1-1-2001

The Utility of Measures of Attention and Situation Awareness for Quantifying Telepresence

Jennifer M Riley

THE UTILITY OF MEASURES OF ATTENTION AND SITUATION AWARENESS FOR QUANTIFYING TELEPRESENCE

By

Jennifer M. Riley

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Engineering
in the Department of Industrial Engineering

Mississippi State, Mississippi

August 2001

THE UTILITY OF MEASURES OF ATTENTION AND SITUATION AWARENESS

FOR QUANTIFYING TELEPRESENCE

By

Jennifer M. Riley

Approved:

David B. Kaber
Assistant Professor of Industrial Engineering
North Carolina State University
(Director of Dissertation)

Larry G. Brown
Professor and Head, Dept. of
Industrial Engineering
(Committee Member)

John Usher
Professor of Industrial Engineering
(Committee Member)

John V. Draper
Oak Ridge National Laboratory
(Committee Member)

Jeffery N. Jonkman
Assistant Professor of Statistics
(Committee Member)

Stanley F. Bullington
Professor of Industrial Engineering
Graduate Coordinator

A. Wayne Bennett
Dean of the College of Engineering

Name: Jennifer M. Riley

Date of Degree: August 4, 2001

Institution: Mississippi State University

Major Field: Industrial Engineering

Major Professor: Dr. David B. Kaber

Title of Study: THE UTILITY OF MEASURES OF ATTENTION AND SITUATION AWARENESS FOR QUANTIFYING TELEPRESENCE

Pages in Study: 152

Candidate for Degree of Doctor of Philosophy in Engineering

Telepresence is defined as the sensation of being present at a remote robot task site while physically present at a local control station. This concept has received substantial attention in the recent past as a result of hypothesized benefits of "presence" experiences on human task performance with teleoperation systems. Human factors research, however, has made little progress in establishing a relationship between the concept of telepresence and teleoperator performance. This has been attributed to the multidimensional nature of telepresence, the lack of appropriate studies to elucidate this relationship, and the lack of a valid and reliable, objective measure of telepresence. Subjective measures (e.g., questionnaires, rating scales) are most commonly used to measure telepresence. Objective measures have been proposed, including behavioral responses to stimuli presented in virtual worlds (e.g. ducking virtual objects). Other research has suggested use of physiological measures, such as cardiovascular responses to indicate the extent of telepresence experiences in teleoperation tasks.

The objective of the present study was to assess the utility of using measures of attention allocation and situation awareness (SA) to objectively describe telepresence. Attention and SA have been identified as cognitive constructs potentially underlying telepresence experiences. Participants in this study performed a virtual mine neutralization task involving remote control of a simulated robotic rover and integrated tools to locate, uncover, and dispose of mines. Subjects simultaneously completed two secondary tasks that required them to monitor for "low battery" signals associated with operation of the vehicle and controls. Subjects were divided into three groups of eight according to task difficulty, which was manipulated by varying the number, and spacing, of mines in the task environment. Performance was measured as average time to neutralize four mines. Telepresence was assessed using a "Presence" questionnaire. Situation awareness was measured using the Situation Awareness Global Assessment Technique. Attention was measured as a ratio of the number of “low battery" signal detections to the total number of signals presented through the secondary task displays. Analysis of variance results revealed level of difficulty to significantly affect performance time and telepresence. Regression analysis revealed level of difficulty, immersive tendencies, and attention to explain significant portions of the variance in telepresence.

DEDICATION

This research is dedicated to my family. Thank you for your endless support.

ACKNOWLEDGMENTS

I express sincere gratitude to the many people without whose selfless assistance this dissertation could not have materialized. First of all, sincere thanks are due to Dr. David B. Kaber, my committee chairman and major professor, for expending so much time and effort to guide and assist me through the intricacies of the doctoral program and the dissertation process. Appreciation is also due to the other members of my dissertation committee for the invaluable aid and direction that they provided: Dr. Larry Brown, Dr. John Usher, Dr. John Draper, and Dr. Jeffery Jonkman.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

For some time, it has been the desire of humans to project their capabilities into environments remote to them, that is, to have the sense of existing in, or acting on, a place other than the one that they are actually physically located in (Tachi, 1999). Many newly developing technologies in the area of teleoperation and virtual reality (VR) are increasing the human's ability to complete tasks remotely. Advanced human-computer interfaces, as a part of these technologies, are thought to be useful for facilitating perception of a remote physical world or a computer-mediated world by using displays to present visual, auditory, and haptic sensory information. The design of these sophisticated interfaces is aimed at providing the impression of being "present" in the remote or simulated environment being experienced by the user (Loomis, 1992). It has been speculated that interfaces that facilitate this sense of "presence" also enhance human task performance, for example, with virtual interfaces/controls used to manipulate teleoperators (Hine et al., 1995; Sheridan, 1992; Bystrom et al., 1999).

In the context of teleoperation, telepresence is described as the sense of being present at a remote robot task environment while physically present at a local control station. Because of its hypothesized benefit to performance, telepresence has been identified and generally accepted as a design criterion for remote manipulation systems or teleoperators. Although some researchers have conducted studies to examine telepresence, and have found empirical evidence to suggest a relationship between telepresence and performance (Witmer & Singer, 1998; Riley & Kaber, 1999; Kaber et al., 2000; Kaber and Riley (in press), there is still confusion in the literature regarding the following: (1) factors that may affect or comprise telepresence; (2) other mental constructs that may be related to telepresence; and (3) how telepresence is related to performance.

1.1 Teleoperators and Teleoperation

The concept of telepresence emerged in the teleoperation arena. Teleoperation involves remote task completion using a teleoperator. Draper (1995) defines teleoperators as "human-robot systems that combine powerful human perceptual and problem solving capabilities with the hardiness of machines." Sheridan (1991) defines a teleoperator as a machine that extends a persons' sensing and manipulation capability to a remote location that is either hostile or remote to him or her. Teleoperators have been further described as a subdivision of synthetic environments, along with VR systems and telecommunications. A synthetic environment provides computer-mediated human interaction with an environment that is physically separate from the user in order to allow human perceptual, cognitive, and psychomotor capabilities to be projected into normally inaccessible, hostile, or simulated environments (Draper et al., 1999). These and other definitions of teleoperators (Jelatis, 1975; Corliss & Johnsen, 1968) describe four major functional components of teleoperation (or remote task performance) offered by Schloerb

(1995): (1) a human operator using a teleoperator system that consists of (2) an operator interface and (3) a teleoperator (machine) to act on (4) a remote environment.

1.2 Applications for Teleoperators

Teleoperators were first developed for the nuclear industry because of the potential for human injury and illness in hazardous environments and the need to maintain human involvement in operations. The increase in demand for work with radioactive materials in industrial nuclear "hot cells" led to a need for remote handling/manipulation systems (Pretlove, 1998). Today, humans may use teleoperators in various other task situations. For example, telerobots are now being used in bomb disposal and minefield remediation. Lemhofer (1999) describes a remotely controlled robotic vehicle equipped with a manipulator arm that allows a user to observe an object, to use various probes for identifying the object, and to use tools to deactivate the object that is identified as a bomb. A similar system is described by Eisenhauer et al. (1999). They discuss the Enhanced Teleoperated Ordnance Disposal System (ETODS) that was developed for humanitarian demining. The system can be used for minefield vegetation clearance, detection and identification of mines, and excavation. This system has been demonstrated to be an effective and efficient demining tool, which minimizes threats to the human operator during excavation of landmines.

The types and functioning of teleoperators have evolved, as well. For example, a teleoperator can be completely controlled by a human operator (teleoperated), or it may be semi-autonomous (operating under shared control between a human and a computer system). Teleoperators can also be either stationary or mobile (able to navigate or to be navigated) (Couture et al., 1998; Battelle & Ridao, 1999; Stone, 1992). A wide range of applications has been considered for each of these types of teleoperators. For example, Norman et al. (1999) describe use of the semi-autonomous mine detection system (SAMS) that allows for safe and efficient landmine detection. The system can be completely manually operated or operated under a semi-autonomous mode using Global Positioning System (GPS) waypoint planning and navigation and Semi-autonomous Control (SAC) software. In general, these systems are desired for use under the following conditions: (1) human health and life may be compromised (e.g., nuclear waste disposal, bomb disposal, minefield clearance and hostage negotiations); (2) humans may adversely affect the environment (e.g., medical applications and clean-room operations); and (3) humans are not able to occupy the physical space (e.g., deep space exploration and nano-robotics) (Pretlove, 1998; Massimino & Sheridan, 1994; Stassen & Smets, 1995; Tcherneshoff, 1998).

Batelle and Ridao (1999) state that, in each of the aforementioned technical fields, a common goal is pursued. This goal is to move robots closer to the human in regard to cognitive, perceptual and psychomotor skills. One advantage of human operators is that they are able to reason and to exercise judgement (Klein, 1997). This helps them to solve problems when facing contingencies which may arise in complex, and potentially cluttered, remote environments. Human operators, especially experts, are able to improvise and alter procedures on the basis of perceived system/environmental states (Klein, 1997). Technology is currently limited in how accurately robots can be designed to model human functioning and behavior. In support of the above, Pretlove (1998) gives two reasons why humans are still needed in the control loop for most telerobotic applications - because of their high level of skill, and because machine technology has not advanced enough to operate completely autonomously and intelligently in complex and unstructured environments. Until computer capabilities (e.g., perception for judgement, environmental navigation and motor control) match, or come close to, those of humans, the human operator is needed during operations, even if only as a system supervisor.

1.3 Anthropomorphism in Teleoperator Design

In an effort to mimic and project human capabilities through the design of teleoperators, some slave robot designs in earlier systems were made to bear a strong resemblance to the human in shape and functional properties of the torso and limbs, as well as sensory capabilities (using sensors for perception of visual and auditory stimuli). The limbs and effectors of the slave devices move in concert with the motions of the human user, while video cameras and audio devices convey visual and audio signals from the remote task environment to the user. These types of systems are considered to be highly anthropomorphic and follow the philosophy that robotic systems function best when approximating human size and shape (Corliss & Johnsen, 1968). As well, they feature output movements easily mapped to user controller inputs (spatial correspondence). This type of design has been aimed at helping to project human functionality or “presence” into a remote environment. In fact, in earlier studies, some researchers observed that with some highly anthropomorphic designs, users reported a “compelling impression of telepresence” or “remote presence” – the sense of being at the location occupied by the slave device (Corker et al., 1980). More recently, however, designs of teleoperators and research in teleoperations have been less concerned with the size and shape of robotic devices as a design methodology critical to good performance, but maintains the need to provide for presence (Draper et al., 1998).

1.4 Telepresence in Teleoperations

Developing technology in VR and virtual displays has led to a move from anthropomorphism to a focus on high-fidelity displays and the provision of rich visual information on the remote site for facilitating a sense of presence for users (Stone, 1992). An example is the interface for the Telepresence Remotely Operated Vehicle (TROV) developed by NASA Ames Research Center. The TROV is a complex robotic agent that has been used in exploration of environments below the sea surface near McMurdo Science Station, Antarctica. The robot is operated through a virtual environment vehicle interface (VEVI) that makes use of four video cameras and advanced sensor technology. Two of the cameras allow for stereo viewing of the remote environment with the use of active eyewear, like light shutter glasses. This type of system is considered to be a "telepresence system", defined by Huber and Davies (1996) as a system that provides the human operator with visual input through a remote video system.

