W. A. (2000). *Human computer interaction for alert warning and attention allocation systems of the multi-modal watch station.* Bellingham, WA: The International Society for Optical Engineering.

Ohman, A. (1997). *As fast as the blink of an eye: Evolutionary preparedness for preattentive processing of threat.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Oviatt, S. L. (2002). Breaking the robustness barrier: Recent progress in the design of robust multimodal systems. *Advances in Computers, 56,* 305-341.

Oviatt, S. L., Cohen, P., Wu, L., Vergo, J., Duncan, L., Subh, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., and Ferro, D. (2000). Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Human Computer Interaction, 15*(4), 263-322.

Penfield, W., Rasmussen, T. (1950). *The cerebral cortex of man: A clinical study of localization of function.* New York, NY: Macmillan.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

Priplata, A. A., Niemi, J. B., Harry, J. D., Lipsitz, L. A., and Collins, J. J. (2003). Vibrating insoles and balance control in elderly people. *The Lancet, ,* 1123-1124.

Rabbitt, P. (1984). *The control of attention in visual search.* New York, NY: Academic Press.

Rogers, R. D., and Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology, 124*(2), 207-231.

Rupert, A. H. (2000). Tactile situation awareness system: Proprioceptive prostheses for sensory deficiencies. *Aviation Space and Environmental Medicine, 71*(Suppl.), A92-A99.

Ruz, M., and Lupiáñez, J. (2002). A review of attentional capture: On its automaticity and sensitivity to endogenous control. *Psicológica, 23*(2), 283-309.

Sarter, N. B. (2000). The need for multisensory feedback in support of effective attention allocation in highly dynamic event-driven environments: The case of cockpit automation. *International Journal of Aviation Psychology, 10*(3), 231-245.

Sarter, N. B. (2002). *Multimodal information presentation in support of human-automation communication and coordination..* New York, NY: JAI Press.

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics, 36*(5), 439-445.

Sarter, N. B. (2007). *Multiple-resource theory as a basis for multimodal interface design: Success stories, qualifications, and research needs.* New York, NY: Oxford University Press.

Sarter, N. B. (2007). *Human-computer interaction: HCI intelligent multimodal interaction environments.* Beijing, China: HCI International.

Sarter, N. B. (2005). *Graded and multimodal interruption cueing in support of preattentive reference and attention management.* Orlando, FL: Human Factors and Ergonomics Society.

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

Sorkin, R. D., Kantowitz, B. H., and Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.

Speier, C., Valacich, J. S., and Vessey, I. (1997). *The effects of task interruption and information presentation on individual decision making.* Atlanta, GA: Association for Information Systems.

Spence, C., and Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception and Psychophysics, 59,* 1-22.

Spence, C., McDonald, J., and Driver, J. (2004). *Exogenous spatial-cuing studies of human crossmodal attention and multisensory integration.* Oxford, UK: Oxford University Press.

Tan, H. Z., Lu, I., and Pentland, A. (1997). *The chair as a novel haptic user interface.* Redmond, WA: Microsoft Research.

Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., and Zelinsky, G. J. (1999). Influnce of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1595-1608.

Trafton, J. G., and Monk, C. A. (2008). *Task interruptions.* Santa Monica, CA: Human Factors and Ergonomics Society.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing, 31,* 156-177.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American, 255,* 106-115.

Treisman, A. (2005). *Synesthesia: Implications for attention, binding, and consciousness—A commentary.* Oxford, UK: Oxford University Press.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., and Crundall, D. (2003). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics, 46*(6), 629-646.

Van Erp, J. B. F. (2000). *Direction estimation with vibro-tactile stimuli presented to the torso: A search for the tactile ego-centre.* Soesterberg, The Netherlands: TNO Human Factors.

Van Erp, J. B. F., and Werkhoven, P. J. (1999). *Direction estimation with vibro-tactile stimuli presented to the torso: A search for the tactile ego-centre.* Soesterberg, The Netherlands: TNO Human Factors.

Weber, E. H. (1978). *The sense of touch.* New York, NY: Academic Press.

Weinstein, S. (1968). *Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality.* Springfield, IL: Thomas.

Wickens, C. D (1980). *The structure of attentional resources.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C. D. (1984). *Processing resources in attention.* Orlando, FL: Academic Press.

Wickens, C. D. (1992). *Engineering psychology and human performance (2nd Edition).* New York, NY: Harper Collins.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences, 3,* 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., and Hollands, J. G. (2000). *Engineering psychology and human performance (3rd Edition).* Upper Saddle River, NJ: Pearson/Prentice Hall.

Wickens, C. D., Mavor, A. S., Parasuraman, R., and McGee, J. P. (1998). *The future of air traffic control: Human operators and automation.* Washington, DC: National Academy Press.

Williams, L. J. (1988). Tunnel vision or general interference? Cognitive load and attentional bias are both important. *The American Journal of Psychology, 101*(2), 171-191.

Wolfe, J. M., Butcher, S. J., Lee, L., and Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 483-502.

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 419-433.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics, 38*(11), 2371-2394.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., and Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *The Journal of Neuroscience, 17*(10), 3956-3963.

Yantis, S., and Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 121-134.

Yantis, S., and Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*(601), 621

Zompa, I. C., and Chapman, C. E. (1995). Effects of cross-modal manipulations of attention on the ability of human subjects to discriminate changes in texture. *Somatosensory and Motor Research, 12*(2), 87-102.