ABSTRACT

WARREN, HEATHER LYNN. Auditory Cueing Effects on Human Performance with an Adaptive System. (Under the direction of David B. Kaber.)

Adaptive automation (AA) is the dynamic allocation of complex system functions to a human operator and/or automated controller on the basis of the state of the human-task-environment and having the objective of optimizing overall system performance. Adaptive automation has been successfully applied to different types of simulated tasks in laboratory settings; however, there have been few field applications and problems remain in providing human operators of adaptive systems with adequate feedback on changing system states and modes of operation. This may result in poor operator situation awareness (SA). Previous AA research on system feedback mechanisms has primarily focused on visual cues of system state changes. Some work has investigated complex auditory icons, but no research has considered the use of vocal cues on adaptive system state changes.

The current research investigated the potential for multimodal interfaces to improve adaptively automated system performance by considering multiple resource theory (MRT) of attention in system interface design. Subjects were provided with feedback on AA states via their visual and auditory senses in order to improve overall system performance and human-automation interaction. An experiment was conducted to compare the use of visual (icons), auditory (earcons) and vocal cues for conveying the state of an adaptive teleoperator (remote-control robot) in a high-fidelity, virtual reality simulation of an underwater mine disposal task. Earcons have been found to be beneficial for cueing operators of automated system states, but there has been no research to investigate earcons as feedback mechanisms in either complex human-machine

interaction or in adaptively automated systems. In this study, modal cues were associated with task phase changes and teleoperator control mode changes.

The type of cue and level of cue complexity (level of detail) was varied between and within subjects, respectively. Operator performance was evaluated in terms of system-state awareness, accurate control commands, and time-to-task completion. The research sought to discover which cue type was most effective for facilitating overall system performance and maintaining operator SA.

Results demonstrated the manner in which humans use visual and auditory sensory cues for feedback when dealing with adaptively automated systems is in agreement with MRT. Vocal cues were identified as being superior for warning operators of system-state changes, maintaining SA (attentional resources) and facilitating overall complex system performance. The results of this study are applicable to the design of future automated systems and may serve to improve efficiency and effectiveness of performance.

AUDITORY CUEING EFFECTS ON HUMAN PERFORMANCE WITH AN ADAPTIVE SYSTEM

by
HEATHER LYNN WARREN

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

INDUSTRIAL ENGINEERING

Raleigh

2002

APPROVED BY:

Chair of Advisory Committee

DEDICATION

I dedicate this thesis to Louise "Crump" Crumpler. Crump served as a motivation not only in my life, but also in the lives of all those she touched. She supported me in educational endeavors right from the start. She taught me things that simply cannot be learned in the classroom. For all the love, I thank you. You will be deeply missed by all.

BIOGRAPHY

Heather L. Warren is originally from Suffolk, VA. She grew up there with her mom, dad, and brother, Russ. She graduated from Nansemond-Suffolk Academy in Suffolk, VA in 1996. She then went on to receive her B.S. in Integrated Science and Technology with concentrations in energy, environment, and biotechnology from James Madison University in Harrisonburg, VA. This thesis served as the completion to her M.S. in Industrial Engineering with a concentration in Ergonomics from N.C. State University in Raleigh, NC. She is interested in an integrated approach to problem solving involving practical solutions from a variety of disciplines. In the past, she has worked at U.S. Army Corps of Engineers, N.C. State University, The International Institute for Energy Technology in Marsaxlokk, Malta, Agua del Pueblo in Quelzaltenago, Guatemala and Solid Surface Technologies.

ACKNOWLEDGEMENTS

I would like to thank my family and friends for support over the past two years. I would also like to thank Dr. David Kaber, Dr. Gary Mirka, and Dr. Michael Wogalter for academic advisement. In addition, I would also like to thank Mohamed Sheiknainar for the development of the virtual environment used in this research. As well, I would also like to thank the United States Navy, Office of Naval Research for support through a research assistantship in Industrial Engineering at North Carolina State University, without which my studies would not have been possible.

TABLE OF CONTENTS

Page

List of Tables

LIST OF FIGURES

GLOSSARY

AA – Adaptive Automation

ATC – Air Traffic Control

BOOM – Binocular Omni-Orientation Monitor

CAVE – Computer Automatic Virtual Environment

CRT – Cathode Ray Tube

DOF – Degree-of-Freedom

HDVD – High Definition Volumetric Display

HMD – Head-Mounted Display

HSD – Honestly Significant Difference

IPD – Inter-Pupillary Distance

LOA – Level of Automation

MR – (Duncan's) Multiple Range

MRT – Multiple Resource Theory

SA – Situation Awareness

SART – Situation Awareness Rating Technique

SGI – Silicon Graphics

SSQ – Simulator Sickness Questionnaire

TLX – Task Load Index

VE – Virtual Environment

VR – Virtual Reality

1. LITERATURE REVIEW

1.1 INTRODUCTION

Adaptive automation (AA) is considered to provide the capability to improve complex systems control over traditional, conventional automation; however, there are many problems currently associated with its implementation (Scerbo, 1996, Kaber, Riley, Tan, & Endsley, 2000). One proposed approach to effective implementation of AA is to incorporate multimodal interfaces into the design of systems to improve human-automation interaction. Kaber et al. (in preparation) investigated the effectiveness of multimodal interface cueing on automation-state changes in adaptive system operation. Specifically, auditory and visual cues of adaptive system state changes were evaluated in terms of operator situation awareness (SA) and performance in an underwater teleoperated virtual reality (VR) demining task. The goal of their experiment was to reduce decrements in system performance that are typically associated with AA, as a result of poor human-automation communication. This problem is generally associated with reduced operator mode awareness and longer duration recovery times in the event of system errors. Both modal and bimodal cues were evaluated. They included musical earcons, semi-abstract icons, and both earcons and icons presented simultaneously. Cues were provided to operators to indicate impending changes in the level of system automation (manual control or supervisory control) and the phase of the task (search for mines, place depth charges, or detonate mines). Kaber et al. (in preparation) found bimodal cues to be superior to modal cues (particularly visual stimuli). As well, cueing with earcons produced significantly fewer errors than cueing with icons. Operator perception of system states, or Level 1 SA, was also superior with bimodal cues. In addition, bimodal cues produced shorter trial times. Although these results are promising for interface design for effective implementation of AA, Kaber et al.'s (in preparation) research did not investigate the use of vocal cues. Vocal cues have the potential advantage of reducing the translation that may be associated with interpreting the meanings of earcons. Also, the meanings of vocal cues may be easier for subjects to learn than earcons or icons because vocal cues require natural language processing. As a result, vocal cues may produce shorter trial times and better performance.

The following literature review covers AA and research on interface design for effective implementation of AA. It also covers earcons, their design, and potential benefits as advanced warning cues in complex systems, as compared to other modal cues. Finally, virtual reality systems are reviewed as test-beds for human factors research, and AA studies in particular, and a suitable experimental setup is identified for this research. The review is intended to serve as a basis for work involving assessment of vocal messages, earcons, and icons for operator cueing in adaptively automated systems.

1.2 ADAPTIVE AUTOMATION

The purposes of automation include performing tasks that humans do not want to do, do poorly, or are not well suited to do (Scerbo, 1996). Traditional automation is defined as the outsourcing of tasks to a machine to perform (Parsons, 1985). Traditional automation can result in performance problems such as reduced operator SA, which may occur when humans are removed from complex system control loops (Kaber, Riley, Tan, & Endsley, 2001). Other problems include an increased monitoring load on operators. Automation often requires a human supervisor to oversee the functioning of many machines simultaneously.

Unfortunately, humans are not effective monitors of information sources for extended periods of time, and this situation often results in vigilance decrements (Parasuraman, 1986). Scerbo (1996) has noted that introducing automation into a system that was previously non-automated may actually increase operator workload due to monitoring requirements. Over the long term, skill decrements can also result from traditional automation because the human adapts to a supervisory control role in the process and does not exercise active decision-making or manual control (Shiff, 1983). If a situation develops that requires the human to intervene in the control loop, he or she may not remember how to perform tasks manually as a result of the automation (Scerbo, 1996).

Adaptive automation is defined as "the dynamic allocation of tasks or control functions between human operators and automated control systems over time based on the state of the human-task-environment" with the objectives of moderating operator workload, yet maintaining situation awareness at the same time (Kaber, 1999). Adaptive automation, by nature, incorporates multiple levels of automation (LOAs) into a single system. Adaptive automation may be beneficial to complex system performance because it considers the human in the design cycle and ensures the human is retained in the control loop during normal operations in order to deal with potential automation errors and failures.

Adaptive automation is a new and evolving technology that is not thoroughly understood. Optimal AA strategies for human-computer-interaction, team scenarios, and operator performance have not yet been realized (Kaber, Riley, Endsley, & Tan, 2001). Several strategies have been presented to create an effective adaptively automated system. The first strategy is to alter the level of automation based upon the human's interactions with the system. The operator's interactions are judged against a situational database to determine the optimum interactions. If the operator is not interacting with the system in the most effective manner, the system may switch to a higher LOA (Morris & Gluckman, 1994). The main problems associated with this strategy are the complexity of building the database and the retroactive response (Scerbo, 1996). A more proactive strategy involves modeling human performance against actual operator performance. The modeled behaviors are compared to the actual behaviors. Like the previous approach, the system would be capable of switching LOAs when needed (Scerbo, 1996). Hancock and Chignell (1987, 1988) proposed a slightly different strategy, in that they used current and future workload of the operator as a means for allocating tasks to the operator or the machine. Another proposed strategy to AA is using biopsychometrics, such as heart rate variability, eye movements, and event-related brain potentials to trigger a change in the LOA (Morris & Gluckman, 1994, Byrne & Parasuraman, 1996). Morris and Gluckman (1994) also advocated monitoring the task and looking for critical events that might serve as a basis for altering the LOA. However, this strategy has proved to be ineffective for complex systems because the event driven change is not directly correlated with performance or workload (Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992).

The ultimate goal of AA is to determine dynamic function allocation strategies that optimize human operator performance, situation awareness and workload (Kaber & Riley, 1999). Many solutions have been proposed to alleviate the problems created by various AA strategies. Among these solutions is the creation of more effective interfaces. Providing feedback on mode changes, such as cues may improve the effectiveness of AA strategies because the human is kept more in the information loop, therefore, the operator's performance and SA may increase (Kaber, Hughes, Wright, Sheiknaniar, & Warren, in preparation).

1.2.1 Adaptive Automation Interfaces

The main problem that occurs as a result of AA in complex systems is a lack of operator knowledge concerning the state of the system at any given point in time because of often inadequate system feedback. One potential solution to this problem is the development of more effective interfaces for AA systems. There are three types of interface solutions that have been proposed in the literature. They include dynamic displays, multimodal displays, and direct manipulation displays.

Weiner (1988) introduced the concept of dynamic displays into the cockpit setting with the objective of improving pilot situation awareness. Specific interfaces are introduced over time based on the LOA to be employed. Displays can also be selected or deselected based on the operator's task requirements. In order to support human operator performance in interacting with AA, these displays must be consistently designed to allow users to smoothly switch among different LOAs without getting lost due to complex changes within the interface. The interface design must ensure that users remain aware of what functions they are controlling and what functions the automation is controlling (Norman, 1990).

It is possible that interfaces displaying the level of information required for operating some AA systems may actually increase the complexity of the system (Kaber, 2001). Consequently, dynamic interfaces have been designed and evaluated for supporting human performance and situation awareness, but constantly changing interface features may confuse operators. One method of addressing this problem would be to incorporate multimodal displays into the system (Kaber, 2001). Such displays may capitalize on MRT by allowing for presentation of all critical system information without overwhelming the human with information along a single modality (Scerbo, 1996). Multimodal sensory cues can be an effective means of notifying the human of LOA changes or other changes within a system (Kaber, Hughes, Wright, Sheiknainar, & Warren, in preparation). In theory, multimodal displays may increase situation awareness and performance, while reducing cognitive load.

Ballas et al. (1991) hypothesized that information processing required to resolve user intentions and adaptive machine states would be reduced when using a direct manipulation interface because the user's mental model is more accurately matched by the interface functions and controls. A benefit of direct manipulation, as stated by Ballas et al. (1991), is that such interfaces maintain a consistent style and may offset the disadvantages of changing controls and displays as part of AA implementations in complex systems. In terms of human performance, the consistency in the interface style may result in increased SA.

Ballas et al (1991) used a dual-task paradigm, including a target confirmation and classification task along with a tracking task, to investigate the effectiveness of direct manipulation interface design over conventional interface design for reducing AA-induced deficits in SA and performance. Deficits were expected as a result of drastic changes in display content as system automation changed from one mode to another. They speculated the direct manipulation interface would increase the users sense of task engagement; thereby, making the transition from one mode of system automation to another easier for operators. The target classification task was either fully automated or performed manually by a human depending upon the level of difficulty of the tracking task. Multiple modalities were used to cue changes in the state of the target classification task automation. The cues included a beep and a graphical border placed around the task window when manual control was required. The border served to highlight the task interface. Ballas et al. (1991) found that performance (or "SA") decrements occurred during manual system control due to advanced automation of the target classification task. They observed that the direct manipulation interface was superior to a conventional interface for ameliorating adverse performance consequences due to system state changes. The relative effectiveness of the beep or graphical border around the target task window for cueing subjects on system state changes was not assessed. However, based on subjective query of subjects, Ballas et al. observedthat the more intuitive direct manipulation interface allowed subjects to better anticipate state changes than the conventional interface.

Ballas et al. even though subjects were cued to the complex system state changes as part of AA, automation-induced SA deficits were still observed during periods of manual system control. This suggests that the type of cueing may not have been entirely effective for informing operators of state changes or the interface design did not allow subjects to easily infer how to modify control actions based on a system state change. In general, the study supports the need to examine explicit cueing of operators to complex system state changes to promote SA and performance subsequent to a state or behavior change (Kaber, Hughes, Wright, Sheiknainar, & Warren, in preparation).

Dynamic displays, multimodal displays, and direct manipulation displays are all proposed solutions to problems incurred as a result of AA; however, more empirical research is needed in order to determine which display type produces the best results under various implementations and strategies of AA.

1.2.2 HUMAN-HUMAN AND HUMAN-AUTOMATION INTERACTION

The issues in implementation of AA become more complex when considering team operations or control of complex systems. When AA is utilized in team scenarios, it may drastically alter the team environment (Kaber, Riley, Endsley, & Tan, 2001). Teams are created to exploit individual abilities of members. One team member may be stronger in a certain area of a task than another member. Team member interaction and communication is crucial for a system to function correctly (Kanki, 1993). In order for the team to be an effective entity, the members must be able to share their knowledge via an effective means of communication, be it vocal or non-vocal. Different situations are more conducive to different forms of communication. For example, in a loud factory, production team members would most likely not be able to hear one another speak, so they may use other forms of non-vocal communication, such as hand gestures or written language (Bowers, 1996). Since automation, in effect, may function like a member of the team, it also must be capable of communicating with other team members. It should also be designed not to compromise critical vocal and non-vocal communication among human operators. Scerbo (1999) suggested that AA should be designed so that multiple forms of communication are possible across multiple modalities allowing more natural communication to result and possibly reduced team member (human, computer controller) workload.

Sklar and Sarter (1999) previously hypothesized that auditory and haptic cues may promote improved interaction between humans and automation in coordinated complex system control; however, they did not investigate an adaptive system. Specifically, they examined the use of visual, tactile, and bi-modal (visual and tactile) cues to indicate unexpected mode transitions in a cockpit simulation. Sklar and Sarter (1999) found the use of tactile, and a combination of tactile and visual cues, to result in higher rates of detection of mode transitions. The rates of detection with tactile cues were particularly pronounced in trials that involved visually intensive displays. It is possible that human-human and human-automation team interaction in the cockpit may benefit from cueing across multiple modalities to facilitate team member comprehension of system states.

As reported earlier, Kaber et al. (in preparation) examined the use of earcons, icons, and a combination of both earcons and icons in an adaptive teleoperation system involving human and computer coordinated control. Auditory cueing produced significantly fewer errors than visual cueing. Time-to-task completion was also greater with visual cues. Kaber et al's (in preparation) study expands upon the previous research by Sklar and Sarter (1999) by providing further evidence of the benefits of multiple modality cueing under AA and improvements in human-automation interaction.

With respect to these results, the auditory modality may be especially beneficial for cueing information because it has the advantage of alerting people when a sound is heard, which can result in faster reaction times (Bly, 1982). Graham (1999) compared earcons to traditional warnings in an automobile to evaluate emergency warnings in vehicle collision avoidance applications. Graham (1999) found in the context of emergency warnings, "Auditory icons have considerable potential advantages over conventional sounds in terms of response time and subjective ratings." In this instance, automobile electronic systems act like a team member to the human operator. The context requires effective communication between the human and the automobile, so collisions do not occur. The automobile communicates by providing cues that a collision is about to occur. The human responds to the automobile's cues by re-directing the vehicle, towards collision avoidance.

In general, all of these studies provide evidence to support the use of multimodal cueing in human-human and human-automation coordinated (team) control of complex systems. Advanced warning of operators (team members) of system state changes with complex auditory, tactile, and bi-modal cues has been proven to be successful in various contexts. Such cueing mechanisms may serve to reduce the number of serious AA implementation issues in complex system design.

1.2.3 Summary of Adaptive Automation

Adaptive automation is likely to be an effective means for implementing automation and improving human performance in the near future; however, optimum methods had not been developed to employ AA over a broad range of disciplines (Kaber, Hughes, Wright, Sheiknainar, & Warren, in preparation). This research evaluated what cue types are most effective in reducing a lack of operator mode awareness problems associated with AA. Scerbo (1999) cited Weiner (1989), who stated, "Automation is neither inherently good nor bad. It does, however, change the nature of the work, and, in doing so, solves some problems while creating others." These problems need to be known, so designers can create ways to improve upon and solve them (Scerbo, 1999). Different interfaces are one proposed method of solving problems with AA. This research evaluated the effectiveness of an interface utilizing different sensory modalities (visual icons, vocal cues, and earcons) in an AA application.

1.3 HUMAN INFORMATION PROCESSING OF AUDITORY STIMULI

Mountford and Gaver (1990) said, "Sound exists in time and over space, vision exists in space and over time." Humans have omnidirectional hearing capability, but only have a small field of view; therefore, sending information to the auditory modality makes sense particularly for system-state warnings. Sound can be provided in a way so as not to distract the user from the visual display, but rather to complement it and to effectively use human cognitive resources.

According to Wickens and Hollands (2000), "Multiple-resource theory would predict relative independence between the perception of music and the involvement with tasks that are manual or vocal." Likewise, Martin et al. (1988) found that reading comprehension and instrumental music could be simultaneously processed without any decrement in understanding; however, reading comprehension did suffer when lyrics were added.

Multiple messages can be conveyed simultaneously by different parameters of auditory cues. King and Oldfield (1997) said broadband signals should be used, so the user can successfully hear all of the parameters of sounds. Albers (1996) found that when he shifted a portion of information to the auditory channel as opposed to the visual channel in an Audible Web application, users were able to gather more information than they did with a traditional web application. Albers (1996) shifted information concerning data transfer, feedback on user's actions and navigational aids through content feedback to the auditory modality because the visual modality was already busy gathering all of the information that is incurred in an internet application. In addition, the users could more accurately describe the status of the system as far as data transfer and activities.

When multiple channels of auditory sources are presented, usually the subject only focuses attention on one channel. However, channels of auditory information that are not paid focused attention may remain in the echoic short-term sensory store for 3-6 seconds (Wickens & Hollands, 2000). If the previously unattended information becomes attended to, then it may be recognized and the information may be sent to working memory and long-term memory (Wickens & Hollands, 2000). According to Banks et al. (1995), "Information presented in an unattended channel is temporarily inhibited for several seconds following presentation, demonstrating 'negative priming'." If the information first presented in the unattended channel is then presented in the attended channel, then there is usually a slower reaction time than if the information had originally been presented in the attended channel.

As humans age, degenerative effects decrease the ability to distinguish sounds, just like other types of stimuli (Barr & Giambra, 1990). Therefore, if auditory cues are designed for older users, care needs to be taken to create earcons that are significantly different from one another.

1.4 EARCONS

In general, the benefits of the auditory modality may be capitalized upon by the use of earcons. Earcons are messages that are created from auditory tones. The properties of earcons are generally pitch, amplitude, frequency, duration, tone, and special effects that may be delivered to musical tones. Earcons can be hierarchical, meaning that one earcon's information may build off of another earcon's meaning. Earcons are useful because they can be internationally understandable. In addition, earcons may prove to be beneficial for people who are blind or have low vision, users of displays that are already visually cluttered, and for people whose clothing inhibits their sight (Brewster, 1998).

Mynatt (1994) described basic rules for incorporating usability into earcons. First, the sound must be identifiable; that it is common enough that the user is familiar with it, but also distinguishable from other sounds. Second, the sound should map well to the action being portrayed. In a study previously conducted by Brewster (1998), many subjects reported attaching meaning to the earcons, so they could better remember them. Third, the parameters of the sound should be appropriate. That is the earcon should be constructed in the normal hearing range of humans, should not be excessively long, and should not create negative emotions upon sensation.