Pretlove (1998) states that to operate efficiently in a remote environment, the human requires sufficient visual information to be able to interpret the remote scene. He continues by stating that a telepresence system is one that displays high-quality visual information about the task environment such that the human feels physically present at the remote site and can effectively and efficiently complete the teleoperation task. Similarly, Sheridan (1992) contends that given a "sufficiently high-fidelity display, a mental attitude of willing acceptance, and a modicum of motor participation”, the human user can experience telepresence during teleoperation (or virtual presence in virtual environments (VEs)). (See Appendix A for all relevant acronyms.)

In an effort to further enhance the experience of telepresence, some teleoperation systems extend the provision of sensory information beyond visual stimulation to include auditory, kinesthetic, and tactile feedback from the remote site. For example, Stone (1992) discusses the Virtual Environment Driving Experiment which involved development of an experimental test bed that provides an interactive audio-visual display by integrating a remote, head-slaved stereo TV-microphone system with a stereo HMD, and an interactive tactile feedback glove assembly.

1.5 Defining Telepresence

In more technical terms, telepresence is a psychological experience of the human that is postulated to enhance sensorimotor, as well as cognitive, performance in teleoperation. As a concept, telepresence has been defined in numerous ways. It has been described in terms of teleoperations - meaning present at a distance or control at a distance. Telepresence has been described in terms of human-machine interface quality and its compatibility to the “behavioral-physiological performance capability and limitations of the human” (Smith & Smith, 1989). It has also been described in terms of realism, that is, how well does system produces seemingly accurate representations of real objects, events, and people: Do they look, sound, and feel like reality (Lomard & Ditton, 1997)? Lastly, telepresence has been described in terms of the experience of the human operator. For example, Sheridan (1992) states that telepresence is the “sense of being physically present with virtual objects at the teleoperator site"; "feeling like you are actually 'there' at the remote site of operation." Tachi (1999) states that telepresence (telexistence) is having a "real-time sensation of being at a remote location" while being able to interact with the remote environment.

1.6 Telepresence and Performance

Telepresence in teleoperations is of continuing interest because of its hypothesized benefit to overall human-system performance. Telepresence has been hypothesized to have an influence on human performance with virtual control interfaces for real system operations (e.g. teleoperation (Hine et al., 1995)). It is thought that if users feel an association with remote task operations (i.e., as if they are actively influencing, or have control over, the remote site), they will experience more effective performance at the control interface. Presence has also been equated to performance. Schloerb (1995) states that presence experiences are based on the probability of completing a (virtual) task successfully. Similarly, Bystrom et al. (1999) postulate that having some sense of presence in an environment is a necessary condition for performance to occur. However, as Welch (1999) points out, although there is a "pervasive belief that presence is causally related to performance," there is actually little empirical evidence to support this claim.

Draper et al. (1998) state that it will be difficult to derive a direct relationship between presence and performance. This may be, as they state, because of the potential complexity of the relationship, or because there has been a lack of "appropriate studies" conducted to specifically elucidate the relationship (Welch, 1999). Some researchers

(Welch, 1999) have proposed types of studies to be conducted in order to determine the relationship between presence and performance. Welch (1999) suggests manipulating some factor speculated to influence presence (e.g., the capacity for stereo viewing of an environment) and carefully measuring an aspect of task performance that is not directly relevant to the manipulated factor. For example, an experimenter could compare effects of stereoscopic and monoscopic vision on performance in locating a two-dimensional (2-D) target among other 2-D objects. This would "unconfound" the manipulation of presence. Then, the researcher can be relatively certain that the increase or decrease that may be observed in performance is directly related to the increase in telepresence and not caused by the addition of stereo vision, which itself might enhance performance in locating three-dimensional (3-D) targets.

The debate regarding a relationship between presence and performance is a critical issue. Ultimately, the importance of telepresence is dependent upon its impact on performance. If there is a relationship (causal or correlational) between presence and performance, then there is a need to establish telepresence based design guidelines for teleoperation systems (e.g., the visual interface, controls, etc.). If it is demonstrated that there is no relationship between the two system response measures, then it may be ill advised to continue to design or modify systems with the goal of increasing the sense of presence.

1.7 Telepresence Measures

Telepresence is a concept based on many factors. For example, technological factors such as stereoscopic cues (Hendrix and Barfield, 1996), response latency (Ellis et al., 1997), limited field of view (FOV) (Prothero and Hoffman, 1995), and display type (Deisinger et al., 1997) are hypothesized to affect a user's degree of association with a remote task and environment. Psychological/cognitive factors suspected to influence telepresence include motivation, concentration, and self-control (Psotka & Davison (1993), susceptibility to immersion (Witmer & Singer, 1994), and willingness to suspend disbelief (Lombard & Ditton, 1997). As well, task factors such as length of exposure (Stanney et al., 1998), difficulty/complexity (Sheridan, 1992; Draper et al., 1999) and, level of automation (Sheridan, 1992) are also thought to be critical to telepresence. Because of its multi-dimensionality, it is a challenge for researchers to develop a valid and reliable measure of telepresence. Further, there is likely no single index that will adequately assess experiences of telepresence (Stanney et al., 1998). There is currently no consensus as to how to measure presence. The only common belief is that a measure of presence should be reliable, repeatable, and robust (Stanney et al., 1998; Sheridan, 1992). Both subjective and objective measures have been used and proposed.

1.7.1 Subjective Measures

Typically, telepresence is measured using subjective methods, like questionnaires aimed at capturing the degree to which users feel “present in”, or “a part of”, the remote/virtual world. Subjective measures of presence require participants to report a conscious judgement regarding their “feelings” or experiences. These methods involve psychological measuring devices such as rating scales (e.g., “on a scale from 1 to 10 rate the amount of presence you felt”) or subjective comparisons (e.g., “which of the two environments produced the greatest amount of presence”) (Stanney et al., 1998).

Subjective methods are usually presented through a paper and pencil survey following the task experience. For example, Draper and Blair (1996) measured telepresence following completion of a teleoperation task as a composite response to a pair of presence items. The items included: (1) “I felt as though I were actually in the remote environment as I performed the task,” and (2) “The experience involved unity or fusion of self with the remove environment.” Subjects in this study responded to each item by rating the occurrence of items as follows: (1) very rarely, (2) rarely, (3) sometimes, (4) frequently, and (5) very frequently. Higher ratings on the items with this measure indicated increased subjective telepresence.

Witmer and Singer (1994) developed a presence questionnaire (PQ) which measures presence along three major subscales: (1) involved control, (2) naturalness, and (3) interface quality. The items on the PQ are aimed at capturing the degree to which characteristics of the environment along the three subscales add to, or take away from, presence experiences. Example PQ items are “How aware were you of events occurring in the real world around you?” and “How much did the visual aspects of the environment involve you?” Each item on the PQ is subjectively scored on a scale from 1 to 7, with higher scores being indicative of higher presence experiences. The PQ is a post-trial technique, requiring subjects to report telepresence experiences subsequent to exposure to the task/VE.

There have also been “on-line” measures taken during task performance using hand-dial potentiometers (Freeman et al., 1999) or using verbal responses (Slater & Steed, 2000). Verbal reports elicited during task performance involve the users’ stating when they perceive changes in their “feelings” of being in the remote/virtual world or the physical local environment. An advantage to such on-line measures as compared to post-trial measures is that they help to eliminate problems associated with memory limitations and failure to accurately recall and report experiences subsequent to task performance.

Another subjective technique is the method of paired comparisons. Schloerb (1995) and others have proposed this method which involves having observers compare a real and virtual environment for dissimilarities. (This method may also be used to compare two virtual worlds.) If the observer is unable to distinguish between the two, then it is supposed that the sense of presence in the virtual world will be just as strong as that in the real environment. At the current level of VR technology, it is unlikely that users will mistake a virtual world for a real one, so the procedure typically involves observation of a real scene that is systematically degraded until the observer can no longer distinguish it from the virtual representation. In this technique, the level of degradation required to reach the point of “just noticeable difference” represents the degree of presence.

Researchers have used questionnaires and other subjective methods in studies to measure telepresence, in part, because they appear to be valid measures of the concept (c.f., Witmer & Singer, 1998). For example, the reliability, validity and utility of the PQ developed by Witmer and Singer (1994) have been iteratively investigated. Results of these studies indicate that the PQ is a reliable scale and presence as measured using the questionnaire is a valid construct (Witmer & Singer, 1998; Stanney et al., 1998). In support of subjective measures, Sheridan (1992) has stated that a multidimensional subjective rating scale with factors or dimensions that are carefully qualified should allow subjects to give reliable responses to experiences of telepresence. Sheridan (1992) cites the success of multi-factor rating procedures used for measuring the construct of mental workload as an example. Subjective rating procedures are also used because they are easy to administer and fairly cost effective. Further, since telepresence has been described as a subjective sensation or mental construct, it may be more amenable to subjective measures as compared to objective ones (Draper et al., 1999; Sheridan, 1992).

Although subjective measures have been generally accepted as a valid means of assessing telepresence, they suffer from limitations. In the case of post-trial measures, the quality of the results is dependent upon the memory of the user for an event. Like other subjective measures, for example measures of subjective workload discussed by Wickens (1992), the costs are related to the uncertainty with which a user's response truly reflects his/her experience. For example, Dillon et al. (2000), reporting on presence studies by Ijsselsteijn et al. (1997; 1998) and Freeman et al. (1999), stated that given the subjective measures used in the studies, it was not clear exactly what the subjects were responding to or how their responses were affected by continuous assessment of their own experiences. Other researchers have also expressed concern that the "act of introspection" may have an unpredictable effect on subject responses (see Resources for the Study of Telepresence, and Presence Measurement, available on-line at: http:/nimbus.temple.edu/~mlombard/Presence/measrue/html). In addition, Stanney et al. (1998), citing Ellis (1996), note the need to be aware of the potential for inconsistent reports across raters and rating situations.

Beyond the above concerns, there is concern that items on telepresence questionnaires may be difficult for subjects to understand. This is particularly a problem if questions explicitly refer to the concept of telepresence, rather than presenting items that are related to characteristics of the system that might affect the sense of presence. Related to this, is the issue of a lack of standardization of items on various questionnaires. Few researchers use the same set, or sets, of presence measurement items. Several questionnaires exist and/or have been proposed, and the items presented through each vary greatly (e.g., Witmer & Singer, 1994; Prothero & Hoffman, 1995; Ditton, 1997; Lombard et al., 2000; Lessiter et al., 2000).

1.7.2 Objective Measures

With the limitations of subjective measures of telepresence in mind, researchers have expressed the need for objective measures (Sheridan, 1992; Welch, 1999). Objective measures involve recording study participants' behavioral and physiological responses in order to correlate them with relevant psychological states. Objective measures may be preferred over subjective measures, in part, because they do not involve any subjective introspection on experiences, and they can typically be administered and recorded during an experience. Several ways to objectively measure telepresence have been proposed in the telepresence literature. Held and Durlach (1987) suggest assessing or observing user responses to unexpected or seemingly threatening stimuli in the remote or virtual environment. This follows the philosophy of behavioral realism (Slater & Wilbur, 1997). The idea here is that when subjects are presented with simulated or remote environments via displays/interfaces that make them feel present, they will respond to the stimuli in that environment in the same way that they would if the stimuli were presented in a real environment. A potential drawback to this type of measure, however, is that it could compromise virtual or teleoperation task performance.

A study by Slater and Usoh (1993) employed such a method for measuring telepresence. They observed subject responses to situations of "relative" danger in a virtual space. They looked for subject reactions such as ducking when virtual objects flew toward the subject's head, physical behavior or verbal responses in reaction to standing on a narrow virtual plank over a precipice, and physical reactions to an outside disturbance (a loud noise not in the VE) caused by the experimenter.

Physiological responses have also been proposed as objective measures for telepresence. People routinely experience some type of physiological response to stimuli in the real world (e.g., increased heart rate with the intensity of an emotional response). Similar responses should occur within virtual worlds that may serve as interfaces to, or representations of, real scenarios if they are designed in a sufficiently realistic manner. Observing changes in physiological states may include looking at changes in posture, muscular tension, cardiovascular behavior, ocular responses, or skin conductance (Barfield & Weghorst, 1993).

Prothero et al. (1999) discuss another alternative to objective telepresence measures they call direct or "Class A" measures of telepresence. These measures involve subjects being exposed to conflicting cues from the real and virtual world. Prothero et al. (1999) state that telepresence is an illusion of position and orientation in an environment. They propose that telepresence involves a switch in the cues one uses to determine his/her position or orientation - from using cues defined by the real world, to using those defined by the virtual or remote world. In this case, telepresence is objectively measured in terms of the degree to which subjects respond to the virtual cues rather than the real ones. The researchers provide an example of an experiment that might be used to test the utility of direct measures of telepresence. Participants in the study could be exposed to visual virtual cues and real inertial cues. The subjects could then be asked to adjust inertial amplitude so that they believe that they are at rest. If the visual cues from the VE are attended to and sufficiently compelling to produce vection, subjects are expected to adjust to a non-zero amplitude for inertial oscillation in order to counteract the perceived motion. The magnitude of inertial oscillation chosen by the user could then be used as an objective measure of telepresence.

As with the subjective measures currently being used to measure telepresence, there are drawbacks to using the above-mentioned objective measures. Prothero et al. (1999) point out that there is not, at this time, any strong evidence to suggest that physiological objective measures correlate well with the mental construct of telepresence. This may, in part, be due to a lack of a clear definition and understanding of telepresence. It is difficult to say just what the quantitative relationship is between these measures and "telepresence". Stanney et al. (1998) state that another drawback is that objective measures tend to be "all-or-none". That is, one either experiences presence or they don't; there are no varying "degrees" of presence to be experienced. Subjective measures, such as the PQ (Witmer & Singer, 1994), however, provide the capability to assess varying degrees of telepresence.

1.7.3 Representing Telepresence through Measurement of Other Mental Constructs

Research on the development of a good measure of presence that is commonly accepted is still needed. To this end, other techniques have been proposed for representing telepresence. Telepresence is hypothesized to be related to mental constructs like attention and SA (Draper et al., 1998). It is suspected that telepresence may be objectively measured through these constructs, as they have similar underlying and influential factors (e.g., human limitations in the perception of multiple stimuli in parallel, the ability to filter stimuli for information processing, and the human tendency to become distracted) (Wickens, 1992).

1.7.3.1 Attention as a Measure of Telepresence

Attention is defined by James (1890) as “taking possession of the mind, in clear and vivid form, out of what seems several simultaneous possible objects of trains of thought”. Attention is a mechanism that organizes human cognitive functioning by including only perceptions that directly relate to the objects of interest. Telepresence may be related to attention in that a human operator of a teleoperation system may focus on information related to the robot task in the remote environment, to the exclusion of all other stimuli (Draper et al., 1998). Fontaine (1992) stated that presence seems to be a matter of focus, which occurs when one directs attention toward an event. Witmer and Singer (1998) say that how sharply users focus attention on the VE may determine, in part, how much presence they will report. This suggests that there might be several degrees of telepresence to be experienced (ranging from none to maximal). Thus, the amount of attention being allocated to either the remote or local environment, in which the user is simultaneously involved, may affect or determine the degree of telepresence. In fact, Bystrom et al. (1999) state that if sufficient attentional resources are allocated to the virtual or remote environment, the user may view them as the place of actual occupancy. There has been no research relating established measures of attentional resource allocation to telepresence measured through subjective or objective techniques. Consequently, there is no evidence at this point to indicate whether or not there is some initial threshold of attention allocation that must be achieved in order to cause telepresence experiences.

Another postulation in the literature regarding telepresence and attention is based on the concept of selective attention. Selective attention refers to the tendency to choose to focus only on information that is meaningful and/or of particular interest (Witmer & Singer, 1998). Meyers (1998) states that selective attention means that at any moment one focuses his/her awareness on only a limited aspect of all that is capable of being experienced. The argument for selective attention as a determinant of telepresence is that the user may consciously select to attend to either the remote/virtual or local site in order to acquire task-relevant information to the exclusion of stimuli in the other location. The idea, as stated by Witmer and Singer (1998), is that, although the VE may present novel information to the user, which may initially attract more attentional resources, telepresence will depend more upon how well these novel stimuli and other virtual objects are connected to task performance over time. Bystrom et al. (1999) state that in order to interact with the VE at all, the user must allocate attentional resources to it and that the amount of resources assigned to the virtual world may be influenced by the nature of the task. For example, a particularly engaging, or cognitively challenging, task at the remote site may lead users to allocate more attention to it and, consequently, may result in increased telepresence.

Attention is thought to be a resource of limited quantity. Wickens (1992) discusses a theory describing a single undifferentiated pool of attentional resources available to all tasks and mental activities. This theory, known as the Single Resource Theory (SRT), suggests that if some processes or aspects of task performance (e.g., perception of critical events or response execution) require a substantial amount of attentional resources, then less of it will be available for other processes. This may consequently lead to decreased performance on certain tasks. Relating SRT to telepresence in teleoperation, if the local control room requires substantial attentional resources, then less will be available for allocation to remote site operations; therefore, telepresence and teleoperation performance may degrade. On the other hand, if more attention is allocated to the remote site than to the local space, telepresence may increase.

Other researchers have argued a Multiple Resource Theory (MRT) of attention, which suggests that, instead of a single supply of undifferentiated resources, people have several different capacities with resource properties (Navon & Gopher, 1979; Wickens, 1992). Wickens (1992) describes three dimensions of attentional resources: (1) the two stage-defined resources (early and late processes); (2) the two modality-defined resources (auditory and visual encoding); and (3) the two resources defined by processing (spatial versus verbal). According to the MRT, tasks demanding separate rather than common resources, along any of the three dimensions, will not experience performance problems associated with limitations of attentional resources. That is, the result will be more efficient time-sharing, and the difficulty or workload experienced in one task will have less influence on performance on other tasks. Relating MRT to telepresence in teleoperation, if the local control room requires attentional resources associated with a different dimension/modality (e.g. auditory) from the remote site (e.g., visual), then telepresence in the remote task and performance in the remote or local task should not be significantly influenced. On the other hand, if both the remote and local sites compete for attentional resources along the same dimension (e.g., both requiring attentional resources for visual perception), telepresence (and possibly performance) in the remote task, and/or performance in the local task, may be negatively affected depending upon where (to which site) attention is allocated.

Draper and Blair (1996) proposed an attentional model for synthetic environments featuring a structured resource approach based on a model by Wickens and colleagues (Wickens, 1980; Wickens et al., 1983). In the Draper and Blair model, the human user, who is simultaneously involved in two environments, may potentially be attending to: (1) a computer-mediated world (e.g., through an interface to a teleoperator) and, (2) a local environment containing controls linked to the computer-based environment. Information from the remote site can be presented to the user in the local space through use of displays (visual, auditory, haptic, etc.). At the same time, information in the local environment is presented. The stimuli presented from either place may or may not be related to operator task performance. Thus, as Draper et al. (1998) point out, the user can be exposed to meaningful data, as well as, distracters (non task-related items). In this situation, there are four types of stimuli the user must process: (1) remote task-related items; (2) remote distracter items; (3) local task-related items; and (4) local distracter items. In the model, telepresence can then be described as the operator selecting the computer-mediated environment or remote stimuli for allocation of attentional resources over the local space.

Attentional resource allocation can be measured in terms of a hit-to-signal ratio in signal detection tasks, and it is hypothesized that such a measure may be useful in quantifying telepresence experiences. The signal detection paradigm is applicable to any situation wherein there are two discrete states of the world that cannot be easily distinguished (signal versus no signal) (Wickens, 1992; Sanders & McCormick, 1993). For example, an operator may be responsible for monitoring for various critical events (e.g., teleoperation system automation failures) across a local and remote robot site using system interfaces and controls. The signal to be detected by the human can have only two possible response categories: (1) "yes, a signal is present" (i.e., there has been a failure in the system automation), or (2) "no, a signal is not present" (i.e., there has not been a failure with the system). Combinations of the two discrete states of the environment (signal or no signal) with the possible user responses ("yes, I detect a signal" or "no, I do not detect a signal") produce a 2×2 matrix with four possible events. The events include: (1) hits – a signal is present and is correctly detected; (2) misses – a signal is present but is not detected; (3) false alarms – a signal is not present but is perceived; and (4) correct rejections – a signal is not present and is not perceived. Perfect performance within this matrix encompasses no misses and no false alarms. Measures of attention in signal detection and monitoring tasks take the form of a ratio of the number of hits (correct detections of signals) to the total number of signals presented to the operator.

The hit-to-signal ratio in SDT can be used to describe the user in terms of his/her sensitivity to signals in the presence of noise. Noise can be external to the human (e.g., "static" in auditory signals) or internal (e.g., hyper-neural activity) (Sanders &

McCormick, 1993). A person who desires to correctly detect as many signals as possible among noise may have a low decision criterion, or level of neural activity, that must be met in order for him/her to respond that a signal is present. This is a risky operator. Conversely, a person may be conservative in his/her approach to signal detection. He/she may have a high criterion of neural activity that must be met before stating that a signal is present. In the first case, the operator will likely detect many of the signals presented to him/her, but also have many false alarms (because of the low criterion for detection). In the second case, the operator will miss some signals due to a high criterion for detection, but will have fewer false alarms.

Some systems will have more external noise than others, and some operators will have more internal noise than others. These issues can have an effect on sensitivity to signals in the environment. As well, an operator's response criterion can be influenced by the likelihood of observing a signal (Sander & McCormick, 1993). This may influence the amount of attentional resources allocated to the detection of signals in the environment. For example, operators who perceive the probability of observing a signal to be high may lower their response criterion level and allocate more attentional resources to the task of monitoring and, as a result, experience a higher hit-to-signal ratio. In general, individual differences play an important role in SDT and how an operator may respond to critical system indicators.

It is hypothesized that the hit-to-signal ratio, which has been used to represent attention allocation, may also be used to represent telepresence in teleoperations. An increase in the hit-to-signal ratio for detecting stimuli in the local task environment would be indicative of a decrease in attention to the remote task and possibly a decrease in telepresence. Conversely, a decrease in the hit-to-signal ratio for detecting stimuli in the local task environment would indicate an increase in attention allocated to the remote environment and possibly an increase in telepresence.

One potential problem with using signal detection as a means of assessing telepresence in teleoperation may stem from the fact that the general construct of attention is comprised of two components, including detection (perception) and recognition (identification) (Wickens, 1992). Human short-term sensory stores (STSS), an aspect of the human information processing system, detect stimuli in an environment and permit the information to be temporarily stored for possible use at a later time. Three important properties of the STSS include: (1) it is pre-attentive, that is, it requires no conscious attention to detect stimuli; (2) it is veridical, that is, it preserves most of the physical details of the detected stimuli; and (3) it decays rapidly, remaining less than one second for visual stimuli (Wickens, 1992). Statements (1) and (3) are the most important in terms of attention as a measure of telepresence. This is because they suggest that it is possible for an operator to detect and temporarily store stimuli in STSS and not consciously attend to, recognize, or comprehend the meanings or relationships of the stimuli to overall task goals. Therefore, assessing operator attention allocation using SDT may not necessarily reflect perceptual processing and it may be limited in its applicability to presence. An operator of a telerobot could potentially detect signals from the remote site, but not comprehend them to cause feelings of association with the task or the environment.