Blattner et al. (1989) described how motives, which are musical fragments consisting of a few notes, can be transformed into complex earcons. Motives form hierarchies by using repetition, variation, or contrast (Brewster, 1998). If repetition is used, a motive is repeated a specific number of times to denote the level at which the user is at within the hierarchy. When using earcons as a mechanism of communication, an important point to remember is not to annoy the user with a particular tune. According to Blattner et al. (1989), "hearing a simple tune ten or more times a day irritates users, and can cause audio fatigue."

With variation, motives can be altered in rhythm, pitch, timbre, register, or dynamics. If contrast is used, the motive changes in pitch, rhythmic content, or both. Brewster (1998) suggests that a change in pitch alone will not provide enough information to the user for them to recall the correct level within the hierarchy. The ideal number of pitches within a single motive is two to four (Gaver, 1986).

There are three basic types of complex earcons according to Blattner et al. (1989). They are combined, transformed, and inherited. Each type has specific properties that differentiate it from the other two. Combined earcons simply combine motives. When the motives are used in conjunction with one another, a message is generated. Transformed earcons alter the sequence of the motives. For example, a transformed earcon may involve a crescendo as the motives are played. Inherited earcons involve hierarchies. Each hierarchy is defined to have specific properties. As new motives are added to increase the hierarchy, new properties must also be added. Blattner et al. (1989) suggests that timbre be used as the distinguishing factor for the root earcon of the hierarchy because it is an important element of sounds.

1.4.1 EARCON DESIGN

There are a few general rules that Brewster suggests using when creating earcons. A specific family is composed of sounds from one instrument or voice alone. The same types of earcons share the same rhythm. Kerman (1980) proved that the majority of listeners respond more readily to rhythm than to any other property of music. If an earcon exists within the same family and type as another earcon, then the register can be altered to denote the difference (Brewster, 1998).

Brewster et al. (2001) later generalized his previous guidelines by saying; many different parameters must be considered, such as timbre, register, rhythm, pitch, duration, intensity, and spatial location. Families of earcons share one or more of the above parameters. Families can be further divided into subgroups by sharing additional features as the families of earcons move down the hierarchy (Brewster, Edwards & Wright, 2001).

Blattner et al. (1989) defined a slightly different method of creating clear, concise earcons. A family of earcons is assigned a single, distinct rhythm. The second tier of the hierarchy incorporates the family rhythm, but then adds a pitch sequence. It is suggested that the second tier be composed in a sine wave because it is a more basic form of sound. The third tier consists of a copy of the second tier; however a different timbre and a higher pitch are used. For the creation of further hierarchies, a change in the register or dynamics can be used as differentiating parameters.

When designing earcons, care should be taken in using synthesized sounds because the qualities of the actual instrument may or may not be realized within the synthesizer. It is also important to consider the pitch when selecting timbres because not all instruments are capable of all pitches. Rhythm and duration of tones within earcons allow users to distinguish and process cues. Similar rhythms can be easily confused, thus the wrong information could be conveyed to the user. In order to distinguish earcons with the same rhythm, it is suggested that a different number of notes be used within different sub-groups of earcons.

It is possible that a user can understand two earcons played simultaneously. An earcon can consist of up to six notes played within 1 second and still be considered an effective method for information transfer. According to Brewster et al. (2001), the first note of the earcon should be louder, and the last note should be longer. In order for an earcon to be successful, it must first grab the attention of the user. Intensity is the most common means for involving the user in earcon perception; however, rhythm, pitch, rapid onset and offset times, irregular harmonics, atonal sounds, or arrhythmic sounds can also be used to attract the user. Compound earcons need to allow at least a 0.1 second gap between different earcons, so the user will clearly understand the starting and ending point of the respective earcons (Brewster, Edwards & Wright, 2001).

For detecting dynamic stimuli, a combination of rhythm and pitch have been found to be effective (Blattner, Greenberg, & Sumikawa 1989). Gaver (1986) suggested using a change in register to delineate vertical location or direction; a change in dynamics to indicate horizontal direction; and a crescendo or decrescendo to relate direction to the user.

1.4.2 Earcon Training

Complex system operators can be trained to recognize the distinct features of specific earcons, and therefore understand their meaning (Brewster, 1998). Blattner et al. (1989) state, "If the basic rules of combination are understood, listeners can infer the meanings of earcons they haven't heard." Also, Blattner et al. (1989) said, "Untrained subjects quickly recognize musical timbres." Brewster (1998) found that musical training does not affect a subject's ability to respond to earcons. However, similar rhythms did cause problems for non-musicians. According to a previous study by Blattner et al. (1989), musical motives are better than flat tones for earcon learning. Dowling (1986) proved that performance involving earcons was increased if the sequence of notes was familiar to the user (i.e., familiar motives were used).

1.4.3 Comparisons of Earcons and Icons

Earcons and icons have similar composition and anticipated effects; however, there are differences that may cause one form of communication to be more desirable than the other (Blattner, Greenberg & Sumikawa, 1989). Deese and Grindley (1947) related melodies to visual shapes, stating they were similar in purpose and effect. Buxton et al. (1987) demonstrated the importance of auditory cues in human-computer interaction applications based on empirical findings that when the sound option was not selected in video games, scores decreased. Both icons and earcons convey information; however, only icons are selectable. Earcons are more transient than icons. Earcons presented simultaneously are more complex for a user to interpret than simultaneously presented icons. However, studies by both Elliot (1937) and Sticht (1969) have shown that vocal messages are more easily understood than textual messages.

Icons and earcons can be classified as representational, abstract, or semi-abstract (Blattner, Greenberg & Sumikawa, 1989). Representational icons are basic pictures of objects, whereas representational earcons are digitized sounds that mimic actual events. Abstract icons combine geometrical shapes and marks to invoke the desired effect on the user. Abstract earcons involve motives that do not directly relate to the meaning initially, but rather meaning is attached to them through training. Meaning may be in the form of vocal or visual and even spatial memory codes. Semi-abstract icons are constructed using a combination of geometrical shapes and marks, as well as pictures of objects. Semi-abstract icons can also be simplified pictures of objects, whereas semi-abstract earcons combine a naturally occurring sound with learned motives. Semi-abstract icons and earcons seem to be the more "internationally-friendly" choice. According to the American Institute of Graphic Arts (1981), most international signs are constructed using semi-abstract icons.

1.4.4 Comparison of Earcons and Vocal Information

Earcons are not language dependent, so unlike vocal cues, earcons are understood by people of all cultures, regardless of their spoken language (Tannen, 1998). Brewster (1998) contended that earcons also have an advantage over vocal cues in that they communicate information more quickly. It may take a longer period of time to actually say the words to convey the information necessary to understand a situation than if the meaning is understood by a single earcon. However, it is possible that the human may be capable of understanding the meaning of the vocal information more quickly than the information presented in the earcon because earcons require an additional translational stage within human information processing. According to Rabiner et al. (1997), "Speech is the most intuitive and most natural communicative modality for most of the user population." Beyond this comparison, only one vocal message can be understood at any given time; however, multiple earcons messages can be understood simultaneously. Buxton (1995) said, "In cases where we need to control more than one process simultaneously, speech (alone) is generally not an effective mode of communication."

1.4.5 SUMMARY ON EARCONS

Earcons have proved to be an effective feedback mechanism compared to visual cues; however, current studies have failed to compare the effectiveness of earcons as a feedback with vocal messages. Multiple experiments have proven earcons can be successful in delivering complex messages to users in cognitively taxing situations. For instance, in telephone-based interfaces, Brewster (1998) used earcons for navigation within a menu hierarchy. He found earcons could be recalled well over time. He proved that using earcons is a viable method of navigating through menu hierarchies. When analyzing non-visual interfaces for wearable computers, Walker and Brewster (2000) found, "Workload was significantly reduced when sound was present…users do not have to devote so many of their cognitive resources to perform the task." Earcons may be successful because they reallocate information that would most likely be displayed in the visual modality to the auditory modality. Multiple resource theory provides justification for the division of information into different modalities for maximum simultaneous human information processing. Earcons that are trained and have meanings associated with them may also allow for more efficient and effective information processing than vocal messages.

1.5 Virtual Reality Systems for Human Factors Research

Unfortunately, the majority of historical AA studies presented in the literature have been conducted using low-fidelity laboratory simulations of flight tasks (e.g., the Multi-Attribute Task Battery (Parasuraman et al., 1996; Parasuraman, 1993; Scallen, Hancock & Duley, 1995)). Only recent AA research has developed and employed high-fidelity synthetic representations of, for example, air traffic control (ATC) tasks (Hilburn, Jorna, Byrne, & Parasuraman, 1997), aircraft piloting tasks (Scallen & Hancock, 2001), and teleoperation tasks (Kaber, Hughes, Wright, Sheiknainar, & Warren, in preparation). In some of this work, VR systems have been used to present simulations to promote task realism and generalizability of results to real-world applications. The work has provided important insight into implementation issues associated with adaptive automation.

This research followed this current trend and made use of a VR system to simulate a complex system and to assess the effects of AA applied to that system on human performance. With this in mind, this section presents a detailed review of VR technology as a basis for equipment selection for the experiment as part of this study.

Aukstakalnis et al. (1992) said, "Virtual reality is a way for humans to visualize, manipulate and interact with computers and extremely complex data." Virtual reality allows users to have an otherwise two-dimensional scene represented in three-dimensions. The addition of depth to the field creates an environment that is a closer representation of reality. Virtual reality may be successful because it provides a more accurate and easier mode of decision analysis than is currently available through traditional 2-D simulation (Kaber, 1999).

There are many benefits for using VR for simulation and training as opposed to more conventional methods. Virtual Reality is relatively inexpensive in comparison to the cost of actual field tests. Virtual reality is also capable of producing an extremely high-fidelity representation of reality. Another benefit is VR serves as an excellent medium for training in task-critical situations because real resources are not lost. In addition, VR can be used when there are limited available resources.

1.5.1 Virtual Reality Equipment

Resolution, color, freedom of movement, weight, number of views, cost, speed, and contrast are all factors that go into the selection of equipment for VR applications. Virtual reality equipment can broadly be divided into two categories, including displays and controls. Displays include items like head-mounted displays (HMDs), active and passive glasses, binocular omni-orientation monitor (BOOM) displays, and autostereoscopic displays. Controls include items such as mice, joysticks, wands, data gloves and exoskeleton master controllers (Youngblut et al., 1996). Within each of these categories, there are different brands and models. Since VR is a relatively new concept, the technology and equipment available is constantly evolving. Some existing technology is discussed below.

1.5.1.1 DISPLAY TECHNOLOGY

Head-mounted displays are primarily used because they allow users a relatively high degree-of-freedom of movement. Depending on the model, HMDs can also offer high resolution (1280 x 1024) and high contrast (100:1). However, HMDs limit the field-of-view of the user, and are often cumbersome.

Certain HMDs are better suited to specific applications. For example, Liquid Image Corporation claims that the MRG 4 is one of the best selling HMDs for gaming. Virtual Reality, Inc., offers the VRI HMD 133 for surgical applications. Kaiser Electro-Optics, Inc. offers the Vision Immersion Module, which can be used by multiple users, allowing more than one person to be immersed in the same environment simultaneously. Head-mounted displays, in general, are beneficial for decision analysis because they allow for multiple viewpoints of the same data. Youngblut et al., (1996) said that they might also focus user attention and promote performance as they isolate the user in the virtual environment, so they have an increased sense of presence.

Active glasses work on the principle of switching the image seen by the viewer from one lens of a pair of lightweight glasses to the other (Belleman et al., 2001). When one lens is working, the other is blocked, and vice-versa. The switching “on” and “off” of the lenses, or flicker rate, is so fast that the human eye cannot readily detect the blocked lens, causing the image to appear both seamless and three-dimensional (Youngblut et al., 1996). Two slightly different views are shown to each eye. An infrared beam controls switching of the lens and allows multiple users to use multiple glasses simultaneously with a single beam emitter (Belleman et al., 2001). Therefore, this technology is well suited to large audiences (Youngblut et al., 1996). Another advantage of active glasses is they are much less cumbersome than Head-mounted displays. However, like HMDs, they only provide a narrow field-of-view for the users.

Passive glasses are similar to active glasses in that they are well suited for large audiences, but they are also used commonly in CAVE systems. Passive glasses operate by creating a checkerboard image in each of the alternately polarized lenses. Half of the pixels resulting are then disregarded and the remaining pixels are used to form an image for the viewer. The flicker rate in passive glasses is lower than in active glasses, and the resolution is reduced as a result of the discounted pixels (Youngblut et al., 1996).

Fakespace, Inc. makes the BOOM 3C, which is a visual display that is mounted on a boom to alleviate viewer discomfort associated with usual HMD weight. A cathode ray tube (CRT) display system equipped with an optical tracker is then used for each eye. The boom is equipped with a six degree-of-freedom motion tracker (Youngblut et al., 1996). The motion tracker assigns floating point angles to specific regions in space. The data is then transferred to the workstation, which in turn simulates the image to be projected in the visual display (Belleman et al., 2001). The main advantage of the BOOM is that it is not cumbersome like an HMD.

1.5.1.2 CONTROL TECHNOLOGY

Mice are often used in VR applications because they are familiar to most users. Humans typically interact with conventional PCs using a mouse, so it is logical that they have been used to facilitate human-virtual environment interaction. Joysticks are used with VR for this same reason. Wands are another type of controller that may be used with VR systems. Wands have the advantage of being slightly more abstract, so they can be used to represent different objects of control within the VR simulation. However mice, joysticks, and wands all may not be the best types of control for VR applications because they tend to steal user attention away from virtual scenes (Davies, 1996). Another downfall for these controllers is users may need to see the controller in order to operate it; however, some of the display technologies, such as HMDs, do not allow users to see anything outside of the virtual environment.

An alternative to joysticks and wands are data gloves and exoskeleton master controllers. These devices fit onto your hands and body. Depending on the particular glove or exoskeleton control, there are varying numbers of flex sensors. Flex sensors relay to the VE exactly how a person moves at a particular joint, therefore, increasing the intuitiveness and ease of control as well as the realism of the VE experience, as if the person were actually in it (Gabbard & Hix, 1997).

A suggestion of Gabbard and Hix (1997) is to incorporate VE input devices into real-world objects that would be used if the task were to actually occur. An example they give is the use of a device incorporated into a fire hose to train firefighters. This was initially used in a VE for the Naval Research Laboratory, but similar technology can now be found in common video games (e.g.), suggesting the fast pace of development of VE technology (Gabbard & Hix, 1997).

Gabbard and Hix (1997) also describe speech recognition systems as a means of input in VR applications. Incorporation of speech into a VR system may increase the reality of the system to the user. Also, it provides a method of selecting objects that does not divide the user's attention by forcing him or her to search for input devices or control panels. However, voice recognition technology may not yet be refined enough to use in complex VE systems. Voice input is likely to be a viable input mechanism in the future, but as of now, it is not.

A specific task may involve fewer movements than another task. For that reason, not all VE interfaces need to incorporate six degree of freedom (DOF) motion (Hinckley et al., 1994). More DOF translates into more information that must be processed, therefore, occupying a more significant portion of the bandwidth and possibly introducing or extending lag in the system. Lag creates a sense of disjointedness between the user's movements with the input device and the results within the VE, in fact Richard et al. (1996) state, "delays of more than 300 milliseconds may decrease user presence and immersion".

Although these technologies may facilitate innovative and exciting applications, researchers are striving to improve many facets of them; so future work can enable more realistic displays. Current VR technology research focuses primarily of new 3-D displays and how these devices can be more effectively designed from a human factors perspective. ATR Communications Systems Research Laboratories is working on correcting the depth error that can result from an incorrect inter-pupillary distance (IPD) and fuzziness of a display, which could drastically improve clarity in HMDs (Youngblut et al., 1996). Dimension Media Associates is manufacturing a High Definition Volumetric Display (HDVD) intended to provide an autostereoscopic 3-D display using concave mirrors and specific frequencies to adjust the index of refraction, which diffracts the beam of light. Anticipated applications of HDVD include VE, gaming, surgical applications, and general public information platforms. Other researchers are studying improvements in virtual depth perception, autostereoscopic visual advancements, CAVE, and higher resolution displays (Youngblut et al., 1996).

1.5.2 VIRTUAL REALITY USER CONSIDERATIONS

In addition to recognizing the limitations of certain VR hardware, one must recognize the constraints of the VR user in order to design effective applications. Gabbard and Hix (1997) developed a taxonomy that lists certain features that they believe should be considered when designing a virtual environment. The first item on their list emphasizes that the user is stressed. Stress can result in decreased attention spans, reduced situation awareness, tunnel vision, and a variety of other cognitive problems.

In general, the taxonomy supports the notion that the technology should be designed around user's characteristics, rather than the user adapting to already existing technology. Individual differences stem from user experience levels both in the field and with similar technology (Egan, 1988). User capabilities will likely play a large role in the overall effectiveness of a VE, such as the user's vision, hearing, and size. The system should be designed so that it will be viable for both left and right-handed people. This could be done either by creating the controls and interfaces so that they are symmetrical, or can be switched to the users preference (Gabbard & Hix, 1997). The size of users impacts not only the equipment, but also the role that the users see themselves playing within the VE. One suggestion made by Boeing, was to incorporate different virtual body sizes into the VE, so that the users could wear the size body that would be most beneficial to their task (Gabbard & Hix, 1997).

Object selection is another category in Gabbard and Hix's taxonomy, and it highlights topics such as feedback, frame rates, and selection of multiple objects in Virtual reality. One of the basic points that Gabbard and Hix (1997) make is that users need to be aware that whatever they desire to select is actually selected. Various methods of feedback were listed, such as "highlighting, outlining, or vocal confirmation" (Gabbard & Hix, 1997).

Acoustic feedback is another division in the Gabbard and Hix's taxonomy. From a presence perspective, VEs should make users feel like they are actually there. In order for this to occur, a VE should sound, as well as look, real (Gabbard & Hix, 1997). Virtual environments should capitalize on user's capabilities to perceive audio signals. If a situation is difficult to judge based on appearance alone, the sounds accompanying the event may help a user understand the Virtual environment. For sound presentation, Gabbard and Hix (1997) suggest using speakers over headsets, particularly if multiple users are involved. This recommendation stemmed from the fact that if headphones were used, multiple users would not be able to communicate in a natural manner. Kinesthetic, tactile, and olfactory cues could also be used to make the VE feel more real to a user. However, in adding all of these sensory cues, it is important not to overwhelm the user, so their attention is not drawn away from the virtual task at hand. With this in mind, one would suggest that the sensory cues only be placed in areas that appear to be problematic. Extra sensory data in areas away from the problems would only distract the user. Since one of the purposes of this research is to evaluate multimodal cueing, different sensory mechanisms utilizing a combination of auditory and visual stimuli will be used.

The ability to select multiple objects at once is a feature that can be used through "rubber band-like" grouping mechanisms (Mapes & Moshell, 1995). The frame rate needs to be high enough that it allows the user to have a seamless display, but not so fast that it disorients the user (Gabbard & Hix, 1997). Ware and Balakrishnan (1994) suggested that certain tasks such as target acquisition are not heavily affected by low frame rates. This suggestion once again stresses the importance of developing a VE specific to the individual needs of the project.

Another important issue that Gabbard and Hix (1997) discuss is the number of users and their mode of evaluating their location in a Virtual environment. The virtual task will normally determine the number and location of the users, but design specifications must be made, so that these characteristics are integrated into the Virtual environment. If the VE focused on only one task, then typically a single-user VE is designed; however, if there is more than one task in question, designs should include a method for interaction between multiple users. Designing for multiple users allows all users to be aware of not only their presence, but that of the other users, as well. Typically, VE equipment is designed with the single user in mind; however, there are many situations that would benefit from multiple user capability. The Responsive Workbench, or similar technology is suggested for multiple-user environments. With respect to user interaction in using the workbench, if it is important for the operators' positions within a multiple-user VE to be known, then it is essential to have some representation of each individual within the overall environment, in order for the users to see their position relative to the other operators.

The perspective of the user in the VE is also an important consideration in application design. In VE's, the perspective can be anything, ranging from that of a human walking through an environment to that of a bird flying. The perspective is critical because it aids in the user's sense of presence, and allows the user a mechanism to investigate their surroundings. Gabbard and Hix (1997) suggest using an egocentric perspective to accentuate presence and an exocentric perspective to obtain a more "detailed relative position information". An aid to users would be an option that would allow them to switch from an egocentric to an exocentric perspective. The task to be used in this research provided operators with the capability to select among different perspectives during task performance.

Another dilemma that must be combated in VE design is field of view versus pixel allocation. If a larger field of view is used, then lower resolution may be necessary in order to maintain high frame rates. Likewise, if a smaller field of view is used, a higher resolution may be possible. Once again, this is related back to the problem of bandwidth (Yoshida, Rolland, & Reif, 1995).