1.7.3.2 Situation Awareness as a Measure of Telepresence

With the limitations of an attentional measure based on signal detection in mind, an alternate mental construct, SA, which has been related to attention allocation in complex systems control, has also been proposed for representing telepresence (Draper et al., 1998). Situation awareness has been described as "generating purposeful behavior", that is, behavior that is directed toward a task goal (Smith & Hancock, 1995). Several other descriptions exist for the concept. Sarter and Woods (1991) believe that SA involves coherent situation representation that is continuously updated on the basis of recurrent assessments. Endsley (1988) has formally defined SA as the "perception of elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future." It has been hypothesized as being critical to operator task performance in complex and dynamic operations (Endsley, 1995; Salas et al., 1995). Smith and Hancock (1995) argue that SA specifies what must be known in order to solve a class of problems when interacting with a dynamic task environment.

Endsley's definition suggests three levels of SA: (1) Level 1 SA, perceiving critical factors; (2) Level 2 SA, understanding what the critical factors mean and their relationship to each other and task goals; and (3) Level 3 SA, understanding and being able to project what will happen with the system in the future. By this definition, Level 1 and Level 2 SA are directly related to attention (perception and recognition). Endsley and Jones (1997) state that attention is a critical mechanism that individuals use to achieve SA. In earlier literature, Fracker (1988) suggested that SA is the knowledge that results when attentional resources are allocated to a "zone of interest." Endsley and

Jones (1997) go on to say that direct attention is needed in order to perceive and process information for achieving SA and for selecting actions and responses. Supporting this statement, Sekuler and Blake (1994) believe that awareness of objects and events in the environment (i.e., situation awareness) guides the actions of individuals.

SA can be limited in dynamic task environments, like teleoperation, by the availability of attentional resources. Allocation of attentional resources to one stimulus over another may mean a loss of SA on certain elements of a task. Relating this concept to teleoperation, because an operator's cognition is distributed across two different places at the same time, he/she may develop SA on the remote or local space alone, or together. An operator's ability to develop good SA on the two environments jointly will be critically affected by their capability to divide attention across the local control station and remote environment (Draper et al., 1998). An increase in operator attention allocation to the remote or local environment for achieving SA may result in loss of SA on the alternate environment. In terms of representing telepresence, this means that as SA (perception, recognition, and understanding) on the remote task environment increases, the degree of telepresence should also increase. Draper et al. (1998) suggest that telepresence is the, "maximization of SA in the remote environment." Conversely, as SA on the local environment increases, telepresence may decrease.

In order to use SA as a measure of telepresence, a good objective measure of SA is required. A "good" measure of SA is one that can be described as follows: (1) sensitive to changes in SA on the basis of changes in performance requirements of a task; (2) unobtrusive to the task (i.e., does not interfere with, contaminate, or alter task performance); and (3) reliable and effective with repeated use. Several techniques for measuring SA have been proposed and used. Endsley (1995) reviews various methods, including:

(1) Physiological methods (e.g., P300 and other electroencephalographic measures) have shown promise for identifying whether information is cognitively processed. This measure, however, fails to assess how much is retained in memory, or whether or not the information was registered correctly and understood.

(2) Performance measures are usually non-intrusive and can be automatically recorded. They can, however, suffer from limited diagnosticity, as poor operator performance can occur for reasons other than poor SA (e.g., high workload and poor sampling strategies).

(3) External task factors usually involve artificially altering the information being displayed to an operator and measuring the time it takes for him/her to react to the change. These measures may lack diagnosticity because experimenters cannot predict how users will react to the addition or removal of information. External task factors may also alter the realism of the task.

(4) Rating techniques usually involve use of scales to assess SA. These techniques are low cost and relatively easy to use. Unfortunately, they are limited by a subject's ability to accurately report his/her own level of SA. They may also be influenced by the performance of the subject. For example, good performance may result in reports of good SA. Examples of this type of measure include the SART (Situation Awareness Rating Technique) by Taylor (1990) and the SWORD (Subjective Workload Dominance) technique by Vidulich (1989).

Questionnaires have also been used to measure SA and are discussed by Endsley (1995). These measures can allow for detailed information about SA to be collected on an element by element basis and evaluated against reality. This can be an objective measure that is administered either post-trial or on-line. The advantage of questionnaires is that subjects have ample time to answer lengthy questions regarding SA following a trial. However, with post-trial methods, the data is subject to limitations in memory. On-line measures can eliminate this problem. The SA questions can be asked during task performance. Problems with this alternative, however, are the potential for increasing operator workload through queries and altering task performance (Endsley, 1995).

As a way to solve the problems with questionnaires, the “freeze” technique was developed. Under this technique, the task/simulation is halted and subjects are queried as to their perceptions of the situation at the time (Endsley, 1987; Endsley, 1988; Fracker, 1991). Subject responses can later be compared to the real situation recorded at the time of a query in order to assess SA. A technique administered in this manner is the SAGAT (Situation Awareness Global Assessment Technique) (Endsley, 1987; Endsley, 1988). The SAGAT was developed for use in aircraft simulations, but the methodology has been used in other applications, including air traffic control (Endsley & Rodgers, 1994) and simulated radar monitoring (Kaber, 1997). The technique involves developing SA queries on the basis of information requirements for successfully completing task performance. Items related to all levels of SA are included in a question list. Questions may also be related to system functionality and status. Queries of the SAGAT should be directly compatible with the knowledge required for performance of the specific task (e.g., specific to aircraft piloting) (Endsley, 1995). Once developed, the queries are presented to subjects during task performance. The simulation is halted at random intervals and questions are randomly selected from a pool of questions, which are then presented in random order. Subjects quickly answer questions regarding their perceptions of the situation and are returned to the simulation. Data gathered on subjective perceptions of SA during testing are later compared to reality.

Endsley (1995) conducted experiments using the SAGAT to examine fighter pilot SA with different cockpit display types. She found the process of querying participants through simulation freezes did not substantially alter SA for subsequent task performance. She demonstrated no negative performance effects with this technique with as many as three stops during a 15-minute performance period. She also found that simulation freezes lasting between 5 and 6 minutes did not intrude on task performance.

The SAGAT technique has several advantages. First, subjects cannot prepare for questions in advance as they are randomly selected from a pool of queries and the timing of their presentation is random. Secondly, queries can provide a detailed assessment of SA as the questions cover a wide range of situation knowledge, including Level 1, Level 2, and Level 3 SA. Thirdly, the methodology eliminates problems with post-trial methods.

It is speculated that this type of measure could be used as an objective one for representing telepresence in a teleoperation task. For example, subjects could be required to respond to queries on the basis of SA requirements for successful remote-task completion, aspects of the teleoperation system and performance (e.g., actions at the interface), as well as current and future system states. The measure could serve to indicate the degree of association with the remote environment experienced by a subject during task completion. For example, high SA on remote events might indicate high levels of telepresence.

CHAPTER II

PROBLEM STATEMENT

At this point, there is limited understanding of the concept of telepresence. Few studies have been conducted to empirically assess the many factors – technological, psychological and task - that have been hypothesized to affect telepresence. Even fewer studies have demonstrated evidence of a direct link between telepresence and performance. Further, to date, there is no commonly accepted model of the concept. This limited understanding of what telepresence is makes it difficult, or near impossible, for us to use telepresence as a predictor of human-machine system design and overall performance.

Most researchers acknowledge that telepresence is a complex and multidimensional variable (Biocca & Delaney, 1995; Kalawasky, 1998; Sheridan, 1992; Welch et al., 1996). Some have devised and executed studies to learn more about the origin of telepresence, its underlying factors, and its utility to performance. In order to address the question of what the defining experiences of the sense of presence are, Schubert et al. (1999) conducted a study involving 246 participants with the goal of decomposing presence in to definable components. First the study participants were exposed to VEs, and following the experiences, they were asked to complete a 75-item survey developed by Schubert and colleagues. The survey was created by combining questions from previously used questionnaires (developed by Carlin et al., 1997; Ellis et al, 1997; Hendrix, 1994; Slater & Usoh, 1994; Towell, 1997; and Witmer & Singer, 1994), questions from their past research, and newly designed questions. Schubert et al. (1999) conducted first- and second-order factor analyses on the data collected from the survey. Results of the study revealed three different telepresence components: (1) spatial-constructive; (2) attention; and (3) reality judgement. Each of these components was separate from five immersion factors. The results from this study offer some information about the factors/components comprising telepresence and support the importance of the role of attention in telepresence experiences. As an ancillary finding, the results supported the contention that telepresence is different from immersion.

Beyond attempts to expound upon what causes telepresence, some studies have been conducted to examine effects of various VR system parameters on telepresence and its relationship with constructs like performance and workload. For example, Hendrix and Barfield (1996) conducted a study which demonstrated that simulation update rate affects the sense of presence. Prothero and Hoffman (1995) have shown that limiting the FOV near the eye results in a decrease in telepresence experiences. Draper and Blair (1996) conducted a study using a teleoperator in a pipe-cutting task that demonstrated telepresence to be significantly correlated with workload and the concept of flow. In a series of simulation studies, Kaber and others (Riley & Kaber, 1999; Kaber, et al., 2000; Kaber & Riley, in press) examined telepresence effects of factors such as display type, control mode, response latency, visual display configuration, and user motivation. They have also used these studies to assess the potential relationship between telepresence and performance. Results of these investigations revealed a significant relationship between presence and performance. They have found that with performance increases with increases in subjective telepresence in simulated teleoperation tasks. Kaber and others have also been able to use results of these studies to provide preliminary design guidelines for teleoperator system display resolution, task duration, and control/display configuration (Riley & Kaber, 1999; Kaber, et al., 2000; Kaber & Riley, in press). This research is important because it serves to verify a relationship between telepresence and performance and to motivate further investigation of the concept including its measurement.

At this point, the most common method of measurement of telepresence is subjective assessment through questionnaires (Nash et al., 2000). Some subjective techniques have been validated as reliable measures of telepresence (e.g., the PQ by Witmer & Singer, 1994), but they suffer from limitations (e.g., memory limitations, poor subject ability to accurately report experiences/feelings, etc.). With this in mind, there has been a call for objective measures of telepresence in the literature. Several objective measures have been proposed, including investigating the potential of SA and attention for effectively quantifying telepresence in remote operations. It has been hypothesized that SA can serve to indicate levels of telepresence, with an increase in SA on remote events and operations indicating an increase in the sense of presence at a remote site. Additionally, attention has been hypothesized to be important to telepresence, with expectations that increases in allocation of attentional resources to remote task information should result in increases in user telepresence experiences.

CHAPTER III

EXPERIMENT

The goal of this investigation was to determine whether measures of SA and attention could be used to quantify telepresence. In an attempt to further knowledge about measurement of telepresence, an experiment involving a virtual teleoperation task was conducted. The specific objectives of this study are listed in the following section.

3.1 Objectives

The objectives of this experiment were as follows:

1. To explore the utility of a measure of SA and an attention allocation measure for predicting the strength of telepresence during teleoperation task performance.
2. To explore the hypothesized relationships between SA and telepresence and attention and telepresence.
3. To further investigate the hypothesized relationship between telepresence and human task performance.

An experiment was conducted during which subjects were asked to complete a simulated teleoperation task. Performance data in the teleoperation task were collected and examined along with observations of user SA using the SAGAT. Performance in two embedded secondary monitoring tasks (performed simultaneously with the teleoperation task) was also assessed. Secondary task performance was recorded as a measure of attention allocation to the virtual teleoperation task or real control environment. Data on telepresence was collected using the PQ developed by Witmer and Singer (1994). Cognitive workload was examined using the Modified Cooper-Harper Scale (MCH) (Wierwille & Casali, 1983). Presence data and workload data were recorded at the end of each test trial.

3.2 Tasks

3.2.1 Teleoperation Task

Both the teleoperation task and secondary monitoring tasks were computer-based and were presented on various computer displays. The simulated teleoperation task involved a mine neutralization scenario in which subjects navigated a robotic vehicle (rover) equipped with a manipulator and several demining tools. The tools included: (1) an airknife used to uncover mines by removing dirt around them; (2) a shotgun used to detonate mines that could not be safely removed from the ground; and (3) a gripper tool used to pick-up mines from the ground with the manipulator. The tasks required subjects to navigate the rover in an outdoor environment and to use the tools to locate, uncover, identify, and neutralize landmines from the space. The subjects viewed the virtual task through a stereoscopic display using Crystal Eyes (stereographic) goggles. A standard mouse controller was used to direct the motion of the rover in the VE. A standard computer keyboard was used to select and activate tools and to manipulate components of the robotic arm.

The mouse controller facilitated motion of the rover in the forward, backward, left, and right directions. The rover could also be brought to a complete stop in the environment. During navigation subjects listened for an auditory signal indicating the proximity of a mine to the rover. The auditory signal was a “ringing bell” that was activated when any part of the rover or robotic arm was directly over a mine. Once a mine was located, the operator’s goal was to use the airknife tool to uncover the mine and identify whether it could be safely removed from the virtual ground.

Subjects used verbal commands for various task activities through a “Wizard of Oz” setup (see Figure 1 for a photograph of the experimental setup). A controller at a separate workstation viewed a computer monitor presenting the VE and executed the verbal commands from a subject by using a standard keyboard. Commands were used to do the following: (1) control the auditory mine detection signal; (2) control rotations of segments of the robot arm; (3) select tools; (4) use tools; (5) control a two-dimensional HUD on or off; and (6) respond to one of the secondary monitoring tasks (“low telerover battery” signal detection task). A complete list of the verbal commands used during task performance can be found in Appendix B.

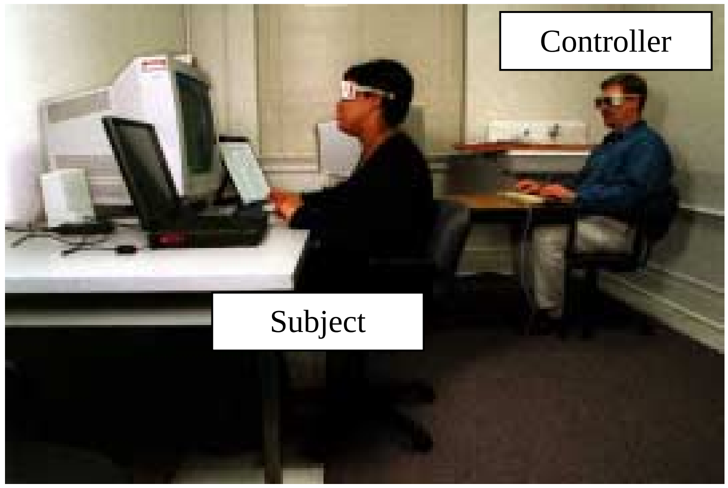


Figure 1. Photograph of "Wizard-of-OZ" experimental setup.

Mine neutralization in the VE required subjects to select the appropriate tool based upon whether or not a mine could be excavated from the ground. There were two types of mines hidden in the outdoor space. One type, an anti-personnel (AP) mine, could be safely excavated from the area. Subjects were mandated to pick-up and remove these mines and transport them to a storage area located in a corner of the VE. For AP mines, subjects used verbal commands to direct the motion of the robot arm to position it over the mine. Subjects used commands to select the gripper tool and use the tool to grasp the mine. Once the mine was removed from the ground, subjects transported the mine to the storage area. Verbal commands were then used to position the arm over a storage box and release the mine from the gripper tool. When the mine was released into the storage box, neutralization was complete.

The second type of mine, anti-tank (AT) mines, could not be safely removed from the ground. Because these mines would explode if picked up using the gripper tool, subjects were required to detonate them where they had been virtually buried. When AT mines were located and uncovered, subjects used verbal commands to select the shotgun tool and position the robot arm to aim the shotgun at the mine. Subjects then detonated the mine by discharging the shotgun. A graphical representation of the task steps is provided in Figure 2.

A finite number of mines were present in the virtual space. Subjects were required to completely neutralize (either detonate or excavate) a total of four mines to complete a teleoperation task trial. Displays and menus were presented to the subject indicating the robot tools that were available and/or in use during various stages of the task. Figure 3 shows a graphic of the displays and environment presented to subjects.

3.2.2 Monitoring Tasks

There were two embedded secondary monitoring tasks (i.e., the tasks were described to subjects as being part of the overall functioning of the system). The tasks required subjects to monitor displays for visual signals indicating a critical event associated with operation of the rover and controls in the teleoperation tasks. One signal was presented in the VE display. It was superimposed over the image of the virtual task environment in the lower left corner of the monitor. The second signal was presented on a separate display in the real environment. The real world (RW) monitoring signal was presented using a portable computer. The signal was presented in the lower right corner of the system display. Both signal displays appeared identical in features and operating characteristics. The monitoring signals consisted of a single “light” indicating the battery level of various components of the teleoperation system for performance of the tasks. The "light" was comprised of a 2-D colored object drawn on the displays.

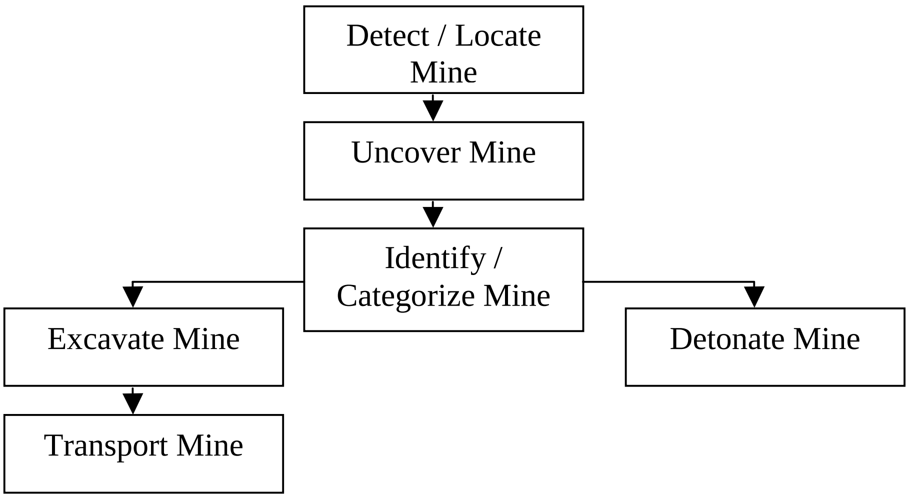


Figure 2. Hierarchical diagram of steps in virtual telerover-assisted mine neutralization.

Figure 3. Graphical representation of the teleoperation task and VE.

Subjects were informed that the system was equipped with two batteries (see Appendix C for a complete set of subject instructions). Each subject was instructed that battery power levels were critical to system operation and should be monitored for the duration of the teleoperation task. The lights representing the battery signals appeared as green squares when the batteries were operating under normal conditions (i.e., full battery power). The lights appeared as green triangles to signal low power levels. Participants were required to respond to all low battery signals by using verbal commands and keyboard strokes. Subjects were advised that low battery signals presented in the VE were associated with power for motion of the rover and functioning of the tools. To respond to this signal, subjects provided a verbal command to the controller. The controller executed the command by striking a key on the controller keyboard to return the battery to its normal state. Subjects were instructed that the second battery (RW signal) was associated with the communications system. To respond to this signal, the subject was required to strike the "b" key on the portable computer keyboard in order to return the battery to normal operating conditions. This portion of the task was psychomotor in nature, requiring participant monitoring, diagnosis, and action. The timing and number of signals were randomized for both displays. The tasks were intended to describe the manner in which (or to which display) subjects allocated their attentional resources. That is, the tasks were used to determine if subjects were focusing on the VE or the real environment. This was accomplished by assessing subject performance in detecting signals across the two displays. The monitoring tasks were displayed in the positions described above in order to ensure that they fell within the same field of view for subjects. Considering these positions, there was essentially an equal likelihood of detecting signals across the two displays. Figure 4 shows a photograph of the virtual and real-world battery displays.

Figure 4. Photograph of virtual and real-world battery displays for monitoring tasks (see shapes inside superimposed circles).

3.3 Experimental Variables

3.3.1 Independent Variable

The independent variable examined in this study was level of difficulty (LOD) in the teleoperation task. Three settings of LOD were used in the study: low, moderate, and high. The LOD was manipulated by varying the total number of mines and spacing of mines in the VE. As the number of mines increased, and the space between mines decreased, the difficulty in locating mines decreased. The real-world area of the VE used in this experiment was approximately 320 feet wide × 320 feet deep. The low, moderate, and high LOD simulations included 20, 15, and 10 mines respectively. The minimum separation distance (spacing) of mines under any LOD setting was substantial enough that subject navigation of the rover was not hindered. That is, the rover was able to move in the environment such that it was not continuously located over the top of a mine. Mines were hidden in the VE such that no mine was located under virtual trees or other objects in the environment (i.e., all mines were accessible to the rover).

3.3.2 Dependent Variables

The dependent variables of interest in this study were teleoperation task performance, telepresence, workload, SA and attention. Subject immersive tendencies were also measured and recorded using the Immersive Tendencies Questionnaire (IQ) developed by Witmer and Singer (1994).

3.3.2.1 Performance

Performance in the primary teleoperation task was measured in terms of time-to-mine neutralization. Time-to-mine neutralization was measured in seconds from the beginning of subject interaction with the VE to the successful neutralization (removal and storage or detonation) of a single mine. This performance measure was then averaged across the four mine neutralizations to obtain a measure of average time-to-mine-neutralization.

3.3.2.2 Telepresence

The PQ was used to assess the degree to which subjects felt a part of the virtual or remote task environment. It included 19 empirically validated questions describing three subscales: (1) involved control, (2) naturalness, and (3) interface quality. The involved control category included questions regarding ability to control events in the remote environment and how aspects of task performance promoted subject involvement. The naturalness subscale included questions regarding the naturalness of control devices and interactions with the environment. The quality of the interface subscale included items regarding the fidelity of sensory modality (visual, auditory, etc.), perception and movement, as well as environment richness. Each item on the PQ was rated on a 7-point scale according to instructions provided by Witmer & Singer (1994). The scores for each item were combined to yield a composite telepresence score, with higher scores on the survey indicating increased telepresence. The PQ, in combination with the IQ, has been found to be a valid measure of telepresence with reliable results over multiple experiments (Nash et al., 2000). The complete list of questions included in the PQ can be found in Appendix D.

3.3.2.3 Workload

The MCH is a univariate workload scale directed at evaluating tasks in terms of accomplishability, number of errors, and level of performance. The MCH consists of a flow diagram with various descriptor terms regarding task difficulty and operator demand level. The terms are associated with, and intended to describe, a subject's perceived mental workload. The mental workload ratings range from “1” (“Very easy/operator mental effort is minimal”) to “10” (“Impossible/instructed task cannot be accomplished reliably”). The MCH Scale has been validated in various types of studies. For example, it has been reliable in measuring cognitive workload in tasks requiring perception of critical events, like detection of critical system indicators during simulated flight (Casali & Wierwille, 1982). It has been reliable in measuring workload in tasks requiring cognition (e.g., simulated aircraft flight with navigation tasks requiring arithmetic and geometric operations at varying levels of complexity) (Rahimi & Wierwille, 1982). It has also been successfully used in communication tasks (e.g., aircraft control and communication requiring detection, recognition, comprehension, and response to flight commands provided by an experimenter) (Casali & Wierwille, 1983). Results of these studies indicate that the MCH Scale and its ratings are statistically reliable indicators of overall mental workload (Wierwille & Casali, 1983). The MCH Scale diagram can be found in Appendix E.