1.5.3 SUMMARY OF VIRTUAL REALITY

Based on the review of literature, this experiment used a Virtual Research V8™ HMD integrated with a Silicon Graphics Zx10™ workstation to present to subjects the underwater virtual demining simulation used by Kaber et al. (in preparation). The Virtual Research V8™ HMD isolated user vision to only the VE. Test trials were short in duration and the weight of the HMD was not expected to be a factor in user performance. In addition, subjects used vocal commands as well as a mouse to control the simulation, which do not require visual attention or that the subject remove the helmet to perform interface control actions. The SGI was chosen because it provides the necessary computational power to run the simulation at relatively high frame rates.

Virtual reality is a highly effective tool that can be successfully used for a variety of applications, including experimental simulations for research purposes. It is also important to recognize that not all tasks are well suited for virtual reality (Stanney, 1995). For those tasks that could be improved upon with the use of VR, the VE should be clearly presented, relevant, and consistent. It should be clear, so the users understand exactly what they are seeing, and are required to accomplish. It should be relevant so the VR is actually useful for accomplishing task goals. And, it should be consistent so users know exactly where to find specific information (Stanney, 1995). The perspective should be adjustable, so egocentric and exocentric views are possible. Common sensory cues should tell the user exactly what problems exist. These and the other user considerations and design principles reviewed above have been considered in the design of the VR system and task for the current AA research. The task and experimental equipment are detailed in the Methods section.

2. PROBLEM STATEMENT

Multimodal cueing on complex system states has been demonstrated to have a positive effect on human operator performance; however, the most effective method of cueing has not been identified for adaptively automated systems, like contemporary teleoperators. There are many different mechanisms that can be used for cueing operators of impending changes in automated system states. The cueing mechanisms examined in this research include earcons, icons, and vocal cues.

The cueing mechanisms that involve sound, earcons and vocal cues, may have the advantage of capitalizing on MRT. Auditory cues may be more effective in visually intense tasks because they may tax different attentional resources than already present visual stimuli. In addition, humans have omnidirectional hearing capability, but only have a small field of view; therefore, transmitting information to the auditory sense may be more appropriate for system state warnings.

Earcons, or messages constructed out of auditory tones, can relay messages quickly through the use of different parameters of sound. Whereas earcons transform sound into a message, vocal cues present a message in a language native to the user. It is possible that vocal cues may be a more effective mechanism for presenting messages in simple systems because no transformation is required to perceive a message. However, complex systems may more effectively incorporate earcons because their messages can be hierarchal in nature. Hierarchal earcons are capable of encoding complex messages in a short duration because the user is able to learn the meaning of messages based on specific musical properties (Blattner, Greenberg, & Sumikawa, 1989). Users then may be able to expand their understanding of novel earcons based on previous messages, when a parameter of the earcon is altered (Brewster, 1998).

Earcons and icons have similar composition and anticipated effects; however, there are minute differences that may cause one form of communication to be more desirable than the other. Both icons and earcons convey information; however, only icons are selectable. Earcons are more transient than icons. Simultaneously presented earcons are more complex for the user to interpret than simultaneously presented icons (Brewster, 1998). In situations where there is little visual clutter, it is possible that icons may be the most effective mechanism for cueing. However, in situations where visual clutter is high, earcons or vocal messages may be the most effective cueing mechanism because they involve the auditory modality, rather than depending solely on the visual modality.

In addition to the modality of presentation, the level of cue complexity may also play a role in the overall effectiveness of the cue on system state changes. The cue that produces the best results during performance in low complexity condition may not be the cue that is most successful in a high complexity condition. For low complexity conditions, earcons may prove to be the most successful because there is not a great deal of information that must be learned by the user. However, in high complexity trials involving earcons, users must learn many more motives, etc.; therefore, vocal cues may be more successful in these trials. Given the need to translate their meanings, the earcons may entail an additional step in human information processing beyond the "steps" required to comprehend vocal cues. This study aids in not only characterizing the most effective cue type, but also characterizes the most effective cue type for various levels of cueing complexity.

3. METHODS

3.1 TASK

The virtual demining task involved three general phases including: (1) searching for mines with the virtual telerover; (2) placing depth charges on the surface of a disposable mine; and (3) detonating the charges in order to destroy the mine. The task was conducted using a virtual teleoperated rover previously used in Kaber et al.'s study (2001) (see *Figure 1*). Using vocal commands, a subject manipulated the rover's arm and wrist to collect a charge from a charge bin mounted on the front of the rover using a magnetic gripper tool. The magnetic gripper was also used to place charges on mines. The subject then used vocal commands to manipulate the position of a sonar dish mounted on the robotic arm in order to aim it at the charge and to detonate the charge with a direct signal. The rover also featured a tool display that showed which tool was active (either the magnet or sonar).

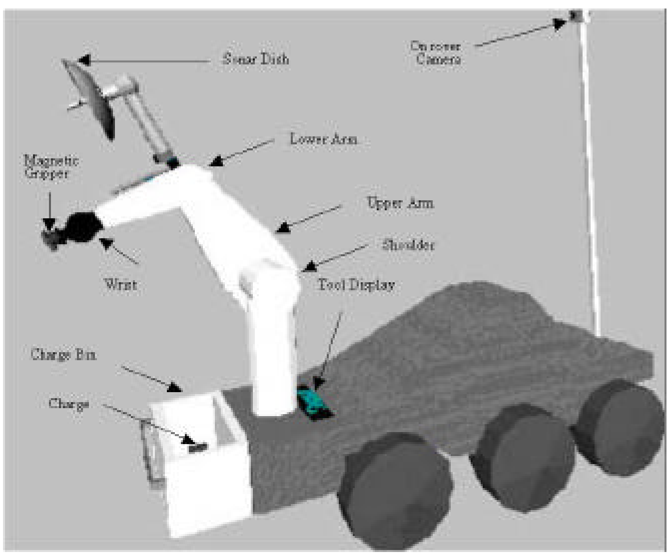

Figure 1. Diagram of rover (Kaber et al., 2001).

There were multiple steps specific to each of the three task phases. In the search phase, the user had to first locate a mine, then, drive to the mine that they previously located. The steps of the place charge phase included removing the charge from the bin and placing the removed charge on the mine. In order to accomplish these tasks, the user must manipulate the telerover's arm to remove the charge, drive to the mine, and then, manipulate the arm again to place the charge on the mine. In the detonate phase of the task, the user had to move back from the mine, position the sonar by manipulating the robotic arm's position, and then fire the sonar at the charge using a mouse click. The user's goal was to successfully detonate two mines within each trial. Images of the VE simulation at the different stages of the demining task are presented in *Figures 2-6.*

There were four different types of mines within the virtual demining simulation. Two types of charges were available inside the bin located on the front of the telerover. Once the subject approached a mine they wished to detonate, they then had to select the corresponding charge from the bin. Each charge is specific to only two mine types, so the correct charge must be selected for the mine to be destroyed. The particular charge must then be placed on the mine. *Figure 7* delineates the correct mine and charge association, a square charge for the two types on the left side and a cylindrical charge for the other two. If the wrong charge is picked, the charge can still be placed on the mine, and the detonation procedure will proceed; however, only the charge will explode. The mine would be left floating in the environment after the explosion. The use of an incorrect charge was counted as an error in user task performance, and was recorded in log files associated with the test trials. This was also the case for other response measures, which will be discussed in detail later in this section.

In the underwater demining task, a Wizard of Oz technique was used in which subjects vocally commanded the experimenter to select a specific tool, either the magnetic gripper or the sonar. The vocal commands were spoken in natural language. Confirmation feedback was provided through a small tool status display on the rover within the virtual environment. The display presents the name of the tool that is active. Instantaneous feedback was provided for vocal commands for rover arm and sonar movements when the specified segment of the manipulator moves within the Virtual environment. Subjects used a mouse controller to direct rover navigation, select viewpoints, and activate the various tools. Feedback was provided on viewpoint selection and rover movement by displayed changes within the virtual environment.

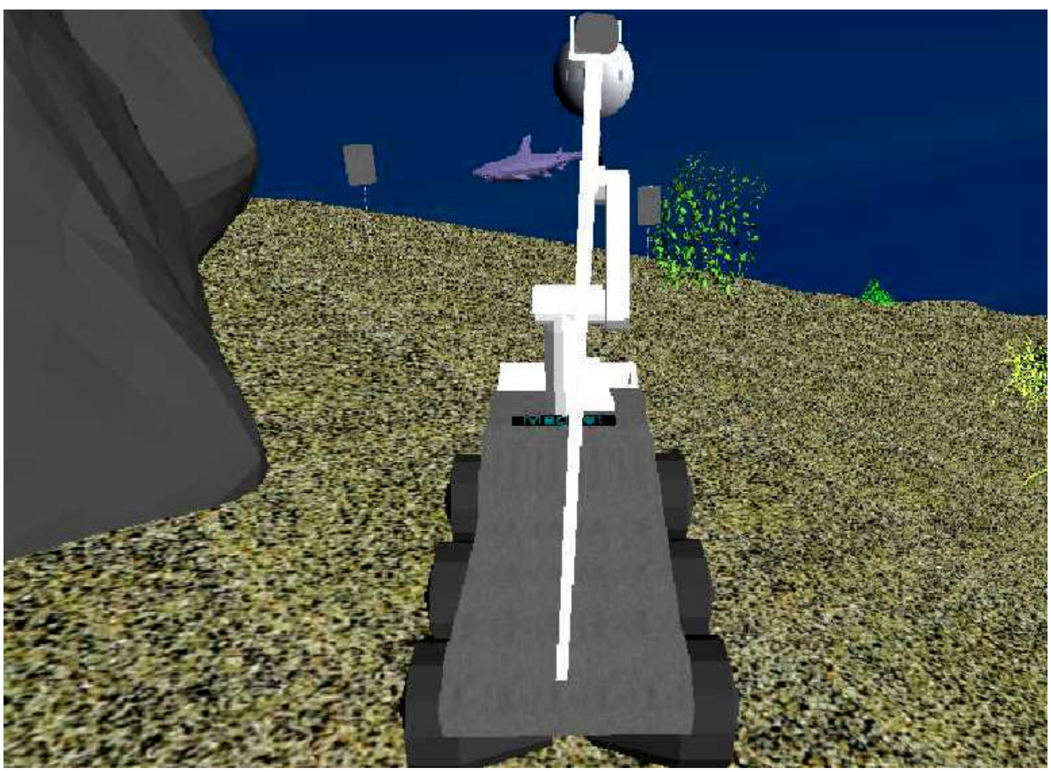

Figure 2. Search phase: Behind rover view.

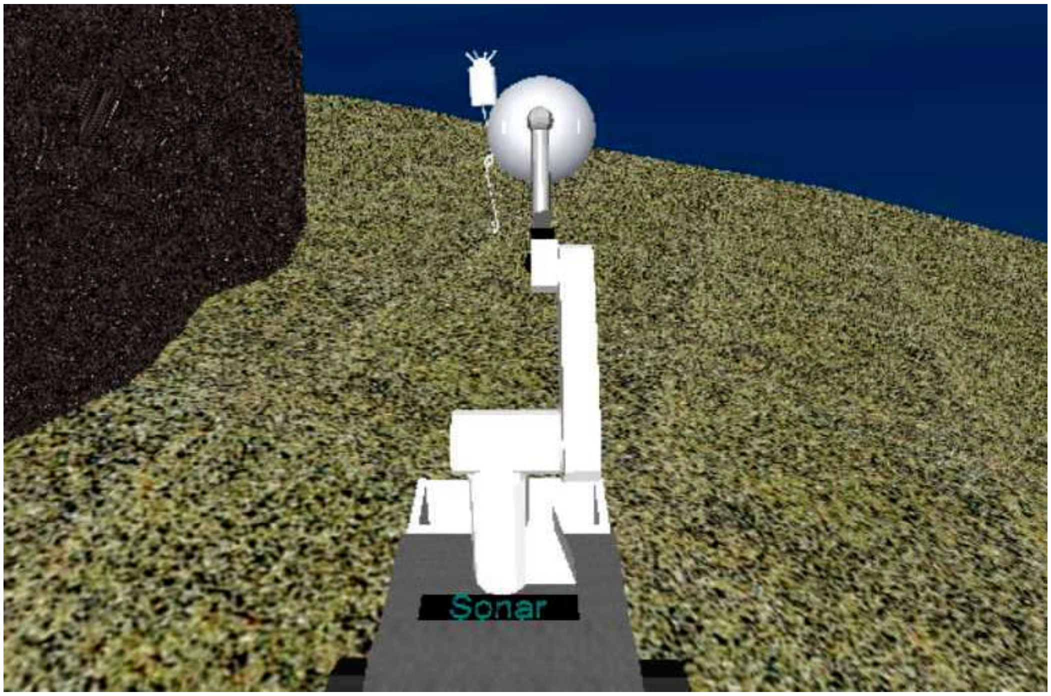

Figure 3. Search phase: On-rover view.

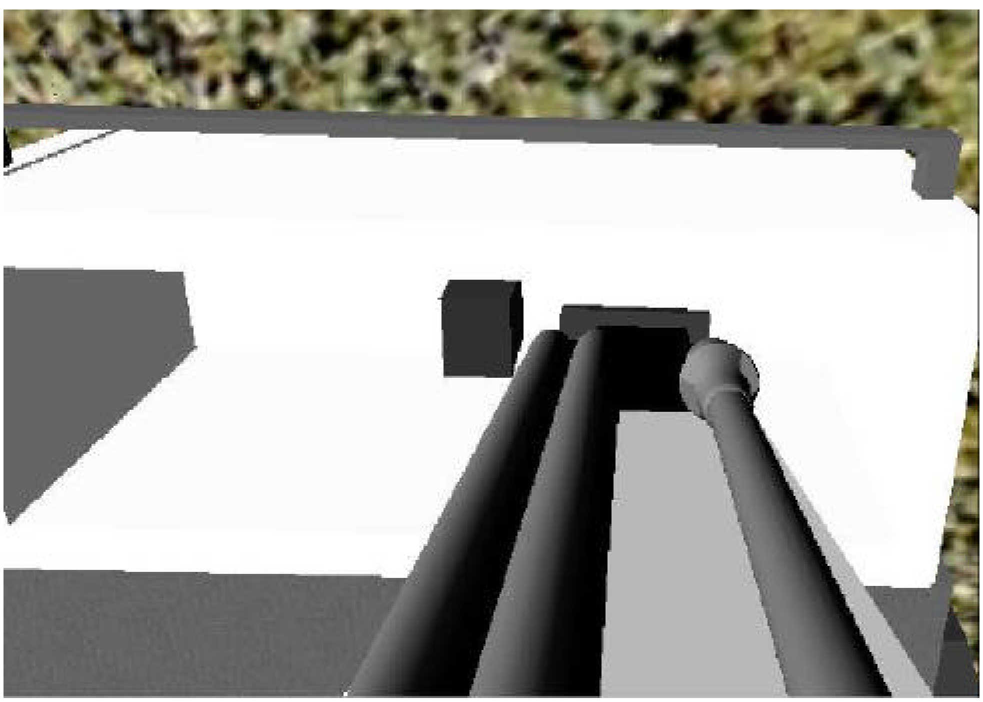

Figure 4. Selecting a charge from the bin on the rover: Rover arm view.

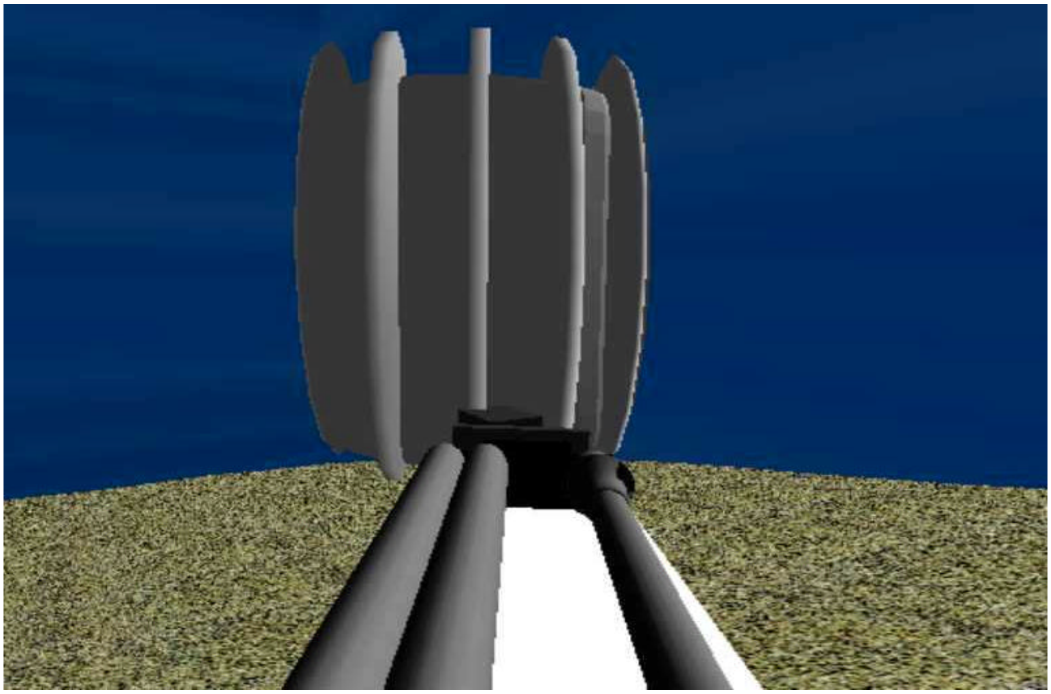

Figure 5. Place charge phase: Rover arm view.

Figure 6: Detonate phase: On-rover view.

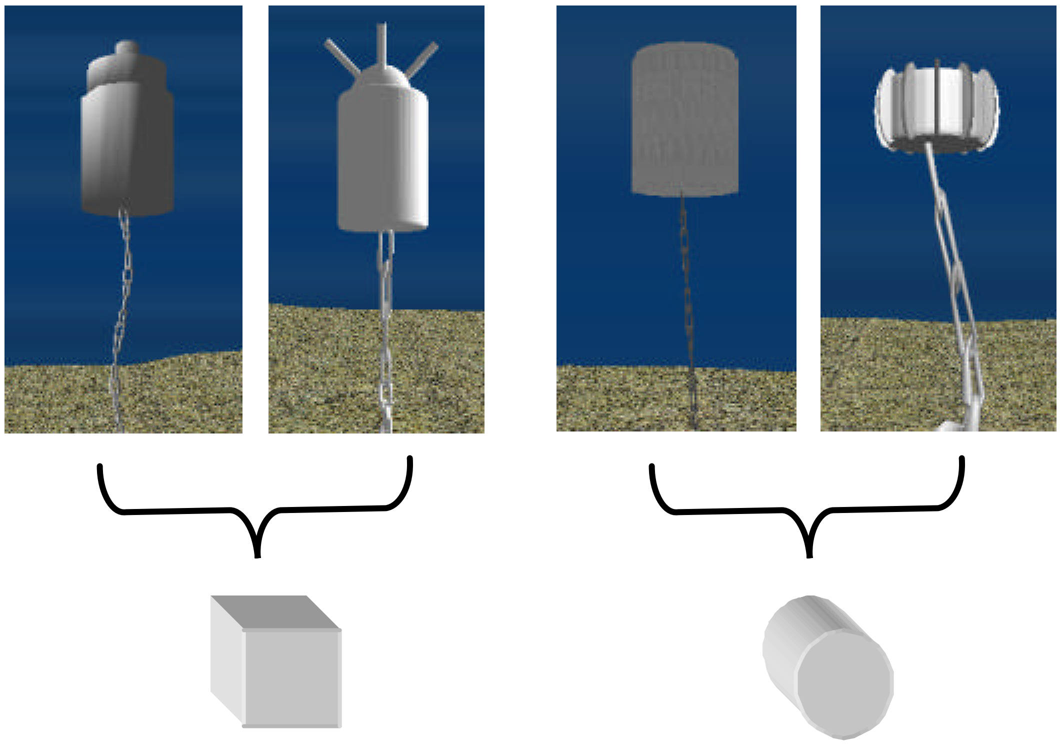

Figure 7. Mine types and associated charges.

3.2 Subjects

There were 32 subjects used in this experiment. The number of subjects was based on the number of participants in Kaber et al.'s study (in preparation). That experiment also used 32 subjects under similar cueing conditions and had highly significant results. The task used in both experiments was identical.

Undergraduate and graduate students were recruited from the NC State University population. Each subject was required to have 20/20 or corrected to normal visual acuity and PC experience (using Microsoft Windows). Subjects ranged in age from 18-30 years.

Subjects were randomly assigned to one of four groups of eight persons corresponding to the different types of cues to be provided on teleoperation task phase changes and system automation state changes (to be described later). At the onset of their participation, subjects were required to read and sign an informed consent form that summarized the task, risks associated with it, and responsible parties, as seen in *Appendix 2*. Subjects were also required to fill out a form detailing some personal characteristics such as VR experience, PC experience, and video game experience (rated on a scale from 1 ("none") to 5 ("frequent"))(see *Appendix 3*). On average, subjects highly rated PC experience (mean = 4.78), moderate ratings were recorded for video game experience (mean = 3.59), and VR experience was rated low (mean = 1.59). All subjects were monetarily compensated for their participation in the experiment (see *Appendix 4*).