3.3.2.4 Situation Awareness

The SAGAT technique involved stopping the teleoperation task at random points in time to administer queries concerning both the current and future states of the teleoperation task. The SA queries were designed to assess Level 1 SA (perception of critical task elements and events), Level 2 SA (comprehension of elements and their relation to the task goal), and Level 3 SA (projection of future environmental or task states). Appendix F presents the list of queries on SA developed in the context of the mine neutralization task. Example queries are listed here for each level of SA:

1. What is the current tool type in use? (Level 1 SA)
2. What type of mine is currently being neutralized? (Level 1 SA)
3. Is there a mine in close proximity to the rover? (Level 2 SA)
4. Can the mine currently being neutralized be safely removed from the area? (Level 2 SA)
5. What will happen if the current mine is picked up? (Level 3 SA)
6. How many more mines will you have to neutralize after the current mine is completed? (Level 3 SA)

Subject responses to the queries administered during task performance were compared to "real" situation data (Endsley, 1995) recorded by the experimenter during task performance. Subject SA was recorded as the percentage of correct responses to queries for each level of SA.

3.3.2.5 Attention

Attention was measured using the secondary monitoring tasks. Participants were required to simultaneously monitor the two battery displays while performing the simulated teleoperation task. The monitoring tasks required subjects to detect when the "light" displays signaled low battery power levels. Attention was measured in terms of monitoring performance using hit-to-signal ratios. Attention was recorded as a ratio of the number of low battery signals detected to the total number of signals presented in the virtual and the real forms of the secondary task display. A ratio of attention allocation across the VE and RW was also computed.

3.3.2.6 Immersive Tendencies

The IQ is an 18-item questionnaire intended to capture the degree to which VE users are predisposed to becoming immersed in a virtual task environment. Many of the items on the questionnaire measure subject involvement in common or everyday activities. Other questions are aimed at assessing a subject's state of alertness or fitness and ability to focus or redirect attention (see Appendix G for a full list of IQ items). Like the PQ, each question on the IQ was scored on a 7-point scale, with higher scores indicating increased tendency toward immersion. Scores for each item were combined to yield a composite IQ score. These scores were recorded prior to task performance and later analyzed to assess potential covariances with other experimental variables. This analysis was conducted to ensure that there was not any one subject group that was predisposed to immersion more so than any of the other experimental groups. If this were the case, a potential bias might have existed in assessing the effects of LOD on presence.

3.4 Experimental Design

A between-subjects design was used with LOD as the control variable. Subjects were randomly assigned to the three groups of LOD: low, moderate, or high. Each group included the same number of subjects. Repeated measures of performance, telepresence, mental workload, SA, and attention were recorded for each subject in each group. The experimental layout used in the study is presented in Table 1.

Table 1. Data collection table for experiment.

	Level of Difficulty (LOD)		
Replications	*Low*	*Moderate*	*High*
	Subjects		
1	1-8	9-16	17-24
2	. .	. .	. .

Statistical analyses included a Multivariate Analysis of Variance (MANOVA) to assess an overall LOD effect and one-way Analyses of Variance (ANOVAs) on all response measures. The ANOVA model used in this research can be written as:

$$y_{ij} = \mu + LOD_i + e_j$$

where,

y_{ij} = the response variable (e.g., performance, telepresence, etc.);

μ = the grand mean;

LOD_i = the fixed effect of the ith level of difficulty; and

e_j = error within each LOD.

In order to draw inferences about the potential significance and strength of relationships among the various response variables, a multiple linear regression (MLR) model was also investigated. The key response variable in this study was telepresence. In an attempt to determine the utility of using measures of SA and attention to describe telepresence, percent correct responses to all SAGAT queries (SA) and the attention allocation ratio across the VE and RW (attention) were included as regressors in the MLR. Because of potential effects of LOD on telepresence and other measures, the independent variable LOD was also included as a regressor in the model of telepresence. A variable representing immersive tendencies was also included in the model to account for pre-existing individual differences regarding susceptibility to telepresence. Both of these variables were added to the model to avoid biases in the parameter estimates for SA and attention that might have occurred if variance due to LOD or IQ were erroneously attributed to SA or attention. The MLR model used in this research can be written as follows:

$$y = \beta_0 + \beta_1(X_1) + \beta_2(X_2) + \beta_3(X_3) + \beta_4(X_4) + \varepsilon$$

where,

y = the response variable, telepresence;

β_0 = the intercept average value when X_{is} = 0;

β_1 = the partial regression coefficient associated with X_1 (SA);

β_2 = the partial regression coefficient associated with X_2 (attention);

β_3 = the partial regression coefficient associated with X_3 (LOD);

β_4 = the partial regression coefficient associated with X_4 (IQ);

X_1 = the first regressor (SA);

X_2 = the second regressor (attention);

X_3 = the third regressor (LOD);

X_4 = the fourth regressor (IQ);

ε = error associated with measurement of y.

Correlation analyses were also conducted to assess relationships among the response measures. Pearson Product-Moment Correlation Coefficients were computed to evaluate the potential relationships listed below:

1. telepresence and performance;
2. telepresence and workload;
3. performance and workload;
4. telepresence and immersive tendencies;
5. SA and attention;
6. immersive tendencies and performance;
7. immersive tendencies and workload;
8. immersive tendencies and SA; and
9. immersive tendencies and attention.

3.5 Subjects

Twenty-four North Carolina State University students were recruited for participation in this study on a voluntary basis. Visual acuity was used as a selection criterion based on the need to visually immerse subjects in the task in order to potentially observe telepresence, particularly when using the PQ. All subjects were required to possess 20/20 or corrected to normal vision. All subjects were required to have experience using personal computers and familiarity with mouse controllers and standard keyboards.

Subjects ranged in age from 19–26 years, with an average age of 20.25. The subject population included 2 females and 22 males. All were right handed in terms of mouse controller use. All subjects reported at least some video game experience.

A power analysis was conducted to determine the expected power of hypothesis tests with 24 subjects (8 subjects per LOD group). This analysis involved using power function tables for fixed-effects models and procedures described by Keppel (1982, pg. 70-72, 549). A calculated quantity used for referencing power charts, (φ), was determined using an equation based on deviations of the treatment means from the grand mean ($\mu_i - \mu$), the population variance (σ^2), and the number of levels of the independent factor. Power levels were calculated for a significance level of α=0.05.

3.6 Equipment

The experimental setup included:

1. an Intergraph TDZ2000GX1 workstation with a high-performance graphics subsystem presenting the VE;
2. two 21-inch graphics monitors operating under frame interlaced stereo mode at 1280 × 1024 resolution;
3. two pairs of Crystal Eyes stereovision goggles;
4. one standard computer keyboard;
5. one standard mouse controller; and
6. a portable Dell computer.

Figure 5 shows a photograph of the overall experimental setup. Monitors A and B presented the virtual teleoperation task and VE battery signal to the subject (A) and the controller (B). Monitor C presented the RW battery signal to the subject. Light shutter glasses were worn by the controller and provided to each subject for exaggerating the 3-D imagery of the VE.

Figure 5. Photograph of experimental setup.

3.7 Procedures

3.7.1 Training and Testing Protocol

The experimental procedures for the study spanned across two days. The first day of experimentation involved an introduction to the study and task training. The second day included subject testing and data collection. At the onset of experimentation, each subject participated in a familiarization period. This period was provided to acquaint subjects with the objectives of the study, the training and testing procedures to be followed, and the equipment to be used. During the familiarization period, subjects completed a consent form and an anthropometric survey. The survey was intended to gather information on the subjects' qualifications for participation in the study. The familiarization period lasted approximately 30 minutes.

Subsequently, subjects were provided with an explanation of the IQ and instructed on how to complete the questionnaire. Subjects were allowed time to answer all questions on the survey. Completion of the IQ required approximately 10 minutes.

Subjects participated in an extensive training session including instruction on how to operate the simulated robotic vehicle using the mouse controller and how to manipulate the robotic arm using the keyboard and verbal commands. The training session was divided into three periods. The first period involved instruction, demonstration, and hands-on training on use of the keyboard and mouse controller for completing the simulated teleoperation task. It included two 14-minute trials in which the subjects were allowed to practice several excavations and detonations of mines in the VE. The trials were separated by a 2-minute break.

The second training period involved instruction, demonstration, and hands-on training on the use of the keyboard for completion of the secondary monitoring task. Subjects were allowed to practice the monitoring tasks without having to simultaneously perform the teleoperation task. A 7-minute period was provided for monitoring the VE battery signal and responding with the verbal command. Another 7-minute period was provided for monitoring the RW battery signal and responding using the portable keyboard. Subjects also trained under multi-task conditions. A 14-minute session was provided for practicing simultaneous performance of the teleoperation task and monitoring tasks. The total time for the second training period was approximately 30 minutes, which was followed by a 2-minute break.

Subjects completed a third training period involving an explanation of the SAGAT queries and the administration of surveys during test trials. Subjects were allowed to practice multi-task performance involving SAGAT freezes and queries for a period of approximately 15 minutes. The total time for the experiment familiarization and training was about 2 hours.

At the start of the second day of experimentation, subjects were briefed on the testing procedures. They were then provided with an overview of, and instructions on how to complete, the PQ and the MCH Scale. Explanation of these forms required approximately 15 minutes.

As part of experimental testing, each subject was required to complete two trials. Subjects were instructed to neutralize 4 mines during a single trial and to respond to as many low battery signals as possible during the course of the teleoperation task. The SAGAT query stops were conducted to assess SA during task performance. A total of three SA stops occurred during each trial. No advance information on the number of SAGAT queries or query presentation times was provided to subjects. Each trial required approximately 35 to 50 minutes. Immediately following task performance, telepresence and workload data were collected. The list below presents a summary of the experimental procedures and the approximate time for completion of each aspect of training and testing:

Stage of Experiment	Step	Duration
Introduction	Familiarization Period	30 minutes
	Completion of IQ	10 minutes
Training Period 1	Teleoperation task training	14 minutes
	Break	2 minutes
	Teleoperation task training	14 minutes
	Break	2 minutes
Training Period 2	VE monitoring task training	7 minutes
	RW monitoring task training	7 minutes
	Break	2 minutes
	Teleoperation task + Monitoring task training	14 minutes
Training period 3	Teleoperation task + Monitoring task training + SAGAT	15 minutes
Experimental Testing	Pre-trial survey explanation	15 minutes
	Trial 1 with 3 SAGAT stops	35-50 minutes
	Trial 2 with 3 SAGAT stops	35-50 minutes

The total experimental time was approximately 4 hours.

3.7.2 Trial Schedule

At the start of each trial, subjects were asked to press the “g” key on the portable computer keyboard to begin task performance. This action recorded the start time of the trial and activated a timer for the RW battery monitoring task. At the same time, the controller pressed the “g” key on the standard keyboard to initiate the start of the VE battery monitoring task and record the start time for the trial. The monitoring tasks were programmed to allow subjects to perform the teleoperation task for two minutes without the occurrence of a low battery signal. This was done to promote subject adaptation to the task circumstances. Low battery signals were presented at random points in time during the course of task performance following this 2-minute period.