3.3 EQUIPMENT

This experiment used a Virtual Research V8™ HMD attached to a Silicon Graphics (SGI) Zx10™ workstation to present the underwater VE to the subjects. The resolution provided to each eye by the HMD was 680x480. Dual Pentium III 1.5 GHz processors power the SGI. The workstation included 512 MB of RAM and used a 3D Labs Wildcat 4210 graphics subsystem. Sennheiser stereo headphones were integrated in the HMD and were used to present auditory cues to the subjects. Since the underwater demining simulation task only included one user, stereo headphones were an effective mechanism to isolate the user's attention to the virtual environment. A standard two-button mouse was used to alter the viewpoint, pick-up a charge, place the charge, and fire the sonar. All required interactions with the mouse were performed by simple, gross movements and mouse clicks.

A standard 101-key keyboard was used by the experimenter to control rover and sonar arm movements (vocal commands were given for movements by subjects).

3.4 EXPERIMENTAL DESIGN

The experiment followed a 4×2 mixed design with four different cue types (earcon, icon, vocal, or no cue) and two different cueing conditions (simple or complex). Each subject assigned to the earcon, icon or vocal cue condition performed one simple and one complex trial using their assigned cue type. The no cue condition served as a control group. Each subject in this group experienced two trials with no cues whatsoever.

Table 1 details the distribution of trials across subjects. In the first trial as part of the feedback conditions, subjects received simple cues that described only the phase of the task and the mode of system automation. The second trial supplied more elaborate cues, which delineated not only the phase and mode of automation of the task, but also the current step of the task. The cue types, the cue category, the phase, mode or step that the cue signaled, and the timing of the cues are detailed in *Table 2*.

Table 1: Distribution of trials across subjects.

	Feedback Conditions	**Vocal**	**Earcon**	**Icon**
Trial 1	Simple Cueing (phase & mode)	Subjects 1-8	Subjects 9-16	Subjects 17-24
Trial 2	Complex Cueing (phase, mode & steps)	Subjects 1-8	Subjects 9-16	Subjects 17-24

	No Cue
Trial 1	Subjects 25-32
Trial 2	Subjects 25-32

The earcons were constructed so that families used the same timbre. There were four families of earcons within the simulation. They were: control mode (manual or supervisory) and three for the phases of the task (search, place charge, and detonate). Since rhythm is an important aspect of earcons, the number of beats was the next parameter that was altered in the earcons. In the phase families, the phase name consisted of three main beats. Each step following the phase within the same phase had one additional beat. For example, "search" has three main beats, "locate mine" has four main beats, and "drive to mine" has five main beats. The same pattern is followed for all three phase families. *Table 3* illustrates the use of particular timbres and the number of beats used for each cue.

Table 2: Cues.

Cue Type	Representing	Cueing	Occurs
Simple	Phases of Task	Search	At start
		Place Charge	Before phase change
		Detonate	Before phase change
	Mode of Control	Manual	Before control shift
		Supervisory	Before control shift
Complex (in addition to simple)	Steps of Search	Locate a mine	At start
		Drive to a mine	When subject moves beyond the center quadrant
	Steps of Place Charge	Pick up charge from bin	Start of Place Charge
		Place charge on mine	When charge is picked up
	Steps of Detonate	Move back from mine	When charge is placed on mine
		Position sonar	When cursor returns to center of screen
		Fire sonar at charge	When sonar arm is moved

Table 3: Earcon composition.

Cue Type	Representing	Cueing	Timbre	Number of Beats
Simple	Phases of Task	Search	Dulcimer	3
		Place Charge	Organ	3
		Detonate	Sitar	3
	Mode of Control	Manual	Rain	N/A
		Supervisory	Rain	N/A
Complex (in addition to simple)	Steps of Search	Locate a mine	Dulcimer	4
		Drive to a mine	Dulcimer	5
	Steps of Place Charge	Pick up charge from bin	Organ	4
		Place charge on mine	Organ	5
	Steps of Detonate	Move back from mine	Sitar	4
		Position sonar	Sitar	5
		Fire sonar at charge	Sitar	6

Adaptive automation was facilitated in the present study through dynamic allocations of manual and supervisory control over telerover functions during the experimental trials. The allocations occurred based on a predetermined schedule and were evenly distributed across the subject population. In the supervisory control mode of the simulation, the rover drove itself to a mine, as well as move and activate the various tools to perform specific functions, such as removing a charge from the bin using the magnetic gripper, placing the charge on the mine using the gripper, and detonating the charge using the sonar. The user acted as a monitor while the computer performed all control functions.

Opposite to this, in the full-manual mode, the user was responsible for all system functions. The user drove the rover using simple mouse movements. Movement and activation of tools and viewpoints, as well as arm movements, were controlled by voice commands given by the user.

All subjects were tested under AA with a control sequence involving manual search, supervisory control of the place charge phase, and manual control of the detonation phase for the first mine of the first trial. The second mine involved supervisory control of the search phase, manual control in the place charge, and supervisory control of the detonate phase. The sequence was then reversed in the subject's second trial.

3.5 INDEPENDENT VARIABLES

The independent variables manipulated in this experiment included cue type and cue complexity. The cue type (earcon, icon, vocal, or no cue) was varied between trials. The cue complexity (simple or complex) was varied within subjects.

3.6 DEPENDENT VARIABLES

The dependent variables observed in the experiment included time-to-mine completion, time-to-task completion, task errors, workload, and situation awareness. Task times were calculated using the SGI system clock and were recorded in the log files associated with trials. Task errors (e.g., selecting the wrong charge to detonate a mine) were also recorded in the SGI log files.

Operator workload was measured using the NASA-Task Load Index (TLX) (Hart & Staveland, 1988). The TLX is defined in terms of mental demand, physical demand, temporal demand, performance, frustration, and effort. Subjects were provided with a sheet of demand factor descriptions. They then completed a subjective comparison form that forced them to choose between two demand factors as to which one they believed more greatly affects performance in the experimental task. Following each trial, subjects completed a subjective rating of perceived workload form, on which they were required to draw a vertical line on a linear scale for each of the demand factors (there are six linear scales in total) (See *Appendix 7* for demand ranking and rating forms). In order to obtain a composite workload score for each subject, the ratings (the actual measurements on the linear scales for the specific factors) were multiplied by weighting factors calculated based on demand component rankings.

Situation awareness was evaluated using an adaptation of the 3-dimensional Situation Awareness Rating Technique (SART). SART is the most widely tested subjective technique (Taylor, 1990). Subjects were given a form consisting of three linear scales, each 100 mm long (See *Appendix 6*). The 0 mm point marked the low end of the scale, and 100 mm point marked the high end (Jones, 2000). The scales were used to measure demand on attentional resources (SART D), supply of attentional resources (SART S), and understanding (SART U). SART D considers factors related to "complexity, variability, and instability" (Taylor, 1995, Taylor & Selcon, 1991). SART S factors in "arousal, concentration, division of attention, and spare mental capacity" (Taylor, 1995, Taylor & Selcon, 1991). Whereas, SART U takes into account all prior knowledge of the system, and information gained from feedback and interaction with people or systems. Several studies have proved SART to be an effective mechanism of evaluating SA in that SART provides enough sensitivity to generate effective diagnosticity (Jones, 1995, Taylor & Selcon, 1991). This information is then combined to create an individual's understanding of the system. The subject's overall calculated SART is obtained by the formula seen in *Equation 1* (Taylor, 1995).

$$\text{SA (calculated)} = \text{Understanding} - (\text{Demand} - \text{Supply}) \text{ or} \quad (1)$$
$$\text{Overall SA} = \text{SART U} - (\text{SART D} - \text{SART S})$$

3.7 PROCEDURES

The research was conducted to determine which method of cueing, vocal, echoic, or iconic, is most beneficial to human performance, workload, and SA. Performance in cueing trials was compared to trials in which no cueing method was used in order to provide additional evidence on the effectiveness of sensory cueing for abating AA-induced performance problems.

Prior to experimental trials, subjects had to meet a criterion for hearing, earcon recognition, and icon recognition. The hearing test presented the subjects with three sets of tones that were played through the PC speakers. For the first set of tones, the subjects had to determine which tone of a pair was higher in pitch. For a second pair of tones subjects were required to say whether the rhythms of the tones were the same or different. For the final set of tones, the subjects had to distinguish whether the tones were played with the same instrument. Each pair of tones was played only once for each subject. In order to pass the hearing test, the subjects had to respond correctly to a minimum of ten out of fourteen tone samples. The hearing test was necessary because if subjects were not able to pass it, they may not have been able to distinguish the earcons and vocal cues.

During the experiment, all subjects were exposed to training in all cue types. In the earcon training, subjects were presented with each earcon via the PC speakers. They were then given ten minutes to listen to the earcons. All of the earcons were available to them, and they could listen to them as many times as they thought necessary, and in any order. After the ten minute training period, subjects were tested on the earcons using five different samples. The subjects had to choose from a list of all the possible earcons (21 in total) to determine which earcon was played. In order to pass the test, subjects had to correctly identify at least four out-of five earcons. If the subjects did not meet the specified criteria, they were given an additional ten minute training period to learn the earcons. They were then retested (using a different test). Again, they had to correctly identify four out-of five earcons. If the subjects did not pass the test, they were then disqualified from the study (no subjects were eliminated).

The icon test was similar in that it all of the icons and their meanings were initially presented to the subjects. The subjects then had five minutes to learn the meanings of the icons. The subjects were then presented with randomly ordered flashcards that showed pictures of the icons. They had to correctly identify four out-of five icons (no subjects were eliminated). Following the hearing, earcon, and icon tests, subjects were exposed to training in the VE. After the extensive training sessions in the VE, subjects then experienced two experimental trials.

The first trial for each subject was used to evaluate which form of simple cue was most effective. The second trial involved more elaborate cueing, and was expected to provide evidence of the implications of MRT on the effectiveness of the various cues in the visually-intense task because the various modalities might have been more taxed than in the first trial.

Each subject experienced two trials under a specific cue type. For instance, one subject received both simple and complex, vocal cues, whereas another experienced both trials with earcons.

The experimental trials posed the same level of task difficulty to all subjects. *Table 4* shows the overall procedure for the experiment and time estimates associated with the various steps. (The specific instructions to subjects are presented in *Appendix 1*.) Subjects were required to eliminate two mines in each trial. Therefore, there were two search, place charge, and detonation phases. The experiment required each subject to perform for approximately 2.5 hours, in total.

Table 4: Overall Procedures.

Step	Estimated Time
1. a brief equipment familiarization and instruction period;	10 minutes
2. a hearing test (See *Appendix 8*);	10 minutes
3. completion of a Sim-Sickness Questionnaire (Kennedy et al., 1994; see *Appendix 5*);	4 minutes
4. an extensive training session to learn and practice the maneuvers and commands of the system;	20 minutes
5. training in the presentation of auditory and visual cues (See *Appendix 9*);	15 minutes
6. an explanation of the HMD;	4 minutes
7. practicing in the demining task in the virtual underwater environment;	15 minutes
8. familiarization with the SART;	4 minutes
9. familiarization with the NASA-TLX workload measurement technique;	4 minutes
10. performance in the VE navigation and mine neutralization task including two trials, each followed by the NASA-TLX, SSQ, and a five-minute break;	50 minutes
11. a debriefing on the study	5 minutes
TOTAL TIME:	141 minutes
*If at any time during the execution of these procedures, subjects experience physical or psychological discomfort or fatigue, additional rest periods will be provided.	

3.8 HYPOTHESIS

In general, multiple sensory cueing was expected to improve SA according to MRT. It was hypothesized that vocal cues would be most effective in terms of performance and SA in the complex trial condition; however, earcons were expected to be more successful as far as performance and SA in the simple trial condition. This is based on the evidence previously presented on MRT and the additional stage of cognitive processing (translation of meaning) necessary for earcons over vocal cues. It was expected that earcons would yield the shortest trial times in the simple trial condition and that vocal cues would produce the shortest trials in the complex condition. In contrast, it was anticipated that trial times would be the longest when icons were used as a feedback mechanism. These trial time hypotheses were based on the subjects acknowledging, processing, and understanding the cues in a timely manner. It was also expected that trial times would be explainable in terms of MRT.

It was hypothesized that task errors and workload would be greatest in trials involving the use of icons because the visual cues would require the use of the already taxed visual modality. In general, it was expected that there would be higher perceived workload in the complex cueing conditions than in the simple cueing conditions. The subjects had more information to manage in the complex cueing conditions.

3.9 Data Analysis

3.9.1 General Statistical Analysis

Statistical analyses included one-way Analyses of Variance (ANOVAs) on all response measures. The simple ANOVA model for the experiment can be written as:

$$Y_{ij} = \mu + CT_i + CC_j + \varepsilon$$

Where:

Y_{ij} = the response variable (e.g., workload)

CT_i = Cue type

CC_j = Cue complexity

ε = Error

A full model incorporating a subject (SUB) term and all two-way interactions was investigated as part of a conservative approach to the analysis. The full model can be written as:

$$Y_{ijk} = \mu + CT_i + SUB\ (CT)_{j(i)} + CC_k + CT*CC_{ik} + CC*SUB(CT)_{jk(i)} + \varepsilon_{l\ (ijk)}$$

Where:

i = 1, 2, 3, 4

j = 1, 2, 3, 4, 5, 6, 7, 8

k = 1, 2

l = replicates under each level of complexity

However, since a repeated-measures design was not used and each experimental unit was observed only once under each cue type × cue complexity condition, the cue complexity × subject interaction in this model served as the error term in the majority of the statistical tests for significance on the other terms included in the full model.

The acceptable alpha level for all analyses was set at 0.05. Duncan's Multiple Range (MR) tests were used to further investigate any significant main effects or interactions, specifically to make comparison of mean responses for the cue type and complexity conditions. The SAS Code used for the primary data analysis can be viewed in *Appendix 10*.

Outliers were identified and removed from the various data sets on the response measures by considering Cook's D statistic for all observations calculated using the SAS GLM procedure. Cook's D values were considered extreme if they were at least one order of magnitude greater than all other Cook's D values. The DFFITs statistic was also calculated for all response observations using the SAS REG procedure in order to objectively identify outliers. Data points were seen as outliers if their DFFITs value was above 0.5 or below – 0.5. This value was determined by using the equation *2v(p/n)*, where, *n* was equal to the number of observations used in the model and *p* was equal to the number of parameters in the model, so 2v(4/64) = 0.5 (SAS, 1994, p.1419). Beyond this, extreme values for the response measures were identified in the main effects plots, and by examining model residuals. Any observations removed from the data sets were identified as extreme values according to the DFFITs statistic, Cook's D statistic, and residual values. The identification of outliers based on the previously mentioned statistics is described in Neter et al. (1990, pp. 392-393, 401-406).

All observations dropped from the data sets were replaced with the mean response for the specific experimental condition, which was represented by the removed observation (the particular cue type at the specified level of complexity), and consequently all response measures were analyzed using the same number of data points. The data points removed from the various data sets are presented in *Table 5*. In total, three data points were removed as outliers from the data on time-to-task completion, two data points were removed from the data on the demand rating as part of SART, two data points were removed from the data on the supply rating for SART, and two data points were removed from the data on the understanding rating for SART. The were no differences in the significant research findings when the outliers were included in the data analysis of time-to-task completion, SART demand, supply, and understanding components. Outliers, according to the influence statistics, were removed primarily to ensure all assumptions of the ANOVA were upheld in the data analysis. The number of outliers removed from the data set only represented a small percentage (1.6%) of the total number of observations.

Table 5: Data points identified as outliers.

Response Measure	Subject #	Trial #	Value	Replaced Value
TTC	10	1	961	488
TTC	10	2	927	505
TTC	25	1	960	506
SART Demand	17	1	0.5	6.32
SART Demand	17	2	0.5	6.44
SART Supply	6	1	8.3	3.85
SART Supply	25	2	1.55	4.93
SART Understanding	11	1	2.75	8.09
SART Understanding	11	2	2.45	7.86

The normality and constant variance assumptions of the ANOVA were evaluated for each response measure. Normality was assessed using normal probability plots and Shapiro-Wilks test. The constant variance assumption was evaluated by examining the residual plots for all response measures by cue type and cue complexity. Linear trends in the normal probability plots were observed for all measures and all Shapiro-Wilks tests were insignificant ($p>0.05$). The residual plots revealed equal variance for the response measures across the settings of cue type and cue complexity.

4. RESULTS

4.1 TIME-TO-TASK COMPLETION

At the outset of the data analysis, an ANOVA was conducted on time-to-task completion for the no cue control condition in order to determine whether there was a significant effect of trial. A simple model in trial number was used for this analysis. The ANOVA result revealed a significant effect of trial on subject performance when no cueing of system state or task phase changes was provided ($F(1, 14) = 5.66$, $p = 0.0321$). Duncan's MR test revealed that subjects took significantly longer ($p<0.05$) to dispose of mines during the second trial as part of their participation in the experiment than in the first trial (as much as 2 min and 37 s longer with average trial times ranging from 7 min and 28 s to 10 min and 7 s). It is possible that a fatigue effect may have occurred across the trials leading to slower mine disposal times towards the end of the experiment.

With this in mind, the data for each no cue condition trial was handled separately in the primary experimental analysis in order to account for a potential fatigue effect. That is, the ANOVA and post hoc tests were used to make comparisons of performance under each cue type × cue complexity condition with mean performance for the first trial as part of the no cue condition, as well as the second.

An ANOVA on time-to-task completion, using the full model presented in Section 4.1, revealed a significant effect of the interaction term for cue type and cue complexity ($F(3, 52) = 3.01$, $p = 0.0382$). *Figure 8* presents an interaction plot for the four cue type and two complexity conditions. However, the main effects of cue type and cue complexity were not significant in influencing time-to-task completion. Given the significant interaction of cue type and cue complexity, post-hoc tests were conducted in order to determine the breakout of the means for the levels of the interaction. Duncan's MR Tests revealed three different classes of means, as presented in *Figure 9*. The no cue condition during the second test trial and the icon condition during the simple trial appeared to produce the slowest times-to-task completion. The icon and earcon conditions during the complex trial and the first no cue condition trial appeared to produce moderate times-to-task completion. Finally, the simple and complex vocal cueing conditions as well as the simple earcon condition all appeared to produce shorter trial times. However, only the no cue condition during the second test trial was significantly worse ($p<0.05$) than the conditions in the final class. In addition, vocal cueing, and the simple earcon condition were significantly superior ($p<0.05$) to the simple icon and performance during the second trial of the no cue control condition.

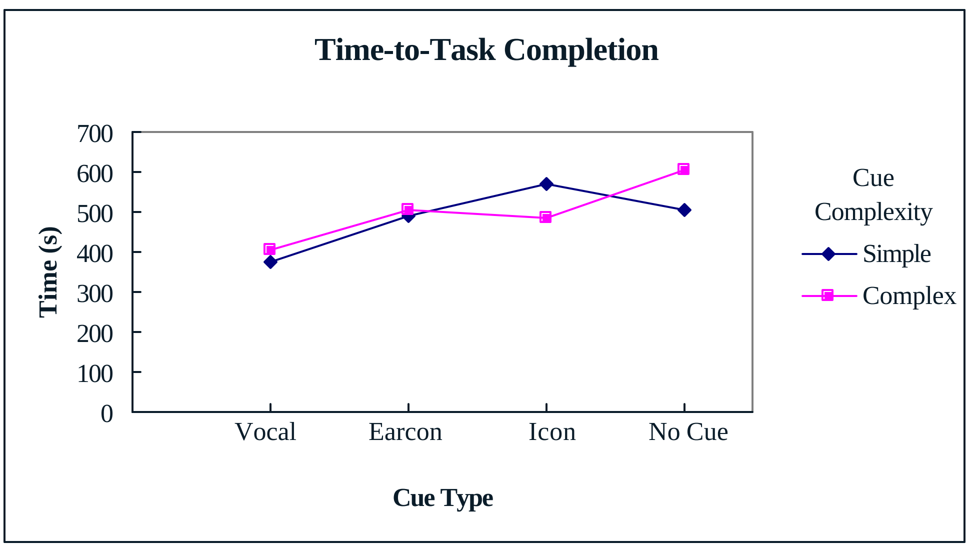

Figure 8. Interaction Plot of cue type and cue complexity on Time-to-Task Completion.

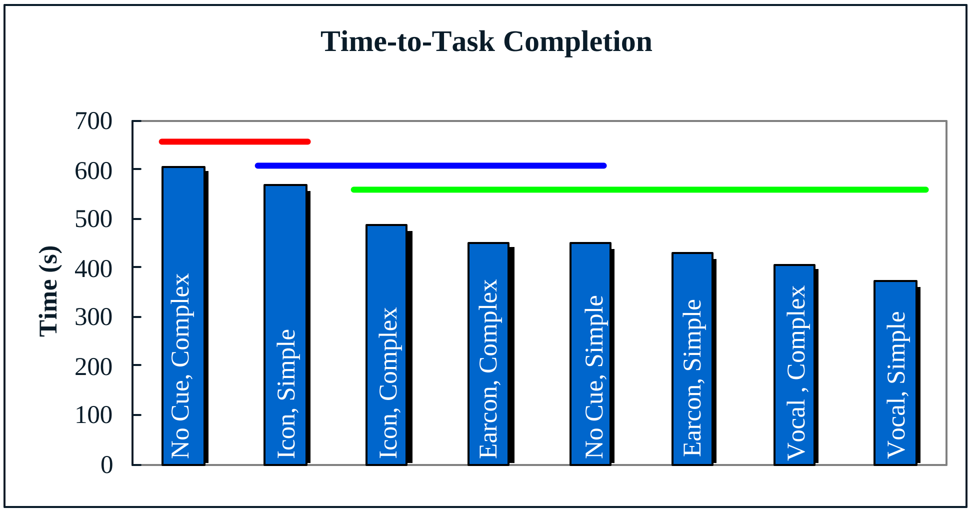

Figure 9. Duncan grouping of time-to-task completion.

To some extent, this supports Graham's (1999) claims that earcons are more effective than icons or no cue for time-to-task completion; however, this appears to only be true when subjects are posed with simple icons and after they have performed a task for an extended period with no cues. The results also provide some evidence of MRT because both vocal cueing conditions and simple earcons produced faster trial times than icons and no cues. This is in accordance with MRT because the task is already visually-rich. Since icons utilize the visual modality, the visual channel may be overloaded, or visual attentional resource competition occurs, which is not a factor when using auditory cues. The lengthier trial times associated with complex earcons in comparison to those involving vocal cues could be due to the amount of information encoded into the earcons and the need for subject interpretation. This effect was previously hypothesized.

As far as time-to-task completion, the results generally indicate that over time, no cueing produces worse performance than any form of cueing whatsoever (vocal, earcons, icons, simple, complex). They also indicate that vocal cueing is superior to simple icons and no cueing, particularly if operators may be fatigued or bored. These findings are consistent with Kaber et al.'s (in preparation) results, which revealed auditory (earcons) and bimodal cueing (earcons and icons) to be superior to visual cueing (icons).

Since there were no main effects in this analysis, significant differences among the various cue types existed only across levels of cueing complexity. It was hypothesized that simple earcons would yield the shortest trial times in the simple cueing condition and vocal cues would yield the shortest trial times in the complex cueing condition. Earcons were expected to produce the shortest, simple trial times because they provide information to the user through the auditory modality without competing with vocal communication. Previous studies have demonstrated that earcons can be a successful mechanism for providing complex system state information to operators (Dowling, 1986; Gaver, 1986; Buxton & Baecker, 1987; Blattner, Greenberg & Sumikawa, 1989; Brewster, 1998; Walker & Brewster, 2000; Brewster, Edwards & Wright, 2001; Kaber, Hughes, Wright, Sheiknainar & Warren, in preparation). However, vocal cues were anticipated to be superior in the complex cueing condition because they convey information without requiring additional mental transformation for comprehension that is required with earcons. Complex earcons require more information to be translated than simple earcons. This study did not reveal simple earcons to be superior to simple vocal cues. According to the statistical analysis, the shortest trial time for both simple and complex cueing conditions was vocal; however, there was no significant difference among these conditions. This could be a result of the simplicity of the experimental task, or the fact that the task was sequential in nature. It is possible that a more complex task, or one requiring time-sharing of subtasks may produce results supporting the original hypothesis.

Although cue type and cue complexity were not independently significant, as hypothesized the average times-to-task completion were lower for both the simple and complex vocal cues. It is possible that if the task had been more difficult, there may have been a more significant breakout of mean performance for the various cue types. It is also possible that a fatigue effect occurred, which is evidenced by the no cue condition's data; however, this effect may have extended beyond the no cue condition to the experimental trials involving the other cue types. Since the simple cue condition was always presented as the first trial and the complex cue condition was always presented in the second trial, it is possible that a fatigue effect may have confounded the significance of a cue type main effect.

4.2 SITUATION AWARENESS

The statistical analysis conducted on the overall SART score data did not yield any significant results. In general, this would indicate that the differences in feedback provided to subjects were not substantial enough to cause differences in high-level cognition. However, it is possible that the SART measure of SA was not sensitive enough to capture the differences between cue types and cue complexity. Kaber et al. (in preparation) used the SAGAT (Endsley, 1995) and did not find a significant difference between auditory or visual cueing in terms of SA, but they did find a significant difference between bimodal cueing and auditory and visual cueing. The SAGAT may have been more sensitive to changes in SA due to system interface manipulations since it is an objective rather than subjective measure.

Since the overall SART data did not reveal any significant differences among conditions, further analyses were conducted on the demand, supply, and understanding components of the SART. An ANOVA on SART understanding ratings revealed no significant main effects or the presence of the two-way cue type by cue complexity interaction.

4.2.1 SART DEMAND

On the basis of MRT, it was hypothesized that auditory cueing would improve SA by placing demands on attentional resources other than visual attention. The analysis of the SART demand rating did not reveal a significant effect of cue type ($F(3, 52) = 2.10$, $p = 0.1118$) when considering the model error as the denominator in the F-test. The main effect was also insignificant when individual differences within cue type group were considered in the F-test ($F(3, 52) = 1.15$, $p = 0.431$). It is important to note that the experimental sample size was sufficient to render any effect of the subject term in the SART demand-rating model insignificant ($F(4, 52) = 1.82$, $p = 0.1383$). However, the variance due to individual differences was greater than that attributable to the model error term.

Despite the lack of a significant cue type effect and a means breakout, the trend of attentional demand across cue types was logical. Average rating appeared to decrease from icon to earcon, to no cue, to vocal cueing as can be seen in *Figure 10.* On average, vocal cueing may be less demanding because it was the only cue type that involved natural language processing; therefore, translating the cue into comprehension of system states may have been less attentionally demanding for subjects. On average, earcons may be more demanding than vocal cues because of mental translations required to comprehend the cues.

On average, icons were more attentionally demanding than the earcon and vocal cue types, which is in line with MRT. Earcons and vocal cues utilize the auditory channel for presenting task information, whereas icons utilize the visual channel. Since the visual channel was already used for presentation of the VE, additional items presented via the visual channel may have overloaded attentional resources.

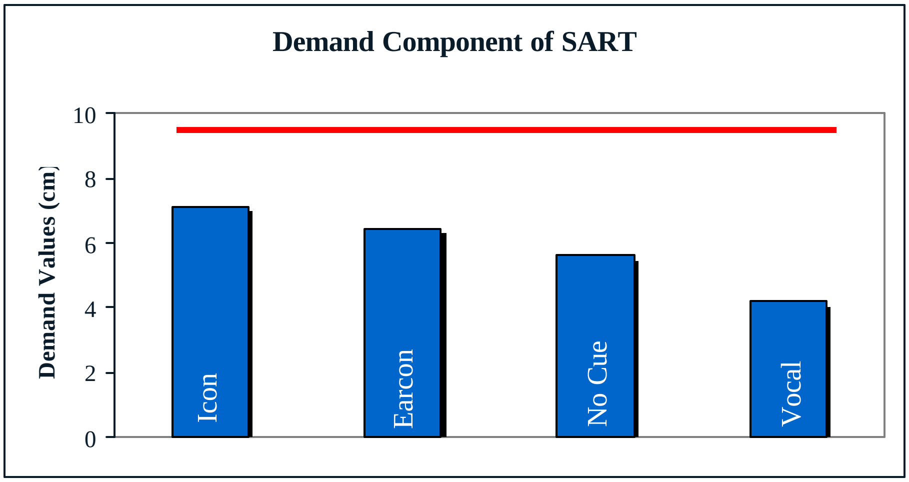

Figure 10. Average Demand Component Ratings as Part of SART.

4.2.2 SART SUPPLY

As in the SART demand rating analysis, cue type appeared to be significant when considering the Mean Square Error in the F-test ($F(3, 52) = 5.07$, $p = 0.0037$). However, there were significant individual differences within cue type group ($F(4, 52) = 4.25$, $p = 0.0047$). When considering the Mean Square Error for the subject term in the F-test on cue type, the main effect proved to be insignificant ($F(3, 52) = 1.19$, $p = 0.4188$).

4.3 WORKLOAD

It was hypothesized that workload would be greatest in trials involving the use of icons because the visual cues would require the use of the already taxed visual modality. In general, it was expected that there would be higher perceived workload in the complex trial conditions than in the simple trial conditions. The analysis of overall workload scores captured using the NASA-TLX did not reveal any significant effects. Kaber et al. (in preparation) found a significant difference in workload associated with cue type, specifically visual cues resulted in a higher perceived workload than auditory cues.

With this in mind, further analyses were conducted on those demand components of the TLX that subjects ranked as being most important to task performance at the beginning of the experiment. Although the mental demand and effort components were ranked as highly important, there was no significant effect of cue type, cue complexity, or individual differences on subject perceptions of mental task load, or the overall effort required in task performance.

4.3.1 PERFORMANCE COMPONENT OF WORKLOAD

Vocal cues were anticipated to be most effective in terms of perceived performance in the complex trial condition; however, earcons were expected to be more successful as far as subject perceptions of performance and SA in the simple trial condition. This is because the cues instructed the user as to what they should do. If the users understood the cues, they should have been confident in their performance. Therefore, the cues that produced the best performance ratings should be those that were best understood.

The analysis of performance ratings revealed a significant effect of cue type when considering the Mean Square Error (F (3, 52) = 3.01, p = .0383). Individual differences were not significant in effect on subject ratings of perceived performance. However, because the cue type effect was close to the significance criterion, Duncan's test did not appear to be sensitive enough to reveal differences among the cue type condition means. *Figure 11* shows the mean subjective ratings of performance using the NASA-TLX measure for each cue type condition, including the no cue control group. On average, icons produced higher performance ratings and may have instilled greater confidence in subjects in their performance. On average, vocal cues may have given subjects more confidence in task performance than earcons and no cueing. Average perceptions of performance under no cueing conditions appeared to undermine subject confidence in performance, as compared to when receiving any form of feedback.

One possible explanation for the icons producing higher average performance ratings is that subjects may have lacked accurate understanding of the state of the task or the automation in order to make an accurate internal assessment of their performance. With the vocal cues and earcons, a subject's attention was immediately drawn to the cues upon presentation and, consequently, their mental model of the task circumstances may have been regularly and accurately updated as a basis for evaluating their own performance. The icons may have been less salient than the vocal cues and earcons as subjects had to direct their visual attention to an icon in order to perceive it and their visual channel was already taxed by the VE interface.

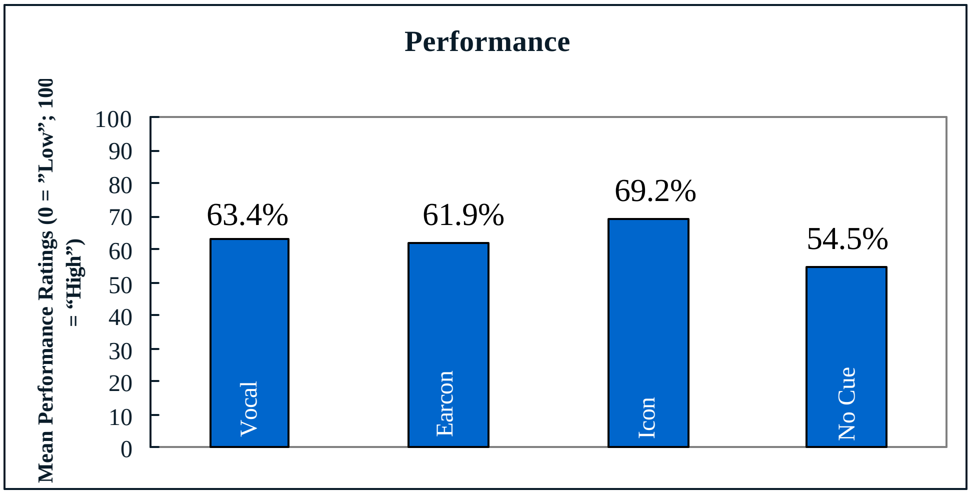

Figure 11. Standard Means Plot for the Performance component of the TLX.

4.4 ERRORS

It was hypothesized that task errors and workload would be greatest in trials involving the use of icons because the visual cues required use of the already taxed visual modality. Errors were recorded when there was an incorrect association of the charge type and the mine type. Very few errors occurred during the experimental trials, in fact, there was only one error per cue type condition. However, all mine classification errors occurred under the simple cueing complexity, so it is possible that complex cueing may prevent errors. Further analysis with a larger data set would be necessary to determine if the errors consistently occur in the simple cueing condition.

Kaber et al. (in preparation) found that significantly more errors occurred in the visual cueing condition than in the auditory or bimodal cueing condition. However, overall there were far fewer errors committed in this experiment than in the study by Kaber et al. (in preparation). The most likely explanation for this is the current study defined errors as selecting the wrong charge type specified for a particular mine type; however, Kaber et al.'s (in preparation) study had a much broader definition of errors, such as inactivity in manual control or selection of the incorrect tool for the task.

5. CONCLUSION

5.1 CAVEATS

Despite the effort to simulate a realistic teleoperation environment, there were several caveats to the experiment. First, the task involved an abstract lab test and certain real-world conditions could not be replicated. In the lab setting, the ambient noise level and subject interactions were controlled. Although the experiment utilized a high fidelity VR simulation, regardless of how real it may have seemed to subjects, it did not create job and temporal stress levels that would be considered comparable to those experienced in an actual teleoperation environment.

Another caveat of the research was that naïve students were used as subjects rather than real teleoperators. The use of real teleoperators could have impacted the degree of subject involvement in the task, and, consequently, the task performance times, errors, operator SA, and ratings of workload.

5.2 GENERAL CONCLUSIONS

Time-to-task completion was significantly effected by the cue type and cue complexity interaction. The vocal cue type was superior to icons for both simple and complex task conditions and no cueing over an extended period of time. Simple vocal cues produced the fastest trials, on average. The current experimental findings are supported by Bly (1982), who demonstrated cues utilizing auditory modalities produced faster trial times; however, Bly's research did not involve study of an adaptive system. Graham (1999) found earcons to produce faster performance in responses to emergency warnings in an automobile collision detection scenario; however, Graham did not investigate the use of vocal cues.

The current research suggests that the superiority of cues utilizing the auditory modality in visually-rich task environments is consistent across both standard and adaptive automation. Simple earcons were also significantly better in terms of performance than icons or no cueing, which is in line with MRT. Multiple Resource Theory suggests that use of a multiple modalities will increase overall performance. Since the teleoperation task was visually intensive, a modality for cueing other than the visual channel was expected to be best in terms of performance. The vocal cueing condition and earcons both involved use of the auditory modality; however, icons and the no cueing condition did not make use of other modalities than visual. Interestingly, complex earcons did not appear to differ from icons and the no cueing condition. On this basis, it may be possible that a certain amount of information can be successfully learned and encoded using earcons; however, after a specific point too much information encoded in earcons is no longer supported by MRT.

The overall SA data produced by the SART technique was insignificant. Both the demand and supply ratings as part of the SART measure were significantly affected by cue type, but this was not pervasive across subjects. It is possible that an objective SA measure, such as SAGAT (Endsley, 1995), may have been more sensitive to truly significant differences in operator cognition among cue type conditions. A previous study by Kaber et al. (in preparation) found a significant difference in cue types when utilizing SAGAT.

On the basis of the present results, it seems that the either the SART measure was not sensitive enough to capture differences in SA or the task was not complicated enough to generate significant differences in perception, comprehension, and projection under the various cueing conditions.

The overall workload data produced by the NASA-TLX was not significantly influenced by cue type; however, further analysis was performed on the three highest ranked demands of the TLX, including mental demand, performance, and effort. Of these three dimensions, only performance ratings revealed significant results. Walker and Brewster (2000) previously found the use of sound to significantly reduce workload in the context of interfaces for wearable computers.

The evaluation of errors did not yield enough data in order for a valid parametric statistical analysis of the cue type and cue complexity manipulations to be performed. However, the available data did reveal that all errors occurred in the simple cueing condition, suggesting more information on system states may be better for effective task performance. However, more data is needed to conclusively state more cueing is better.

5.3 Design Recommendations

Given the results of the current research, several design recommendations can be offered to improve the overall design of future teleoperation systems utilizing adaptive automation. Since the types of cues used in this research were generic in form and only specific in content, the results of the work can be extended beyond the specific teleoperation task and have ramifications for not only other teleoperation tasks, but also other tasks employing adaptive automation. The recommendations listed below are general, and could be tailored to specific circumstances, as well as specific task requirements, which is often the case with teleoperation systems

- In the present task, both simple and complex vocal cues were superior. Therefore, similar tasks involving visually-rich system interfaces should employ vocal cueing, if there is not already voice communication required by the task (such as human-human or human-machine voice communication).

- Earcons are a beneficial cueing mechanism that when used in simple cueing circumstances are superior to simple icons and no cueing whatsoever. It is possible that tasks requiring more voice interaction may conflict with the use of vocal cues on system states and may be better suited to utilizing earcons, particularly if icons are considered as a feedback alternative.

- When using earcons, it is important that the users are well trained so they can effectively recognize and infer the meanings of the earcons. Any advantage of earcons over icons and no cueing only appear to exist when the information content of the earcon is simple and easy to interpret.

- When designing a multimodal teleoperation interface, it is important to ensure that the bandwidth of the communication link between the system and the user is sufficient to provide timely feedback on system states (Chou & Wang, 2001). If the bandwidth insufficient, then the system may not perform effectively and efficiently regardless of the type of operator cueing that is used.

- Performance without cueing appears to become worse over time. This suggests that some form of cueing is always better than none (with the exception of simple icons in a visually-rich task environment). Given this result, interfaces should be designed to use some form of cueing via a modality other than the most prominent sensory channel used by the system interface (e.g. if the task is visually-rich, use auditory, or if the task is auditorily-rich, use visual, etc.).

5.4 Future Research

As with all experiments, there are aspects that could be improved upon in future research to more effectively analyze and expand upon the results. For example, a more complex task could be utilized in which the actions are not necessarily sequential. Therefore, the subject would be forced to pay close attention to the cues at all times. With such a task, an objective SA measure like SAGAT (Endsley, 1995) may be more appropriate because it could be used to evaluate SA at different points in time during a trial and not as a rating of average operator SA at the close of a trial. However, if SAGAT were to be used, it would be helpful if it could be administered without distracting the user from the experimental task. Otherwise, task freezes required for administering SAGAT queries could result in increased operator vigilance that would not necessarily occur in the real task.

Since this task was designed to simulate a teleoperation task in the real-world, the most effective cue type for the task would likely be highly dependent upon the task environment and individual operator preferences. In an environment similar to the one used in this experiment involving a single teleoperator and a computer, vocal cues are highly effective because there is little voice communication between the operator and the machine.

However, since a Wizard of Oz technique was used in this experiment to control tool selection and the rover arm movements, there was some human-human voice interaction. It was suspected that this communication might conflict with the vocal task cues and cause the earcons to be a superior form of feedback. This did not prove to be the case. However, if the task environment required a significant amount of vocal communication, or involved an interactive team scenario, then vocal cues might not be as effective because of resource competition with ambient vocal noise.

It may also be important for future studies to compare specific frequencies (pitches) of speech and earcons for cueing dynamic, complex system states. Certain voices may be more appropriate in certain working environments. For example, female voices have been found to draw greater attention in complex task warnings than male voices (Sanders & McCormick, 1993). Properties of voice like volume, inflection, and emotion could also be altered to provide the best cue possible.

Another property of auditory cues that may be beneficial to investigate is the direction of auditory cues. This could be performed either through stereo headphones or well-placed stereo speakers. Johannsen (2001) conducted an experiment involving maneuvering a teleoperated robot around a grocery store setting. The subjects had to determine the direction of the robot's trajectory and the system state of the robot using earcons (although the author never specifically called them "earcons"). The path that the subject was to move the robot along was described by directional earcons. Directional earcons proved to be effective in terms of performance for both musicians and non-musicians (Johannsen, 2001). Directionality could be added to both the earcons and the vocal cues investigated in the present research to perhaps improve performance and possibly create significant differences between auditory cues and visual cues and/or no cueing conditions.

Another potentially beneficial step that could be taken to improve upon the findings of this research would be to combine bimodal cues with the use of vocal cues and earcons. Kaber et al. (in preparation) found bimodal cues using earcons and icons to be superior in a condition comparable to the simple cueing condition of this experiment. Scerbo (1996) also supports the use of bimodal cues in adaptive automation. For example, it would be interesting to see how the use of vocal cues plus icons compares to earcons plus icons. Another possible combination of cue types that may significantly improve performance in visually-rich environments is haptic cues plus either earcons or vocal cues. Sklar and Sarter (1999) support the use of bimodal cues involving haptic stimuli; however, research has yet to be done involving bimodal haptic cues in an adaptive system.

Finally, it would also be interesting to see if the results found in this study are robust across different age groups. The students used in this experiment ranged in age from 18-30 years. Barr and Giambra (1990) found as humans age, degenerative effects decrease the ability to distinguish sounds. Given this information, it is possible that the superiority of cue types may change with age. If the vocal cues or earcons were not found to be significantly better than icons for an older population, it is possible that simple cues could be distinguished more radically from one another to create a significant improvement (e.g., increased deviations in frequency and timbre, more drastic rhythm changes from one cue to another, etc.).