Subjects were allowed to perform the teleoperation task for five minutes without a SAGAT query stop. This was done to promote subject achievement of awareness of the simulated task situation (Endsley, 1995). Endsley has stated that an informal rule for administering the SAGAT is to ensure that no query freezes take place earlier than 3–5 minutes into the task to allow time for the participants to build-up a picture of the situation. In the present experiment, the first SA measure occurred at random during the second five- minute time block of testing. Once the first query stop took place, a 1-minute time period with no query was allowed (this time period was based on research by Endsley (1995)). The second query stop occurred at random within the next 9-minute period, followed by another 1-minute period without a query stop. The last query stop occurred at random within the following 9-minute period. Each SA query stop was approximately 3 minutes in duration. As previously reported, Endsley (1995) has found that freezes lasting 5 and 6 minutes do not cause memory decay or intrude on task performance.

Following the last SA stop, subjects continued with task performance until four mines were completely neutralized. Depending upon the difficulty level experienced by the subjects, the task continued approximately 5 (for low LOD) to 20 (for high LOD) minutes beyond the last SA query measure. Subjects were required to monitor the VE and RW battery displays for signals until the teleoperation task was completed. Performance in terms of time-to-mine-neutralization was recorded during the trials at the completion of each neutralization (i.e., a detonation or disposal). Subjects were asked to press the "q" key on the portable computer keyboard at the end of each trial. This was done to record the stop time for the trial and end the RW monitoring simulation. The controller pressed the "q" key on the standard keyboard to stop the VE monitoring and teleoperation tasks. At the end of each trial, the measures of telepresence and workload were recorded. Figure 6 shows a graphical representation of a typical trial schedule. (NS denotes no low battery signal. NQ denotes no SA query. Numbers represent minutes in trial.)

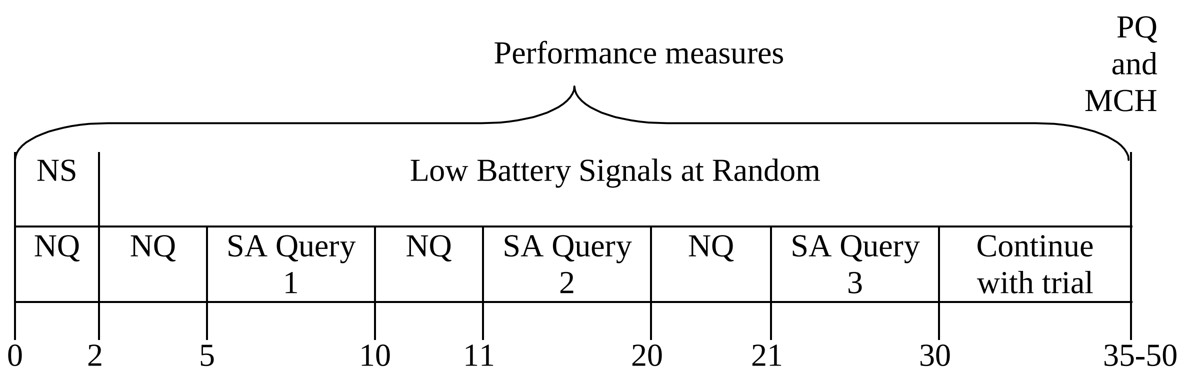


Figure 6. Trial schedule.

<u>3.8 Data Collection</u>

Performance time in seconds was recorded by the computer system for each mine neutralization. Time for SA query stops during mine neutralizations was subtracted from the performance time. Neutralization time was averaged for each, because average performance was better associated with the post-trial measures of telepresence and workload, as well as the attention allocation ratio, which was calculated on the basis of hit-to-signal data for an entire trial (not individual mine neutralizations). Consequently, a total of 48 observations on average performance were recorded (2 trials × 8 subjects × 3 levels of difficulty).

Telepresence and subjective mental workload were measured immediately at the end of each trial. The composite PQ scores were calculated and used for data analysis. Forty-eight observations of telepresence and the same number of subjective reports of mental workload were collected during the experiment (2 trials × 8 subjects × 3 levels of difficulty).

Situation awareness was recorded at each SAGAT query stop. Subjects were presented with a 6-item questionnaire developed by randomly selecting 2 questions for each level of SA from the overall pool of 16 SA queries. During a query stop, subjects were instructed to discontinue task performance and turn away from the display. They were then instructed to move to a separate experiment station so that they could not view the task while responding to the SA queries. Subjects were allowed time to complete all SA questions. Subject responses were compared to the real situation as recorded by the controller. The percent correct responses to all SA queries administered at each stop was recorded, with values ranging from zero to 100% (0% (0/6), 16.7% (1/6), 33.3% (2/6), 50% (3/6), 66.6% (4/6), 83.3% (5/6), or 100% (6/6)). Average SA was determined by calculating the mean percent correct responses across the three SA stops. This resulted in a total of 48 observations on average overall SA (2 trials × 8 subjects × 3 levels of difficulty). Situation awareness was also assessed for each level, as defined by Endsley (1995). Since there were two questions representing each level of SA on each survey, percent correct responses for any level of SA could have values of 0% (0/2), 50% (1/2), or 100% (2/2). Average SA for each level was determined by calculating the mean percent correct responses for each level of SA across the three query stops. This resulted in a total of 48 observations on average SA for Level 1, Level 2, and Level 3 (2 trials × 8 subjects × 3 levels of difficulty).

Attention was measured in the VE and in the RW. The monitoring task was designed such that the “batteries” for the system had a 20% failure rate. (Subjects were not aware of the battery failure rates.) Subjects were presented with low battery signals at random points in time with the potential for a low battery signal at 30-second intervals. A hit was recorded each time subjects detected and responded to a low battery signal (i.e., a green triangle). The occurrence of a signal was automatically recorded by the computer system. The number of hits and signals was recorded and used to calculate the ratio of hits-to-signals in the VE and hits-to-signals in the RW for each 1-minute interval during teleoperation task performance. The total hit-to-signal ratio for an entire trial was recorded for data analysis for the VE (HSVE) and RW battery (HSRW) monitoring tasks. Also, the total hit-to-signal ratio in the VE was compared to the total hit-to-signal ratio in the RW to determine subject attention allocation across the two environments. The attention ratio was calculated as HSVE/HSRW. The result was 48 observations on attention to the VE and attention to the RW (2 trials × 8 subjects × 3 levels of difficulty). Similarly, there were 48 observations of the attention ratio across the VE and RW (2 trials × 8 subjects×3 levels of difficult).

Immersive tendencies, as measured using the IQ, were recorded prior to training and task performance. A composite IQ score was recorded for each subject. Consequently, there were 24 observations on IQ (8 subjects × 3 levels of difficulty).

3.9 Statistical Analysis

3.9.1 General Linear Model (GLM)

The PROC GLM procedure of the SAS software system (SAS Institute, 1990) was used to investigate effects of LOD on the various response variables. This procedure uses a least squares method to fit general linear models. It was used because it can specify classification variables that have discrete levels and it can be used for both MANOVAs and ANOVAs.

Initially, a MANOVA was conducted on all response measures. Multivariate Analyses of Variance permits study of more than one dependent variable in an experimental design to test for an overall independent factor effect. A MANOVA on LOD was conducted because it can help to control for experiment-wide error rates when several different dependent variables are of interest.

Using one-way ANOVAs, with LOD coded as a between-subjects variable, the following data were analyzed: average performance (avgperf), telepresence (PQ), overall average SA (avgsa), average Level 1 SA (avgsa1), average Level 2 SA (avgsa2), average Level 3 SA (avgsa3), mental workload (mch), ratio of attention across the VE and RW (attnrat), attention to the VE (HSVE), and attention to the RW (HSRW). All ANOVAs were conducted using the PROC GLM procedure of SAS. The F-tests on the LOD main effect incorporated the mean square error for SUB as a denominator. Therefore, individual differences within each LOD were accounted for in the analyses. All significant ANOVA results reported in this document are at the α=0.05 level of significance.

Duncan's Multiple Range (MR) tests were used to further investigate any significant main effects revealed through the ANOVAs. Duncan's MR test controls the Type I comparisonwise error rate (Chew, 1977). However, it does not control the experimentwise error rate (EER). Chew (1977) notes that some researchers object to the use of Duncan's test because of large experimentwise error rates that can occur with increases in the number of means that are compared. Large Type I error rates make Duncan's test a very sensitive test. That is, there is a large probability of detecting differences when differences actually exist. As well, Duncan's MR test is reported to be reliable when used after the overall F-test is found to be significant (Chew, 1977). All post hoc tests were conducted at a significance level of α=0.05.

3.9.2 Multiple Linear Regression Model

The PROC REG procedure of SAS was used to fit a linear regression model of telepresence by least squares estimates in order to assess the capability of SA and attention to predict telepresence in the experimental task. The PROC REG procedure was used to conduct an F-test for model significance/adequacy, provide estimates for the model parameters, conduct t-tests on the significance of model parameters, and to provide variable/model selection techniques. The procedure also allowed for conducting an analysis of multicollinearity among regressors.

It was suspected that multicollinearity might exist in the model because of the influence of attention allocation on ability to develop SA during task performance. The COLLIN technique as part of the PROC REG procedure of the SAS software was used to reveal any evidence of multicollinearity in the model. Collinearity diagnostics are suggested when explanatory variables are thought to be near linear combinations of each other (Hayter, 1996; Mendenhall & Sincich, 1995). It is more desirable to have regressors in a model that are not correlated with each other (Hayter, 1996), because variables that are related can negatively affect the capability of the model to adequately describe a response and inhibit capability to estimate regression coefficients accurately. The collinearity diagnostics used to reveal multicollinearity in this research included assessing the variance inflation factor (VIF), the model tolerance, and a condition index (CI) (SAS Institute, 1990).

The VIF is a measure of the extent to which the variances of model parameter estimates are inflated by the addition of a particular component to a model. The higher the VIF, the lower the precision may be in estimation of parameter coefficients. A VIF value greater than 10 indicates high multicollinearity among regressors (Hayter, 1996). The tolerance value is a measure of the strength of interrelationships among the explanatory variables in the model. The tolerance value is equal to the quantity $1-R^2$ that results from the regression of all other variables in the model on the explanatory variable of interest. If the variables are orthogonal to each other, the tolerance has a value of 1. If a variable is very close related to other variables, the tolerance value goes to zero (SAS Institute, 1990). Finally, the CI is based on the eigen structure of the X′X matrix. A CI value falling in the range of 30 to 100 indicates moderate to strong collinearity among regressors. The CI is used in conjunction with the proportion of variance measure provided by the COLLIN procedure. The collinearity diagnostic table presents the proportion of variance of the estimates accounted for by each principle component. A collinearity problem exists when there is a critical value CI (that is, CE>30) and there is a principle component that strongly contributes to the variance of two or more parameter estimates (SAS Institute, 1990; Myers, 1990).

These three collinearity diagnostic measures were used in combination to identify whether multicollinearity existed in the regression model. It should be noted that even if multicollinearity exists, it might not severely affect the quality of the fitted model. Furthermore, prediction of the response may also be fairly good, as long as the values of the independent variables (e.g., SA and attention) used to predict the response follow the same pattern of multicollinearity exhibited in the sample data (Mendenhall & Sincich, 1995).