6. WORKS CITED

Albers, M. C. (1996). Auditory Cues for Browsing, Surfing, and Navigating [Online]. *The Proceedings of International Conference on Auditory Display*. Available: http://www.santafe.edu/~icad/ICAD96/proc96/albers5.htm [2001, July 18].

American Institute of Graphic Arts. (1981). *Symbol signs*. New York: Hastings House.

Aukstakalnis, S. and Blattner, D. (1992). *The Silicon Mirage: The Art and Science of Virtual Reality.* Berkeley, CA: Peach Pit Press.

Ballas, J. A., Heitmeyer, C. L., and Perez, M. A. (1991). Interface styles for adaptive automation. In *Proceedings of the Sixth International Symposium on Aviation Psychology* (Vol.1, pp. 108-113). Columbus, OH: The Ohio State University.

Banks, W. P., Roberts, D. and Ciranni, M. (1995). Negative priming in auditory attention. *Journal of Experimental Psychology: Human Perception & Performance*, 21(6): 1354-1361.

Barr, R. A., and Giambra, L. M. (1990). Age-related decrement in auditory selective attention. *Psychology and Aging*, 5, 597-599. In Birren, J. E., & Schaie, K. W. (Eds.) *Handbook of the psychology of aging* (pp.181-203). San Diego, CA: Academic Press.

Belleman, R. G., Stolk, B., and de Vries, R. (2001). Immersive Virtual Reality on commodity hardware. *Proceedings of the 7th annual conference of the Advanced School for Computing and Imaging* (pp. 297-304). Amsterdam, the Netherlands: ASCI.

Blattner, M., Sumikawa, D. and Greenberg, R. (1989). Earcons and Icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11-44.

Bly, S. (1994). What do we learn when several researchers sonify the same data? In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*, (pp. 405-416). Reading, Mass: Addison Wesley.

Bowers, J. A., Oser, R. L., Salas, E. and Cannan-Bowers, J. A. (1996). Team performance in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application* (pp. 243-263). Mahwah, NJ: Lawrence Erbaum Associates, Inc.

Brewster, S. A. (1998). Using Nonspeech Sounds to Provide Navigation Cues [Online]. In *ACM Transactions on Computer-Human Interaction*, 5(3), (pp. 224-259) [Available: http://www1.acm.org/pubs/contents/journals/tochi/1998-5/] .

Brewster, S.A. and Walker, V.A. (2000, November 29). Non-Visual Interfaces for Wearable Computers. Presented at the *IEE Workshop on wearable Computing.* London, U. K.: IEE Press.

Brewster, S., Wright, P.C, and Edwards, A.D.N. (2001). Guidelines for the creation of earcons [Online]. *Glasgow Interactive Systems Group*. [Available: http://www.dcs.gla. ac.uk/~stephen/earcon_guidelines.shtml].

Buxton, W and Baecker, R. (Eds.) (1987). *Readings in Human Computer Interaction: A Multidisciplinary Approach*. Los Altos, CA: Morgan Kaufmann Inc.

Buxton, W. (1995). Speech, Language & Audition. In R.M. Baecker, J. Grudin, W. Buxton and S. Greenberg, S. (Eds.) *Readings in Human Computer Interaction: Toward the Year 2000*. San Francisco: Morgan Kaufmann Publishers.

Byrne, E.A., and Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, 42, 249-268.

Chou, W. and Wang, T. (2001). Design of Multimodal Human-Machine Interface. *Proceedings from IEEE International Conference on Systems, Man, and Cybernetics*, 5, 3187-3192.

Davies, C. (1996). Cognition, Perception, and Experience in the Virtual Environment: Do You See What I See? *Panel Session at the SIGGRAPH '96 Conference*, New Orleans, LA.

Deese, V. and Grindley, G. C. (1947). The transposition of visual patterns. *British Journal of Psychology*, 37, 152-163.

Egan, D. E. (1988). Individual differences in human-computer interaction. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 543-568). Amsterdam: North-Holland.

Elliot, F. R. (1937). Eye versus ear in moulding opinion. *Public Opinion Quarterly*, 1, 83-86.

Gabbard, J. L. and Hix, D. (1997). *A Taxonomy of Usability Characteristics in Virtual Environments.* (Grant ONR-N00014-96-1-0385). Blacksburg, VA: Virginia Tech.

Gaver, W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2, 167-177.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics*, 42(9), 1233-1248.

Hancock, P. A. and Chignell, M. H. (1987). Adaptive control in human-machine systems. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 305-345). North Holland: Elsevier.

Hancock, P. A. and Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man and Cybernetics*, 18, 647 658.

Hancock, P., and Scallen, S. (2001). Implementing Adaptive Function Allocation. *The International Journal of Aviation Psychology*, 11(2), 197 – 221. Lawrence Erlbaum Associates, Inc.

Hart, S. G. and Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload, 139-183. Amsterdam, The Netherlands: Elsevier.

Hilburn, B., Jorna, P. G., Byrne, E. A., and Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua and J. Koonce (Eds.), *Human-Automation Interaction: Research and Practice*. (pp. 84-91). Mahwah, NJ: Erlbaum Associates.

Hinckley, K., Pausch, R., Goble, J. C., and Kassell, N. F. (1994). Design hits for spatial input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software & Technology* (UIST '94), (pp. 213-222). New York: ACM.

Johannsen, G. (2001). Auditory displays in human-machine interfaces of mobile robots for nonspeech communication with humans. *Journal of Intelligent and Robotic Systems*, 32, 161-169.

Jones, D. G. (2000). Subjective Measures of Situation Awareness. Available in *Situation Awareness Analysis and Measurement*, eds. Endsley, M. R. & Garland, D. J, (pp.113-122). Daytona Beach, FL: Embry-Riddle Aeronautical University.

Kaber, D. B. (1999). Virtual Reality for Scientific Visualization. In D. Harris (Ed.), *Proceedings of the 3rd International Conference on Engineering Psychology and Cognitive Ergonomics*. Aldershot, UK: Ashgate.

Kaber, D. B. and Riley, J. M. (1999). “Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement.” *International Journal of Cognitive Ergonomics*, 3(3), 169-187.

Kaber, D. B., Riley, J. M., Tan, K., and Endsley, M. R. (2001). “On the Design of Adaptive Automation for Complex Systems.” *International Journal of Cognitive Ergonomics*, 5(1), 37-57.

Kaber, D. B., Hughes, L.E., Wright, M.C., Sheiknainar, M., and Warren, H.L. (in preparation). Automation state changes and sensory cueing in complex systems control (Final Report: Office of Naval Research Award #N0001401010402). Arlington, VA: Office of Naval Research.

Kanki, B. G., and Palmer, M. T. (1993). Communication and crew resource management. In E. L. Wierner, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 99-136). San Diego, CA: Academic Press.

Kerman, J. (1980). *Listen*. New York: Worth.

King, R. B. and Oldfield, S. R. (1997). The impact of signal bandwidth on auditory localization: Implication for the design of three-dimensional audio. *Human Factors*, 39 (2), 287-295.

Mapes, D. P. and Moshell, J. M. (1995). A two-handed interface for object manipulation in virtual environments. *Presence: Teleoperators and Virtual Environments*, 4(4), 403-416

Martin, R. C., Wogalter, M. S., and Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*. 27, 382-398.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychology Review*, 63(2), 81-97.

Morris, J. G. and Gluckman, J. P. (1994). Definitions and prospective guidelines for the application of adaptive automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 256-263). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mountford, S. J. and Gaver, W. W. (1990). Talking and listening to computers. In B. Laurel (Ed.), *Auditory display: Sonification, audification, and auditory interfaces*. Reading, Mass: Addison Wesley.

Mynatt, E. D. (1994). *Auditory Presentation of Graphical User Interfaces*. Santa Fe. Addison-Wesley: Reading MA.

Neter, J, Wasserman, W., and Kutner, M. H. (1990). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Boston, MA: Irwin.

Norman, D. A. (1990). *The problem of automation: Inappropriate feedback and interaction not overautomation* (ICS Rep. 8904). La Jolla, CA: Institute for Cognitive Science, University of California, San Diego.

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* (Vol.2, pp. 43.1 43.9). New York: Wiley.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., and Barnes, M. (1992). Theory and design of adaptive automation in aviation systems (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R., Molloy, R., and Singh, I. L. (1993). Performance Consequences of Automation-Induced "Complacency". *The International Journal of Aviation Psychology*, 3(1), 1-23.

Parasuraman, R., Mouloua, M., and Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38, 665-679.

Parsons, H. M. (1985). Automation and the individual: Comprehensive and comparative views. *Human Factors*, 27, 99-111.

Rabiner, L. R., Kamm, C. A., and Walker, M. A. (1997) The role of speech processing in human-computer intelligent communication. *Speech Communication*, 23, (4), 263-278.

Richard, P., Birebent, G., Coiffet, P., Burdea, G., Gomex, D., and Langrana, N. (1996). Effect of frame rate and force feedback on virtual object manipulation. *Presence: Teleoperators and Virtual Environments*, 5(1): 95-108.

Rouse, W. B. (1976). Adaptive allocation of decision making responsibility between supervisor and computer. In T. B. Sheridan & G. Hohannsen (Eds.), Monitoring behavior and supervisory control (p. 295-306). New York: Plenum.

Sanders, M. M., and McCormick, E. J. (1993). *Human Factors in Engineering & Design*. 7^{th} ed., McGraw-Hill, NY.

Sarter, N. B., and Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37, 5-19.

SAS Institute, Inc. (1994). *SAS/STAT User's Guide.* Cary, NC: SAS Institute, Inc.

Scallen, S. F., Hancock, P. A., and Duley, J A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics*, 26 (6), 397-403.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Human performance in automated systems: Theory and applications* (pp. 37-64). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Shiff, B. (1983). An experimental study of the human-computer interface in process control. Unpublished thesis, University of Toronto, Toronto.

Sklar, A.E. and Sarter, N.B. (1999). Good Vibrations: The Use of Tactile Feedback In Support of Mode Awareness on Advanced Technology Aircraft. *Human Factors*, 41(4), 543-552.

Stanney, K. (1995). Realizing the full potential of virtual reality: Human factors issues that could stand in the way. In *1995 IEEE Virtual Reality Annual International Symposium Proceedings* (pp. 28-34). New York: IEEE.

Sticht, T. G. (1969). *Learning by listening in relation to aptitude, reading, and controlled speech* (Tech. Rep. No. 69-23). Alexandria, VA: Human Resources Research Organization.

Tannen, R. S. (1998). Breaking the sound barrier: Designing auditory displays for global usability [Online]. *Proceedings of the Fourth Conference on Human Factors and the Web*. [Available: http://www.research.att.com/event/hfweb/proceedings/tannen/].

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations* (AGARD-CP-478; pp. 3/1-3/17). Neuilly Sur Seine, France: NATO AGARD.

Taylor, R. M., and Selcon, S. J. (1991). Subjective measurement of situational awareness. In Queinnec & Daniellou (Eds.), *Designing for everyone, Proceedings of the 11th Congress of the International Ergonomics Association* (pp. 789-791). London: Taylor & Francis.

Taylor, R. M. (1995). Experimental Measures: Performance-Based Self Ratings of Situational Awareness. Available in *Experimental Analysis and Measurement of Situation Awareness*, eds. Garland, D. J. & Endsley, M. R. (pp. 83-91). Daytona Beach, FL: Embry-Riddle Aeronautical University.

Ware, C. and Balakrishnan, R. (1994). Reaching for objects in VR displays: Lag and Framerate. *ACM Transactions on Computer-Human Interaction*, 1(4): 331-356.

Weiner, E. L. (1988). Cockpit automation. In E. L. Wierner & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 433-459). San Diego, CA: Academic.

Weiner, E. L. (1989). Glass Cockpit. *Human Factors of Advanced Technology* Transport Aircraft, NASA, Washington, D. C., Technical Report 117528.

Wickens, C. D. & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd Edition). Upper Saddle River, NJ: Prentice Hall.

Yoshida, A. Rolland, J. P., and Reif, J. H. (1995). Design and applications of a high-resolution insert head-mounted display. In *1995 IEEE Virtual Reality Annual International Symposium Proceedings*, (pp. 84-93). New York: IEEE.

Youngblut, C., Johnson, R. E., Nash, S.H., Wienclaw, R.A., and Will, C.A. (1996). Review of Virtual Environment Interface Technology, (IDA Paper P-3186). Institute for Defense Analysis.

APPENDIX 1: SUBJECT INSTRUCTIONS

INTRODUCTION

Thank you for volunteering to participate in this experiment. The goal of this experiment is to determine the performance effects of advanced notification of changes in the level of automation of a complex system. The experimental task will require you to use a standard 2-D mouse and vocal commands to control a simulated, semi-autonomous virtual mobile robot (or rover) in a virtual underwater mine neutralization operation. The rover is equipped with a robotic arm that can be used to pick up and move objects. The experimental task will require you to navigate the vehicle to locate mines and to use virtual demining tools on the robotic arm to detonate mines. The level of automation of the rover will periodically change throughout the operation between manual control and full automation. These changes in control mode will be signaled to you through visual and auditory cues. The virtual environment simulation will be presented to you through a head-mounted display, or HMD. During the experiment, you will complete an extensive training session and two test trials.

OVERVIEW OF PROCEDURES

The procedures we will follow during the experiment include:

1. A brief equipment familiarization and instruction period;
2. A hearing test;

3. A sim-sickness questionnaire;
4. An extensive training session to learn and practice the commands and maneuvers of the robotic system;
5. Training in the presentation of auditory and visual cues;
6. Explanation of the HMD;
7. Practicing the demining task in the underwater environment;
8. Familiarization with the Situational Awareness Rating Technique (SART)
9. Familiarization with the NASA-Task Load Index (TLX), a workload measurement technique; and
10. Performance in the virtual environment navigation and mine neutralization task. There will be two test trials. Each trial will be followed by the SART form, NASA-TLX workload rating, sim-sickness questionnaire, and a five-minute break. At the conclusion of the trials, there will be a debriefing on the study. If at any time during our execution of these procedures you experience physical or psychological discomfort or fatigue, a rest period will be provided. The total approximate time for completion of the experiment is 2 hours.

INFORMED CONSENT AND ANTHROPOMETRIC DATA SHEET

[Give the subject the informed consent form. Summarize the informed consent for the subject and encourage them to read the form.]

This form summarizes the information that has been presented to you thus far and identifies the persons responsible for the study. The form also addresses University liability to the experiment. I encourage you to read the form. This form will not be associated with any of the other survey forms used in this experiment. In order to participate in this study you must have 20/20 or corrected to normal vision, and you must not have epilepsy or use a pacemaker. You may experience sim-sickness from using the head-mounted display, but precautionary measures will be taken to ensure your well-being. Please sign and date this form.

[Ask subject if they would like a copy of the informed consent form.]

[Present the subject with the anthropometric data sheet.]

This form asks about your personal characteristics and will serve to verify your qualifications for the study. Please take a few moments to complete the survey. If you have any questions, I will be happy to address them. This form, like the informed consent form, will not be associated with any of the other survey forms used in this experiment.

[Present payment form.]

This is the payment form that will be used to calculate your compensation for participating in this experiment. Please fill out the information. Your Social Security number must be included on the form for tax purposes. The income you earn from this experiment is taxable and you should report it to the IRS.

INSTRUCTIONS

I will present all instructions to you orally. If you do not understand certain instructions, you will be able to ask questions before completion of each step in the procedure. You may also ask questions about the experiment during the familiarization, training, and rest periods. I need you to follow all instructions given before and during the testing carefully. If you fail to follow instructions or the equipment malfunctions, I will stop the testing procedure. You will then be allowed to read the task instructions and ask questions you may have, or the system malfunction will be corrected. Subsequently, we will resume testing.

EQUIPMENT FAMILIARIZATION

The equipment to be used in this experiment includes a high-performance graphics visualization workstation presenting the virtual environment and task. The system is integrated with a conventional 2-D mouse controller to cause motion and direction control of the rover in the virtual environment. The mouse is configured for motion in all directions. The system is also integrated with a conventional keyboard to control the robotic arm. Keys on the number keypad will be used to rotate joints of the arm, however; I will execute control commands at the keyboard for you. In the test trials, a HMD will be used to present the VE to you. I will explain the HMD to you in further detail later in the instructions.

[Check to see if the subject has any questions about the equipment or setup.]

HEARING TEST

[Open hearing test, CD tracks in RealPlayer.]

To ensure that you are qualified to participate in this experiment, we will conduct a quick test of hearing. This test only serves to verify your qualifications, and the results will not be used in conjunction with any other testing in this experiment. **[Hand subject hearing testing sheet and pencil.]** For this test, I will present sounds to you through the PC speakers. You will be asked to mark your responses to the questions on the sheet I have handed you.

For the first section, I will present two musical tones. You should mark on your sheet whether the second tone is higher or lower in pitch than the first tone. Do you have any questions?

[Play each track once. Do not repeat. Play CD tracks for tone 1 through tone 5. (Tracks 1-5)]

The next section will present two rhythms. You should mark on the sheet whether these rhythms are the same or different. Do you have any questions?

[Play CD tracks rhythm 1 through rhythm 5. (Tracks 6-10)]

The last section presents the same melody played twice. You should mark whether the two melodies are played with the same instrument or with a different instrument. Do you have any questions?

[Play CD tracks timbre 1 through timbre 4. (Tracks 11-14) When subject finishes, grade responses. Subject must have at least 10 of 14 correct overall or have a perfect score in one of the three sections.]

([Read if subject does not meet criteria.]

Thank you for coming today. This concludes your participation in the experiment. You will be compensated for the time you have spent here.

It is possible that you may experience simulator sickness when using the immersive VE displays including the HMD in this experiment. Therefore, procedures will be employed to assure your safety and well-being. Please inform me if, at any point in time you begin to experience motion sickness-like symptoms.

[Hand subject sim-sickness form and let them fill it out. Calculate sim-sickness score using "HelperSheet" tab in SimSickAndErrorSheets.xls in Excel. If scores exceed criteria, dismiss subject.]

In order to determine the possible presence of simulator sickness symptoms, the Simulator Sickness Questionnaire (SSQ) will be administered to you immediately before and after an experimental trial. If your score indicates that you may be suffering from sim-sickness, the questionnaire will also be administered at a 20-minute interval after a trial for up to 1 hour. If pre-exposure scores on the SSQ indicate that you are not currently in good health, you will not be permitted to participate. If post-exposure scores do not return to pre-exposure levels, you will be required to remain here for an additional 20 minutes at which time the SSQ measure will be administered again. This procedure will be repeated until the SSQ scores return to pre-test levels. If scores do not return to pre-test levels within 1 hour after an experiment, you will be advised not to drive a motor vehicle for 24 hours, and a ride will be provided to you. It will also be recommended that you seek medical counsel for motion sickness-like symptoms. This first sim-sickness form will be used as a baseline for comparison with your post-trial scores. Please fill out this form carefully. Do you have any questions?

TRAINING SESSION

[Make sure the mouse is on the left hand side of the computer, and the keyboard is on the right.]

As discussed, you will now complete a dedicated training session. The session will be divided into three major training periods. These periods are provided to ensure that you are able to use the vocal commands and mouse to complete the experimental tasks.

[Open WorldUp Player simulation. This simulation is presented in the clutter-free underwater environment. C:\simdata\Vango_HeatherTraining.wup Use the no clutter option ("None"). Hit Z to start.]

I will now provide you with instructions on how to complete the teleoperated mine neutralization task and you will practice this task. Your goal during the actual teleoperation task will be to locate mines in the virtual environment. To navigate through the environment, you will need to control the motion and direction of the rover. I will now explain to you how to navigate the rover in the virtual environment using the mouse controller. You will start from a pre-established position in the environment. You will use the mouse to facilitate motion of the rover in the forward, backward, left and right directions. An “arrow” cursor will be presented on the VE display corresponding to the motion of the mouse. Moving the cursor to the upper portion of the display will result in forward motion of the rover. Moving the cursor to the lower portion of the display will cause a reverse motion. Positioning the mouse in the center of the display will cause all motion in the virtual environment to stop. Moving the cursor to either the left or right halves of the display will cause the rover to turn in the corresponding direction.

[Demonstrate the motion of the rover in all directions. Allow subjects to navigate the rover in the environment.]

Four viewpoints will be available to you in the virtual environment. One viewpoint will be from a position several feet above the ocean floor and several feet behind the rover. This viewpoint will give you a view of the entire rover and some of the surrounding environment. We will refer to this view as the “behind rover” view.

The other three viewpoints will be from virtual cameras mounted on the rover. One viewpoint is from a camera that is located on the rear of the rover. This viewpoint is from several feet above the rover and allows you to view a portion of the rover and the robot arm, as well as some of the surrounding environment. This view will be referred to as the “on rover” view.

The second camera is located on the lower arm of the rover. This viewpoint will provide a close-up view of the tools on the arm and the mines to be neutralized. Only a small portion of the surrounding environment will be visible in this view. This view is referred to as the "rover arm" view.

The fourth viewpoint is the "sonar view". The sonar view provides a viewpoint from the sonar antenna. This viewpoint will be helpful in positioning the sonar. You may toggle between the 4 viewpoints by using the right mouse button on the mouse controller.

[Demonstrate use of the mouse button for toggling viewpoints. Show subject pictures of each viewpoints.]

The phases to complete in order to detonate mines is the following:

1. Search Phase: In this phase you will first, locate a mine, and then, drive to the mine.
2. Place Charge Phase: When you are within close proximity of the mine, you will pick up a charge, which is stored in a bin on the front of the rover, using a magnetic gripper tool. The magnetic gripper tool is located on the end of the rover arm. You will specify the magnetic gripper tool by saying, "SELECT MAGNET" to me. Once the tool is inside the box, you will activate this tool by clicking the left mouse button. Next, you will place the charge on the mine by lifting the arm of the rover and placing the charge on the mine by again clicking with the left mouse button.

[Switch to the Rover Arm view.]

3. Detonate Phase: Move back from the mine. You will now need to specify the sonar tool by saying, “SELECT SONAR” to me. The sonar is located on the sonar arm. Position the sonar arm to aim at the charge on the mine. Then, fire the sonar at the charge in order to detonate the charge by clicking the left mouse button after the sonar is aimed at the mine.

[Switch to the Sonar view.]

In addition to these steps, there are also two levels of automation under which the simulation may run. Manual mode means that you have complete control over the motion and actions of the rover. Supervisory mode means the computer has complete control over the motions and actions of the rover. During supervisory mode you will still be required to watch the actions of the rover so that you are aware of everything that is happening within the simulation. The mode may shift at any point in the simulation, and you will be notified by earcons, icons, vocal cues, or no cue as discussed in the next training session.

So to summarize, when a mine is located, you will pick up and place a charge on a mine using the magnetic gripper tool, and then you will detonate the mine using the sonar. You will need to manipulate both the rover arm and the sonar arm. The magnetic gripper and the sonar can be activated by pressing the left mouse button. You will use the sonar only after you have found the mine, and positioned the sonar to aim at the mine. The keyboard will be used to position the rover arm and the sonar arm. I will control the motion of the arms with the keyboard, and you will simply give me vocal commands to move the arms.

[Show the rover in the simulation and point to various parts of the rover arm as you go over the commands.]

The commands that you will use are as follows:

"UPPER ARM UP" – this command will be used to rotate the upper arm upward.

"UPPER ARM DOWN" – this command will be used to rotate the upper arm downward.

[8 and 2 keys move the upper arm.]

"LOWER ARM UP" - this command will be used to rotate the lower arm upward.

"LOWER ARM DOWN" - this command will be used to rotate the lower arm downward.

[7 and 1 keys move the lower arm.]

"SHOULDER RIGHT" - this command will be used to rotate the shoulder to the right.

"SHOULDER LEFT" - this command will be used to rotate the shoulder to the left.

[4 and 6 keys rotate the shoulder.]

"STOP" – this command will be used to stop motion of any component of the manipulator in any direction.

[Check to see if the subject has any questions.]

The first training period will allow you view the mine neutralization process in a supervisory mode. I will explain the procedures of the process as they occur. You will then be allowed to ask any questions you may have regarding the simulation before the next training period.

[Check to see if the subject has any questions.]

You will now detonate one mine manually in the underwater environment using the vocal commands and mouse movements learned.

[Allow subject to practice in neutralization of one mine. Total training time should be about 5 minutes. If subject is unable to "shoot" a mine, just go on to the next part. Use C:\simdata\Vango_HeatherTraining.wup Select the no clutter "None" condition. Use SSSMMM automation sequence.]

You have completed the first training session. Please take a short break before we begin the second training period.

[Following training, allow about 5 minutes for a break.]

EARCON TRAINING

I will now introduce you to the auditory icons that will be used in the simulation to alert you to changes in modes of automation and phase changes. During the simulation the sounds will be presented to you through the headphones of the HMD and will occur directly before a change happens.

[Point to the headphones.]

There are 12 possible sounds that may be played. Each sound corresponds with a specific piece of information: a change to manual mode, a change to supervisory mode, a shift to the search phase, a shift to the place charge phase, a shift to the detonate phase, (steps) locate a mine, drive to a mine, pick charge up from bin, place charge on mine, move away from mine, position sonar, and fire sonar at charge. I will play each of these earcons for you while explaining the information that each one conveys.

[Play tracks 15-27. Refer to CD cover for associated names. Play each earcon and tell the subject what each earcon means. Change volume to the top blue line on the speakers. Play mode changes first. Then, phases. Next, play phases and steps.]

If both the mode and phase of the simulation are about to change simultaneously, you may be presented with two earcons back-to-back. For instance, if the simulation is moving from manual search phase to supervisory place charge phase, earcons that represent the place charge phase and supervisory mode will be played. These are referred to as compound earcons. The phase change (search, place charge, or detonate) will always be presented first with the mode change (manual or supervisory) following. If as many as three earcons are played in conjunction the order will be phase change, step within phase, and mode change. The mode changes are played with the same instrument. All main phase changes cues (search, place charge, and detonate) use 3 major beats, and for each additional step within a phase an additional beat is used. I will now play each of the possible compound earcons for you and go over their meanings.

[Play tracks 28-35. Refer to CD cover for associated names. Play each earcon and tell the subject what each earcon means.]

You will now be allowed to study the earcons on your own for 10 minutes. Knowledge and understanding of these earcons are essential to the successful completion of the demining task. Therefore you are encouraged to thoroughly learn these earcons. You may use the mouse and keyboard (the up/down arrows and the space bar) to play the earcons you wish to hear. The name corresponding with the earcon is shown in the Real Jukebox display.
[Make sure both playlists are visible on the screen. Demonstrate using the mouse and keyboard to play earcons. Wait for 10 minutes while user explores sounds or until subject is finished.]

[Allow a short break while you set up the earcon test. Use end tracks on CD: Test1 (Tracks 36-40), Test2 (Tracks 41-45), or Test3 (tracks 46-50).]

This is a quick quiz to make sure you have successfully learned the earcons. **[Hand subject earcon test sheet and pen.]** I will present five earcons to you. These earcons may be simple, or composed of one "message", or compound, meaning they convey more than one message. Choose what message the earcon tells you from the list at the top of the page, and write the corresponding number in the appropriate blank.
[Play each track once. Do not repeat. Present earcons one at a time, waiting for subject to mark response before moving to next earcon. When finished, make sure subject answered 4 of 5 correctly.]

([If subject did not meet proficiency, repeat the process of playing earcons and allowing 5 minutes of study. Give them a different TestX playlist. If they again do not get 4/5 correct, dismiss them from the study.])

ICON TRAINING

In addition to using earcons to alert you to upcoming changes in the simulation, visual icons will also be used during some trials. These icons, like the earcons, each convey a specific message that corresponds to an upcoming change in the level of automation. There is an icon for each possible change, including: change to manual mode, change to supervisory mode, shift to search phase, shift to place charge phase, shift to detonate phase. There are also icons for each step of the task, including locate a mine, drive to a mine, pick charge up from bin, place charge on mine, move away from mine, position sonar, and fire sonar at charge. I will now show you the icons that will be used.

[Show subject icons, and explain each icon.]

As with the earcons, icons may also be combined back-to-back when level of automation and task phase changes are about to occur.

You will now be allowed a few minutes to study these pictures. With respect to the second set of icons, this figure represents the charge **[point to stick of dynamite]**. The hand pointing to the charge represents placement of the charge on the mine. Do you have any questions? **[Allow user up to 5 minutes to study the icon pictures.]**

[Test subject using flash cards…flip through pictures and have subject identify what each picture means. If they cannot perform this satisfactorily, let them study the pictures more. Test using 5 randomly selected flash cards.] [If user cannot learn icons, dismiss them from the study.]

[Allow a 2-min break, and then start the last training period.]

UNDERWATER ENVIRONMENT TRAINING

[Have the subject turn toward the SGI. Open up Worldup Player with the training simulation. Open up PowerPoint. Open up the environment slides.]

This is the underwater environment in which we will be conducting the actual experiment. There are many types of objects within the environment, such as rocks, plants, fish, subs, etc.

[Show subject pictures of the environment. (First 3 slides of Environment and Mines.ppt)]

There are four different mine types in the environment. Two mine types represent Type 1 mines, and two mine types represent Type 2 mines. You may be required to detonate either Type 1 or Type 2 mines, so it is important that you can classify them. Charges are located in different compartments on the front of the rover.

[Show the pictures of the mines and their associated charge. Shift to WorldUp.]

Remember, the viewpoints available are “Rover Arm,” which looks down the arm of the rover; “Sonar,” which looks out from the sonar dish; “On Rover,” which shows the view from the rover itself; and “Behind Rover,” which shows the complete rover from behind. **[Allow user to toggle through the viewpoints.]**

HMD

You will use the HMD to explore the underwater virtual environment. The HMD will allow you to see and hear directly through the helmet.

You will now be allowed to practice the detonation process in this environment using the HMD. You may complete the detonation of one mine. For this practice trial, you will be presented with both icons and earcons. However, the experimental trials will only expose you to one cue type: earcon, icon, vocal cue, or no cue.

[Open underwater training simulation. Set for trial session, all manual mode. Help subject adjust HMD. Allow subject to detonate one mine. After one mine is detonated manually, have subject remove HMD. Reset the training simulation to run in all supervisory. Have subject put HMD back on, and let them watch the simulation detonate one mine under supervisory mode. Make sure to remind subjects of IPD and focal distance adjustments. Main volume should be between the 1st and 2nd tick mark

on the computer's volume control panel. Wave volume should be at the middle tick mark.]

Now, I'm going to describe SA and workload measures you're going to complete after each trial. Situation awareness can be described as perception of elements in an environment, including their current and future status. Situation awareness will be measured after your task performance in this study. The measure of situation awareness to be used for this study will require you to subjectively rate levels of attentional supply and demand and your knowledge of the task. This training period is provided to ensure that you are aware of how the situation awareness measure is administered and how to respond using a rating form.

The Situation Awareness Rating Technique (SART) will be used to measure situation awareness. This technique requires that you complete a form at the close of each experimental trial.

[Show subject SART form. Explain terms.]

The form presents three scales on attentional demand, supply, and understanding. You will mark on the scales where you believe your level of the particular variable is (be it attentional demand, attentional supply, and understanding). The terms are defined on the SART form.

[Check to see if the subject has any questions.]

NASA-TLX RATINGS

In order to assess the task workload that you experience during experimental testing, you will complete a subjective comparison of various mental and physical demand factors. **[Hand subject the NASA-TLX demand comparison form. Enter the correct subject number, trial number.]** A sheet of descriptions of each of the demand factors will be provided to you. **[Give subject NASA-TLX factor description sheet.]** Please take the time to complete the workload comparison form by referring to this sheet for the definitions of various demands. On each line of the demand comparison form, circle the demand that you believe to be more important to performance in the mine neutralization task.

You will also rate the demand factors upon the completion of task performance during experimental testing. Both your comparisons and ratings of these factors will be used to compute a composite score of workload for the task.

At the end of each test trial, you will be required to complete subjective ratings of perceived workload. You will rate task workload using the form presented to you. **[Show subject a hard copy of NASA-TLX.]** You will complete the NASA-TLX form by indicating the level of demand you experienced during the task by marking an line on the scale directly below each of the factors. A sheet of descriptions of each of the demand factors to be rated will be provided. **[Give subjects NASA-TLX factor descriptions sheet.]** Please make reference to this sheet when rating the various demands. **[5 minute break]**

EXPERIMENTAL TESTING

Now, you will complete 2 test trials in the task that you just trained on. The experimental procedures are as follows:

(1) Participation in the first trial;

(2) Completion of the SART form;

(3) Completion of the NASA-TLX;

(4) Completion of the sim-sickness questionnaire; and

(5) A short break.

This sequence will then be repeated 1 time.

[Check to see if the subject has any questions.]

You will now begin your first experimental trial. Your task is to navigate the rover in the environment to locate and neutralize mines. Please recall that the mode of automation may shift from manual to supervisory control throughout the test trial. You should pay attention to the interface to maintain your awareness to these changes. You should neutralize a total of two mines in the environment.

Due to the nature of the experiment, keeping your attention focused on completing the task is important. I ask that you now remove any watch or timepiece from your person and place it in a location out of sight. I also ask that you refrain from talking during the testing periods. If you have any difficulties, however, please do not hesitate to bring them to my attention and I will assist you.

[Make subjects look away from the interface.]

<u>Experimental Procedure</u>

[Set up the trial.]

1. **Help subject put on HMD and make sure it is adjusted correctly.**

[Ask subject if they are ready to proceed. Click on the appropriate simulation and enter subject data. Help subject put on HMD. Start simulation by pressing "z" on the keyboard.]

[When subject has completed a total of 2 mines, stop task performance. Help subject take off HMD. Close both windows by pressing "q" on the keyboard.]

[Deliver SART, NASA-TLX and Sim-sickness questionnaire.

1. **Deliver SART**

   Please rate the level of each of the SART variables you feel you experienced during the trial at this time.

2. **Deliver NASA-TLX.**
3. **Deliver sim-sickness questionnaire**
4. **Calculate sim-sickness score.**
5. **If sim-sickness score=pretest score, skip step 5.**
6. **If sim-sickness score is higher than pre-test score, subject MUST wait for 20 minutes.**
   a. **After 20 minutes, give sim-sickness test again.**

b. (Go to step 4. This sequence may be repeated for an hour. If scores are not back to normal after an hour, dismiss subject.)]

You will now have a short break before your next experimental session.

[Allow subject a 5-min break. Set up next trial.]

You will now complete your 2^{nd} experimental trial. The method of navigating the rover and manipulating the robot arm is the same. The task will be the same; however, your cueing condition may be different. Following this trial, as in the previous test, you will complete the SART, NASA-TLX and Sim-sickness questionnaire.

This concludes your participation in the experiment. As part of the test trials, you were exposed to _______ conditions. [Explain purpose.] Other subjects will be or have been exposed to __________. Thank you for your time. Do you have any questions?

[Give subjects payment form and directions on how to obtain payment.]

APPENDIX 2: INFORMED CONSENT FORM

Informed Consent Form
Department of Industrial Engineering
North Carolina State University

Auditory Cueing Effects on Human Performance with Adaptive Systems

---

I hereby give my consent for voluntary participation in the research project titled, "Auditory Cueing Effects on Human Performance with Adaptive Systems." I understand that the person responsible for this project is Heather Warren, under the advisement of Dr. David B. Kaber, who can be telephoned at (919) 515-3086. She has explained to me the study objective of assessing the effect of the method of remote control of a simulated robotic rover on my performance in a mine disposal task, including manual and supervisory control. She has also explained to me the objective of examining the effectiveness of visual and auditory cues for informing me of changes in the method of control during my performance. She has agreed to answer any inquiries I may have concerning the procedures of the research and has informed me of my right to refuse to answer any specific questions asked of me. She has also informed me that I may contact the North Carolina State University (NCSU), Institutional Review Board for the Protection of Human Subjects by writing them in care of Dr. Matt Zingraff, Chair of IRB, Research Administration, NCSU, 1 Leazar Hall, Box 7514, Raleigh, NC 27695, or by calling (919) 515-2444.

The researchers for the study have explained to me the procedures to be followed and the potential risks and discomforts. In summary, the procedures include: (1) an equipment familiarization period; (2) a hearing test, (3) a sim-sickness questionnaire; (4) an extensive training session to learn and practice the commands and maneuvers of the robotic system; (5) training in the presentation of auditory and visual cues as part of the simulation; (6) an explanation of the HMD; (7) a practice session involving an underwater demining task; (8) familiarization with the Situational Awareness Rating Technique (SART); (9) familiarization with the NASA-Task Load Index (TLX), a workload measurement technique; (10) two test trials in the underwater virtual environment using an HMD; (11) and a debriefing on the study. All training and testing will be conducted during a single experimental session that will require approximately 3 hours of my time.

The risks have also been explained to me as follows: (1) a potential exists for visual strain and/or fatigue in viewing the virtual environment displays through immersive displays including the HMD and desktop VR display; (2) a potential exists for soreness of the hand and forearm muscles from extensive use of a mouse in controlling the simulated telerover. These risks are not substantially different from those associated with my everyday PC use. In the event that I experience fatigue or discomfort, I will inform the experimenters immediately. In addition, I will be tested for motion sickness symptoms before and after the experiment. I understand that if the symptoms have not gone away after 1 hour, I will be advised not to drive a car for 24 hours and a ride will be provided.

The researchers for the study have the right to terminate my participation if I experience discomfort or fatigue, or I am not cooperative.

I understand that if this research project results in any physical or mental harm to me, treatment is not necessarily available at the NCSU, Student Health Services, nor is there necessarily any insurance carried by the University or its personnel applicable to cover any such injury. Financial compensation for any such injury must be provided through my own insurance program. Further information about these matters may be obtained from the Institutional Review Board at (919) 515-2444, 1 Leazar Hall, NCSU Campus.

I understand that I will not derive any therapeutic treatment from my participation in this study. I understand that I may discontinue my participation in this study at any time without prejudice. I understand that all data will be kept confidential and that my name will not be used in any reports, written or unwritten. I have received a copy of this consent form for my personal records.

Signature of Subject: **Date:**

______________________________ ______________________________

Signature of Authorized Representative: **Signature of Witness to Oral Presentation:**

______________________________ ______________________________

APPENDIX 3: ANTHROPOMETRIC DATA SHEET

SUBJECT SURVEY

Auditory Cueing Effects on Human Performance with Adaptive Systems

The intended purpose of this form is to establish a subject profile based on volunteered anthropometric data. Please complete the sheet to the best of your knowledge following the example formats indicated in the parentheses adjacent to each data field label.

Age (XX-yr.): _________ Gender (M/F): ____ Handedness (Left/Right): ____________

Corrected Visual Acuity: Left Eye (XX/XX): _______ Right Eye (XX/XX): _________

	1	2	3	4	5
Video Game Experience:	1 None	2	3 Occasional	4	5 Frequent
PC Experience:	1 None	2	3 Occasional	4	5 Frequent
VR Experience:	1 None	2	3 Occasional	4	5 Frequent

Do not write below this line. Experimenter use only.

Subject # _________ Cue: E I V N

APPENDIX 4: PAYMENT FORM

PAYMENT FORM FOR SUBJECT

AUDITORY CUEING EFFECTS ON HUMAN PERFORMANCE WITH ADAPTIVE SYSTEMS

Name: ____
First Last

Address: ____
Street Address Apartment Number

____________________ ____________________ ____________________
City State Zip

Social Security Number (XXX-XX-XXXX): ________________________________

Date (XX/XX/01): ____

Start Time (XX:XX): ____

Finish Time (XX:XX): ____

Total Time (XX:XX): ____________________ x $7.50 = ____________________

APPENDIX 5: SSQ

Simulator Sickness Questionnaire

Department of Industrial Engineering
North Carolina State University

Auditory Cueing Effects on Human Performance with Adaptive Systems

Instructions: Circle the items that apply to you RIGHT NOW.

SYMPTOM	RATING			
General Discomfort	None	Slight	Moderate	Severe
Fatigue	None	Slight	Moderate	Severe
Headache	None	Slight	Moderate	Severe
Eye Strain	None	Slight	Moderate	Severe
Difficulty Focusing	None	Slight	Moderate	Severe
Increased Salivation	None	Slight	Moderate	Severe
Sweating	None	Slight	Moderate	Severe
Nausea	None	Slight	Moderate	Severe
Difficulty Concentrating	None	Slight	Moderate	Severe
“Fullness of the Head”	None	Slight	Moderate	Severe
Blurred Vision	None	Slight	Moderate	Severe
Dizzy (eyes open)	None	Slight	Moderate	Severe
Dizzy (eyes closed)	None	Slight	Moderate	Severe
Vertigo	None	Slight	Moderate	Severe
Stomach Awareness*	None	Slight	Moderate	Severe
Burping	None	Slight	Moderate	Severe
Other. Please describe.				

“Stomach Awareness” is usually used to indicate a feeling of discomfort, which is just short of nausea

Do not write below this line. Experimenter use only.

Subject # _________ **Trial: B 1 2** **Cue: E I V N**

APPENDIX 6: SART FORM

SITUATION AWARENESS RATING TECHNIQUE (SART) FORM

The purpose of this form is to evaluate situation awareness. Situation awareness can be defined as the perception of elements in environment, comprehension of meaning, and projection of future states within a volume of time and space (Endsley, 1988).

SART DEMAND includes any factor that requires a demand of your attention, or that is associated with complexity, variability, or instability in the simulation. (Example: How taxed do you feel your attention was during the last test trial?)

SART SUPPLY includes any factor that requires a supply of your attention, or that is associated with arousal, concentration, division of attention, or spare mental capacity in the simulation. (Example: If you had to allocate attention to another task while still performing the demining task, how much attention do you think you could devote to the new task?)

SART UNDERSTANDING includes any factor that requires understanding, or that is supported by prior knowledge, information feedback, or interaction between you and the system. (Example: How would you rate your overall understanding of the demining task?)

Please rate your level of each variable below by marking a vertical line on the scale where you feel your level of the particular variable (SART DEMAND, SART SUPPLY, or SART UNDERSTANDING) was during the last trial.

SART DEMAND

|————————————————————|

Low High

SART SUPPLY

|————————————————————|

Low High

SART UNDERSTANDING

|————————————————————|

Low High

Do not write below this line. Experimenter use only.

Subject # _________ **Trial: 1 2** **Cue: E I V N**

APPENDIX 7: NASA-TLX

NASA-TLX Workload Factor Definitions

Mental Demand

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand

How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand

How much time pressure did you feel due to the rate at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance?

Frustration

How insecure, discouraged, irritated, and annoyed versus secure, gratified, content and complacent did you feel during the task?

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Subjective Comparison of Demand Factors: NASA-TLX Survey

Auditory Cueing Effects on Human Performance with Adaptive Systems

---

Indicate the demand of greater importance by circling its label on each line directly below.

Mental Demand / Physical Demand

Mental Demand / Temporal Demand

Mental Demand / Performance

Mental Demand / Effort

Mental Demand / Frustration

Physical Demand / Temporal Demand

Physical Demand / Performance

Physical Demand / Effort

Physical Demand / Frustration

Temporal Demand / Performance

Temporal Demand / Frustration

Temporal Demand / Effort

Performance / Frustration

Performance / Effort

Frustration / Effort

Do not write below this line. Experimenter use only.

Subject # _________ **Cue: E I V N**

Subjective Rating of Perceived Workload: NASA-TLX Survey

Auditory Cueing Effects on Human Performance with Adaptive Systems

---

Indicate the level of demand experienced during the navigation task for each of these factors by drawing a straight vertical line on the scale directly below.

Mental Demand

Low High

Physical Demand

Low High

Temporal Demand

Low High

Performance

Low High

Frustration

Low High

Effort

Low High

Do not write below this line. Experimenter use only.

Subject # __________ Trial: 1 2 Cue: E I V N

APPENDIX 8: HEARING TEST

HEARING TEST

Auditory Cueing Effects on Human Performance with Adaptive Systems

Mark the appropriate response to the sentence:
The second tone is (higher/lower) than the first tone.

Set 1 higher lower

Set 2 higher lower

Set 3 higher lower

Set 4 higher lower

Set 5 higher lower

Score: /5

Mark the appropriate response to the sentence:
The rhythms are the same. (yes/no)

Set 1 yes no

Set 2 yes no

Set 3 yes no

Set 4 yes no

Set 5 yes no

Score: /5

Mark the appropriate response to the sentence:
The two series of notes were played with the same instrument. (yes/no)

Set 1 yes no

Set 2 yes no

Set 3 yes no

Set 4 yes no

Score: /4

Minimum acceptable score 10/14 overall or a perfect score on one section.
Subject #: ________

APPENDIX 9: EARCON COMPREHENSION TEST

EARCON COMPREHENSION TEST

Auditory Cueing Effects on Human Performance with Adaptive Systems

After hearing each earcon, identify what information it conveys.
Choose from the following:

Single Earcons

1. Manual Mode
2. Supervisory Mode
3. Search Phase
4. Locate Mine Step
5. Drive to Mine Step
6. Place Charge Phase
7. Pick up Charge Step
8. Place Charge on Mine Step
9. Detonate Phase
10. Move Away from Mine Step
11. Position Sonar Step
12. Fire Sonar at Charge Step

Two Earcons

13. Search Phase, Locate Mine Step
14. Place Charge Phase, Pick up Charge Step
15. Detonate Phase, Move Away from Mine Step

Three Earcons

16. Search Phase, Locate Mine Step, Manual Mode
17. Search Phase, Locate Mine Step, Supervisory Mode
18. Place Charge Phase, Pick up Charge Step, Manual Mode
19. Place Charge Phase, Pick up Charge Step, Supervisory Mode
20. Detonate Phase, Move Away from Mine Step, Manual Mode
21. Detonate Phase, Move Away from Mine Step, Supervisory Mode

Mark the number corresponding with the information presented.

Earcon 1 __________ Earcon 2 __________ Earcon 3 __________
Earcon 4 __________ Earcon 5 __________

Score: /5

Do not write below this line. Experimenter use only.

Subject # __________ Cue: E I V N

APPENDIX 10: SAS CODE

```
/*SAS Code template for PROC GLM and residual analyses.*/


Options linesize=80 pagesize=50 nodate pageno=1;
data hlw_ct_cc;

input sub_num trial_num cue_type cue_comp ttc sart nasa_tlx mental perf
effort sd ss su
;
Label
      sub_num = 'Subject Number'
      trial_num = 'Trial Number'
      cue_type = 'Auditory/Visual Cue Type'
      cue_comp = 'Complexity of Cueing'
      ttc = 'Time-to-task Completion'
      sart = 'Composite SART Rating'
      nasa_tlx = 'Overall Workload Score'
      mental = 'Mental Component of TLX Score'
      perf = 'Performance Component of TLX Score'
      effort = 'Effort Component of TLX Score'
      sd = 'Demand Component of SART'
      ss = 'Supply Component of SART'
      su = 'Understanding Component of SART'
      ;

/*In the case that you observe a significant interaction of CT*CC, you
will need to calculate a new variable based on the data in order to
represent the interaction as a main effect and conduct a means breakout on
the levels of the new variable.*/

CTCC = (cue_type*10)+ cue_comp;
Label
      CTCC = 'New Variable Calculated Based on Data Representing
Interaction as Main effect';

/*Cue type can be described by a number 1-4:  Vocal=1, Earcon=2, Icon=3,
No Cue=4.  Cue complexity can be described by a number 1 or 2:  Simple=1,
Complex=2*/

/* Outliers have been removed from this data set for TTC.
S=10, T=1, TTC=961, RTTC=488
S=10, T=2, TTC=927, RTTC=505
S=25, T=1, TTC=960, RTTC=506

Outliers have been removed from this data set for SD.
S=17, T=1, SD=0.5, RSD=6.32
S=17, T=2, SD=0.5, RSD=6.44

Outliers have been removed from this data set for SS.
S=6, T=1, SS=8.3, RSS=3.85
```

```
S=25, T=2, SS=1.55, RSS=4.93

Outliers have been removed from this data set for SU.
S=11, T=1, SU=2.75, RSU=8.09
S=11, T=2, SU=2.45, RSU=7.86
*/


cards;
1      1      1      1      387    7      66     20     19     12     3.4    5.4   5
1      2      1      2      392    6.1    65     18     17     15     5.7    5
       6.8
2      1      1      1      360    2.9    55     25     3      18     6.1    2.85
       6.15
2      2      1      2      380    10.35  46     16     9      14     4.35   6.85
       7.85
3      1      1      1      486    8.1    57     4      20     18     4.8    3
       9.9
3      2      1      2      366    13.55  50     7      20     16     2.75   6.35
       9.95
4      1      1      1      350    2.85   73     31     15     18     9.1    4.75
       7.2
4      2      1      2      436    6.2    64     24     15     13     7.2    6.15
       7.25
5      1      1      1      341    9.85   68     8      24     10     3.35   6
       7.2
5      2      1      2      408    13.65  61     8      23     7      2.5    7.4
       8.75
6      1      1      1      328    16.15  38     1      10     10     0.55   3.85
       8.4
6      2      1      2      418    16.2   32     1      7      9      0.3    7.75
       8.75
7      1      1      1      382    2.2    72     31     16     21     8.15   0.5
       9.85
7      2      1      2      402    6.65   66     24     16     21     5.75   2.45
       9.95
8      1      1      1      352    6.3    18     15     1      1      0.8    0
       7.1
8      2      1      2      454    5.85   24     20     1      2      2.5    1.15
       7.2
9      1      2      1      358    2.15   66     17     24     17     7.6    1.15
       8.6
9      2      2      2      413    3.35   63     16     21     17     8.75   2.2
       9.9
10     1      2      1      488    4.7    39     13     9      10     7.35   3
       9.05
10     2      2      2      505    6.6    27     9      9      5      5.85   3.55
       8.9
11     1      2      1      607    2.3    45     18     5      16     6.5    6.05
       8.09
11     2      2      2      508    1.9    55     20     9      17     6.2    5.65
       7.86
12     1      2      1      398    2.25   43     10     17     6      7.4    1.9
       7.75
12     2      2      2      355    6.35   38     12     17     4      5.9    4.15
       8.1
```

13	1	2	1	402	11.45	46	3	4	13	5.05	7.1	9.4
13	2	2	2	407	13.25	45	3	13	11	3.5	7.6	9.15
14	1	2	1	473	5.7	62	30	22	10	10	5.7	10
14	2	2	2	503	5.6	72	34	22	16	10	5.6	10
15	1	2	1	306	8.1	65	12	29	9	5.5	4.6	9
15	2	2	2	475	7.9	63	14	23	11	5.05	5.6	7.35
16	1	2	1	401	10.25	66	18	14	13	2.45	4.55	8.15
16	2	2	2	455	5.9	61	14	15	11	6.15	5.05	7
17	1	3	1	698	11.75	42	16	20	4	6.32	4.9	7.35
17	2	3	2	560	9.4	41	1	26	13	6.44	0.45	9.45
18	1	3	1	369	3.8	60	9	16	18	5.9	3.1	6.6
18	2	3	2	461	4.55	61	11	13	17	7.05	5.2	6.4
19	1	3	1	307	3.75	60	4	26	15	7.45	1.45	9.75
19	2	3	2	330	3.3	58	9	26	7	7.15	1.45	9
20	1	3	1	710	-0.35	70	25	25	15	9.75	0.4	9
20	2	3	2	615	-2.5	60	25	19	15	9.7	0.2	7
21	1	3	1	812	12.55	58	12	21	13	4.7	7.25	10
21	2	3	2	417	7	50	10	17	10	6.25	3.95	9.3
22	1	3	1	462	8.05	88	31	26	3	9.6	7.75	9.9
22	2	3	2	422	11.2	79	30	26	3	6.75	8.25	9.7
23	1	3	1	437	3.4	60	17	22	0	5.7	4.25	4.85
23	2	3	2	394	4.95	59	16	23	0	7.35	6.35	5.95
24	1	3	1	755	6.4	60	15	17	13	6.95	6.25	7.1
24	2	3	2	685	7.3	59	16	14	13	6.8	6.05	8.05
25	1	4	1	506	4.05	38	24	9	5	4.5	2.4	6.15
25	2	4	2	845	1.35	25	15	6	4	5.1	4.93	4.9
26	1	4	1	472	15.45	38	24	9	5	1.15	7.55	9.05
26	2	4	2	528	11.95	25	16	6	4	1.4	9.25	4.1
27	1	4	1	423	5.55	64	20	5	22	7.45	6.15	6.85
27	2	4	2	835	-3.55	78	24	2	23	9.15	1.55	4.05

```
28      1       4       1       322     5.3     63      23      14      13      7.3     2.6
        10
28      2       4       2       580     6.75    72      24      17      18      7.05    4.3
        9.5
29      1       4       1       336     12.9    20      13      3       3       3.95    6.85    10
29      2       4       2       649     13.6    18      9       3       3       3       6.65
        9.95
30      1       4       1       573     12.5    31      4       24      0       1.95    7.25
        7.2
30      2       4       2       499     13.1    42      9       24      6       3.9     7.75
        9.25
31      1       4       1       528     -2.5    53      8       10      15      8.35    0.85    5
31      2       4       2       365     7.05    48      9       18      6       5.75    5.7
        7.1
32      1       4       1       430     -5.05 85      33      20      25      10      0
        4.95
32      2       4       2       551     -5.15 69      33      15      13      10      0.1
        4.75
;

/*This procedure will produce main effects plots in order for you to look
at the trends of your observations.*/

proc print;
proc plot data=hlw_ct_cc;
plot ttc*cue_type = '*';
plot sart*cue_type = '*';
plot nasa_tlx*cue_type = '*';
plot ttc*cue_comp = '*';
plot sart*cue_comp = '*';
plot nasa_tlx*cue_comp = '*';
Plot mental*cue_type = '*';
Plot mental*cue_comp = '*';
Plot perf*cue_type = '*';
Plot perf*cue_comp = '*';
Plot effort*cue_type = '*';
Plot effort*cue_comp = '*';
Plot sd*cue_type = '*';
Plot sd*cue_comp = '*';
Plot ss*cue_type = '*';
Plot ss*cue_comp = '*';
Plot su*cue_type = '*';
Plot su*cue_comp = '*';
Title3 'Response/predictor plot';
Title5;

/*The following are procedures to analyze your performance data.*/

proc glm;
Title3;
Title5;
class cue_type cue_comp CTCC;

/* CTCC */

model ttc = cue_type sub_num(cue_type) cue_comp CTCC/P;
```

```
output out=hlw_ct_cc_ttc P=pttc R=rttc cookd=cookd_ttc;

/* Here you include the predicted values, residuals and cook's D stat (for
ttc) in your data set.*/

test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

/*This procedure will output the mean values for the levels of the
interaction. If the interaction is significant, you can use this output to
produce interaction plots in Excel.*/

means CTCC/duncan;

proc plot data=hlw_ct_cc_ttc;
plot rttc*(pttc cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_ttc plot normal;
var rttc;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

/* The proc univariate generates a basic normality plot. */

/* Plotting cook's D stat by subject to identify outliers. */

proc plot data = hlw_ct_cc_ttc;
plot cookd_ttc*sub_num;
Title3 'Plot of Cooks D';
Title5;

/* This is proc reg procedure to generate influence stats. */

proc reg;
Title3;
Title5;
model ttc = cue_type cue_comp /influence;
output out = hlw_ct_cc_ttc_reg DFFITS = DF_ttc;

proc plot data=hlw_ct_cc_ttc_reg;
plot ttc*DF_ttc = '*';
Title3 'Plot of DFFITS against time-to-task completion';
Title5;

proc plot data = hlw_ct_cc_ttc_reg;
plot sub_num*DF_ttc = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

/*The following are procedures to analyze your SART data.  They are
identical to the procedures for analyzing your performance data, except
the response measure is different.*/
```

```
proc glm;
Title3;
Title5;
class cue_type cue_comp;
model sart = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;
output out=hlw_ct_cc_sart P=psart R=rsart cookd=cookd_sart;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_sart;
plot rsart*(psart cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_sart plot normal;
var rsart;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

/* Plotting cook's D stat by subject to identify outliers. */

proc plot data = hlw_ct_cc_sart;
plot cookd_sart*sub_num;
Title3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model sart = cue_type cue_comp /influence;
output out = hlw_ct_cc_sart_reg DFFITS = DF_sart;

proc plot data=hlw_ct_cc_sart_reg;
plot sart*DF_sart = '*';
Title3 'Plot of DFFITS against sart';
Title5;

proc plot data = hlw_ct_cc_sart_reg;
plot sub_num*DF_sart = '*';
Title3 'Plot of DFFITS values against subject numbers';
Title5;

/* New analyses on SART dimensions.  SART dimensions are Demand, Supply,
and Understanding. */

/* SART Demand */

proc glm;
Title3;
Title5;
class cue_type cue_comp;
model sd = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;
output out=hlw_ct_cc_sd P=psd R=rsd cookd=cookd_sd;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
```

```
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_sd;
plot rsd*(psd cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_sd plot normal;
var rsd;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

proc plot data = hlw_ct_cc_sd;
plot cookd_sd*sub_num;
Title3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model sd = cue_type cue_comp /influence;
output out = hlw_ct_cc_sd_reg DFFITS = DF_sd;

proc plot data=hlw_ct_cc_sd_reg;
plot sd*DF_sd = '*';
Title3 'Plot of DFFITS against Demand component of SART';
Title5;

proc plot data = hlw_ct_cc_sd_reg;
plot sub_num*DF_sd = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

/* SART Supply */

proc glm;
Title3;
Title5;
class cue_type cue_comp;
model ss = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;
output out=hlw_ct_cc_ss P=pss R=rss cookd=cookd_ss;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_ss;
plot rss*(pss cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_ss plot normal;
var rss;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';
```

```
proc plot data = hlw_ct_cc_ss;
plot cookd_ss*sub_num;
TItle3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model ss = cue_type cue_comp /influence;
output out = hlw_ct_cc_ss_reg DFFITS = DF_ss;

proc plot data=hlw_ct_cc_ss_reg;
plot ss*DF_ss = '*';
Title3 'Plot of DFFITS against sart';
Title5;

proc plot data = hlw_ct_cc_ss_reg;
plot sub_num*DF_ss = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

/* SART Understanding */

proc glm;
Title3;
Title5;
class cue_type cue_comp;
model su = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;
output out=hlw_ct_cc_su P=psu R=rsu cookd=cookd_su;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_su;
plot rsu*(psu cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_su plot normal;
var rsu;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

proc plot data = hlw_ct_cc_su;
plot cookd_su*sub_num;
TItle3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model su = cue_type cue_comp /influence;
output out = hlw_ct_cc_su_reg DFFITS = DF_su;

proc plot data=hlw_ct_cc_su_reg;
plot su*DF_su = '*';
```

```
Title3 'Plot of DFFITS against Understanding Component of SART';
Title5;

proc plot data = hlw_ct_cc_su_reg;
plot sub_num*DF_su = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

/*The following are procedures to analyze your NASA-TLX data.  They are
identical to the procedures for analyzing your performance and SART data,
except the response measure is different.*/

proc glm;
Title3;
Title5;
class cue_type cue_comp;
model nasa_tlx = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;

output out=hlw_ct_cc_tlx P=ptlx R=rtlx cookd=cookd_nasa_tlx;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_tlx;
plot rtlx*(ptlx cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_tlx plot normal;
var rtlx;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

proc plot data = hlw_ct_cc_tlx;
plot cookd_nasa_tlx*sub_num;
Title3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model nasa_tlx = cue_type cue_comp /influence;
output out = hlw_ct_cc_nasa_tlx_reg DFFITS = DF_nasa_tlx;

proc plot data=hlw_ct_cc_nasa_tlx_reg;
plot nasa_tlx*DF_nasa_tlx = '*';
Title3 'Plot of DFFITS against NASA TLX scores';
TItle5;

proc plot data = hlw_ct_cc_nasa_tlx_reg;
plot sub_num*DF_nasa_tlx = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

/*Output will now be produced for the three demand components of the NASA-
TLX that had the highest mental, performance, and effort.*/
```

```
proc glm;
Title3;
Title5;
class cue_type cue_comp;
model mental = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;

output out=hlw_ct_cc_mental P=pmental R=rmental cookd=cookd_mental;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_mental;
plot rmental*(pmental cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_mental plot normal;
var rmental;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

proc plot data = hlw_ct_cc_mental;
plot cookd_mental*sub_num;
TItle3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model mental = cue_type cue_comp /influence;
output out = hlw_ct_cc_mental_reg DFFITS = DF_mental;

proc plot data=hlw_ct_cc_mental_reg;
plot mental*DF_mental = '*';
Title3 'Plot of DFFITS against NASA TLX scores';
Title5;

proc plot data = hlw_ct_cc_mental_reg;
plot sub_num*DF_mental = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

proc glm;
Title3;
Title5;
class cue_type cue_comp;
model perf = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;
output out=hlw_ct_cc_perf P=pperf R=rperf cookd=cookd_perf;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_perf;
plot rperf*(pperf cue_type cue_comp) = '*' /vref=0;
```

```
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_perf plot normal;
var rperf;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

proc plot data = hlw_ct_cc_perf;
plot cookd_perf*sub_num;
TItle3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
Title5;
model perf = cue_type cue_comp /influence;
output out = hlw_ct_cc_perf_reg DFFITS = DF_perf;

proc plot data=hlw_ct_cc_perf_reg;
plot perf*DF_perf = '*';
Title3 'Plot of DFFITS against Performance Component NASA-TLX scores';
TItle5;

proc plot data = hlw_ct_cc_perf_reg;
plot sub_num*DF_perf = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

proc glm;
Title3;
Title5;
class cue_type cue_comp;
model effort = cue_type sub_num(cue_type) cue_comp cue_type*cue_comp/P;
output out=hlw_ct_cc_effort P=peffort R=reffort cookd=cookd_effort;
test h= cue_type e= sub_num(cue_type);
means cue_type/duncan e = sub_num(cue_type);
means cue_comp/duncan;
means cue_type*cue_comp;

proc plot data=hlw_ct_cc_effort;
plot reffort*(peffort cue_type cue_comp) = '*' /vref=0;
Title3 'Residual Plots';
Title5 'Model in cue_type and cue_comp';

proc univariate data=hlw_ct_cc_effort plot normal;
var reffort;
Title3 'Residual Analysis';
Title5 'Model in cue_type and cue_comp';

proc plot data = hlw_ct_cc_effort;
plot cookd_effort*sub_num;
TItle3 'Plot of Cooks D';
Title5;

proc reg;
Title3;
```

```
Title5;
model effort = cue_type cue_comp /influence;
output out = hlw_ct_cc_effort_reg DFFITS = DF_effort;

proc plot data=hlw_ct_cc_effort_reg;
plot effort*DF_effort = '*';
Title3 'Plot of DFFITS against Effort Component NASA-TLX scores';
Title5;

proc plot data = hlw_ct_cc_effort_reg;
plot sub_num*DF_effort = '*';
Title3 'Plot of DFFITS values against subject numbers';
title5;

run;
quit;
```