Following the analysis for multicollinearity, variable selection methods were used to determine the optimal set of regressors for the regression model. The goal of variable selection methods is to minimize the error variance by choosing the set of regressors that yield the smallest mean square error (MSE) value. These methods are also used to minimize the variability of prediction and the variance of the parameter estimates provided by the regression model. The forward selection technique was used to identify which regressors should be included in a final model of telepresence based upon data collected for this study. Forward selection is an algorithmic technique that begins with a model containing only an intercept term and then adds the regressor that results in the largest increase in the coefficient of determination (R^2). A partial F-test is then conducted to decide if the regressor should stay in the model. These steps are repeated until the procedure fails to identify a variable that meets the significance level criterion for the F-test. The significance level for the forward selection procedure used in this investigation was $\alpha=0.05$.

3.9.3 Correlation Analysis

Pearson Product-Moment (PPM) Correlation analyses were also conducted on all response measures using the PROC CORR procedure of the SAS software. The PPM correlation coefficients were computed to provide a cursory look into the strength of linear associations between response measures. The technique is a simple way of summarizing the degree of association between two variables (Hayter, 1996). It is important to note, however, that correlation coefficients are not necessarily indicative of a causal relationship between variables. That is, although a significant correlation may be revealed, for example between telepresence and performance, it does not necessarily mean the telepresence causes performance or vice versa. As well, simple correlation analysis with Pearson's method does not account for effects of other experiment variables, for example individual differences due to the subject effect, on relationships between variables. Thus, a partial correlation procedure was also conducted to analyze relationships among variables while accounting for the variance due to subjects. The PARTIAL statement was used with the PROC CORR procedure to conduct the partial correlation analysis.

3.9.4 Residual Analysis and Influential Diagnostics

To prevent potential ANOVA assumption violations in the PROC GLM and PROC REG procedures of the SAS software, transforms were applied to various response measures according to the procedures described by Neter et al. (1990, pg. 142-146) and Hayter (1996, pg. 725-726). Transforms were applied to average performance (avgperf) and telepresence (PQ) modeled in terms LOD and SUB within each LOD for the ANOVA procedures. Table 2 summarizes the evidence of potential assumption violations revealed through diagnostics such as plots of the residuals versus the predicted values, normal probability plots (NPP), and Shapiro-Wilks tests. The evidence is provided on models of untransformed response measures driven by untransformed predictors/regressors. Specific actions taken to account for any violations are also presented in Table 2. Log transformations were applied to avgperf and PQ to account for non-normality and non-constant variance. It should be noted that although some the results reported in this document are on transformed responses, all plots present dependent variable means at various effect settings in original units to promote understanding and interpretation of results (Neter et al., 1990, pg. 147).

Table 2. Summary of evidence of potential assumption violations.

Response Variable	Non-Linearity	Non-Constant Variance	Non-Normality
Average Performance	Evidence: none. Corrective action: none.	Evidence: Plot of residuals against predicted avgperf revealing expanding/collapsing variance. Corrective action: Logarithmic transform applied to response.	Evidence: Significant Shapiro-Wilks test ($p<0.05$). Normal probability plot revealing slightly "heavy-tailed" error distribution. Corrective action: Logarithmic transform applied to response.
Telepresence Scores	Evidence: none. Corrective action: none.	Evidence: plot of residuals against predicted PQ scores revealing expanding/collapsing variance. Corrective action: Logarithmic transform applied to response.	Evidence: Significant Shaprio-Wilks test ($p<0.05$). Corrective action: Logarithmic transform applied to response.

Influential points and outliers are observations of response values for particular input variable settings which have unusually large influence on a statistical model (Hayter, 1996). Influential diagnostics were used to identify such points so that they could be evaluated and removed from the data set used in the PROC GLM and PROC REG procedures, if there was evidence in the data (e.g., atypical values for individual trials) to suggest that points were problematic. Several measures were used to identify points of high influence. These procedures and results are provided through PROC REG and

PROC GLM of the SAS software. Explanation of these procedures and their cutoff values for categorizing data points as influential are provided here:

1. DF BETAS – This is a measure of the number of standard errors the parameter estimates are changed when a point is removed from the estimation procedure. The cutoff value for the DF BETAS is calculated as $2/\sqrt{n}$. Here, n equals the sample size (Myers, 1990).
2. DF FITS – This is a measure of the number of standard errors a fitted value changes when a point is omitted from the estimation procedures. The critical value for the DF FITS is calculated as $2/\sqrt{p/n}$. Here, p equals the number of predictors included in the model (Myers, 1990).
3. COOK'S D – This is a measure of the change in the parameter estimates caused by deleting an observation. Comparatively large values of COOK'S D suggest that a point may be influential.
4. COVRATIO – This is a measure of how MSE is affected by removal of one point or a measure of the changes in covariance properties of the parameter estimates with or without a data point. Points are expected to exert an unusual amount of influence on the generalized variance if the COVRATIO is greater than 1+3*p/n or less than 1-3*p/n.

Residual values can also be used to identify individual data points that do not fit the model well (Myers, 1990). Residual values exceeding 2.0 in magnitude are considered to be suspect (Hayter, 1996). It should be noted that large residual values could occur by chance. However, it is also important to note that such points should be investigated to determine whether there is any evidence (e.g., measurement or performance issues) to suggest that they should be removed from the data set (Hayter, 1996). These diagnostic techniques were used in combination to identify points that were candidates for removal from the experimental data set. All outliers and influential points were removed from the data set for the reported results.

Bonferroni's method, which is commonly used in multiple comparison procedures, may also be used to evaluate points for model influence and was considered in the analysis. This method may be used to reduce the probability of observing significance in statistics (e.g., as observed with studentized residuals) by chance. Bonferroni's method can be used to adjust the significance level so that the type I error rate is controlled at the advertised level (0.05). Specifically, Myers (1990) states that use of the $(\alpha/n) * 100\%$ point of the t_{n-p-1} distribution will produce a significance level that is not larger than α. This method is described as useful, but also as a conservative method of outlier analysis (Myers, 1990). The influential diagnostics listed above were selected over the Bonferroni method because these tests may be more sensitive for detecting "real" existing outliers.

3.10 Hypotheses

The specific postulates examined in this study included:

1. The LOD is expected to significantly affect performance in terms of average time-to-mine-neutralization, mental workload measured with the MCH, and telepresence measured with the PQ. As LOD decreases (i.e., number of mines increases and distance between mines decreases), telepresence and performance

are expected to increase. Workload is expected to increase with increases in LOD. This hypothesis is based on previous research findings on effects of task difficulty on performance and telepresence in teleoperation (Kaber & Riley, in press).

2. Telepresence scores will be significantly, positively related to performance in the virtual teleoperation task. As observed in previous telepresence studies (Riley & Kaber, 1999; Kaber et al., 2000), it is expected that time-to-mine-neutralization in the teleoperation task will decrease with increases in PQ scores.
3. Telepresence scores will be significantly, positively related to SA. The degree of perceived association with the virtual task and environment is anticipated to increase with increased percent correct responses on SA queries. This result is anticipated on the basis of the hypothesized relationship between SA and telepresence discussed by Draper et al. (1998).
4. Telepresence scores will be significantly, negatively related to mental workload. As perceived mental loading increases, telepresence experiences reported with PQ are expected to decrease. The relationship between telepresence and workload is expected to be consistent with findings of previous research (Riley & Kaber, 1999; Kaber et al., 2000).
5. It is hypothesized that as performance improves in the monitoring task displayed in the real world, telepresence scores will decrease. As more attention is allocated to the visual display in reality, less should be available for sensing stimuli and focusing on the virtual task. Consequently, subject perceived association with the

VE should decrease. This postulate is based on the hypothesized relationship between attention and telepresence discussed by Draper et al. (1998).

6. Immersive tendencies (IQ) scores will be significantly related to telepresence, SA, and attention. It is anticipated that as IQ scores collected prior to task performance increase, telepresence and SA in the VE will increases. As well, as IQ scores increase, attention to the RW display is expected to decrease and attention to the VE display is expected to increase, resulting in a positive relationship between IQ and the attention ratio. The observed relationship between IQ scores and PQ scores is expected to be consistent with findings of previous research (Witmer & Singer, 1998; Kaber & Riley, 1999; Kaber et al., 2000).

CHAPTER IV

RESULTS AND DISCUSSION

A one-way ANOVA was performed with LOD as the independent variable and IQ ratings as the dependent variable. The ANOVA was used to determine whether IQ was confounded with LOD group assignment. That is, the results were expected to reveal whether randomly assigning subjects to the various experimental groups lead to a bias in, for example, assigning persons with high IQ scores predominantly to a particular group. No significant statistics ($F(2,21) = 0.13$, $p = 0.8824$) were observed in this analysis.

4.1 LOD Effects on Response Measures

The MANOVA was conducted first. Results of the MANOVA on all response measures revealed a lack of a significant effect of LOD ($F(2,21) = 1.6595$, $p = 0.1179$) based on Wilk's Lambda. The lack of significance observed on the family of response measures may be because the significance level for certain variables was very low. For example, the p-values associated with the effect of LOD on average overall SA (avgsa) and attention to the RW display (HSRW) were 0.8603 and 0.9257, respectively. As well, the F-ratios varied significantly across the various response measures. Individual ANOVAs were conducted on response measures to further evaluate influences of LOD. Table 3 summarizes the results of the univariate analyses.

4.1.1 Situation Awareness, Attention, and Workload

ANOVA results on SA, attention, and workload revealed no significant influence ($p>0.05$) of LOD. Analysis of subject SA was performed using overall average SA and average SA for each level, as defined by Endsley (1995). The difficulty level in the demining simulation was manipulated by varying the total number of mines and spacing between mines in the VE. Manipulating this component of the task increased the difficulty of locating mines, but did not, however, alter the information requirements of the task or the specific steps to be taken for mine neutralization. Changes in mine density also did not alter the elements of the task to be perceived, or, likely, how they were comprehended. Lack of significance of LOD on SA measures may be due to low correlation between the basis of the LOD manipulation and SA information requirements, especially given the content of items as part of the SAGAT questionnaires.

Table 3. Summary of results on LOD effects on response measures.

Response Measure	Predictor Variable
	LOD
Average performance (avgperf)	p = 0.0093 **
Telepresence (PQ)	p = 0.0214 *
Average SA (avgsa)	p = 0.8603
Average Level 1 SA (avgsa1)	p = 0.4054
Average Level 2 SA (avgsa2)	p = 0.8211
Average Level 3 SA (avgsa3)	p = 0.3319
Mental workload (mch)	p = 0.1730
Attention to VE (hsve)	p = 0.3770
Attention to RW (hsrw)	p = 0.9257
Attention ratio (attnrat)	p = 0.7172

* - significant at the $\alpha = 0.05$ level.
** - significant at the $\alpha=0.01$ level.

Regarding the lack of a LOD effect on attention to the VE, the real environment, or the ratio of the two, the range of difficulty or the differences between LODs may not have been large enough to affect subject ability to detect low battery signals in the VE or the RW. In general, subjects were able to detect on average 64% and 47% of the monitoring task signals in the VE and RW, respectively. It may be that at very low difficulty levels, for example 30-40 mines in the virtual space, subjects would have experienced decrements in monitoring task performance due to a need to frequently attend to mine neutralization subtasks. That is, an increase in the probability of finding mines may have required more active subject participation in the neutralization task, like manipulating the robotic arm, using demining tools, or identifying mine types. This might have reduced attentional resources available for detecting and responding to critical low power levels. Similarly, at higher LODs, for example 5 mines in the virtual space, subjects may have experienced increases in monitoring performance in the VE and/or RW. This could potentially be caused by a substantial decrease in active involvement in the teleoperation task. Very few mines in the virtual space would mean long durations of rover navigation as part of the neutralization task, and very little activity in robot manipulation and task object handling. In this case, the result might have been an increase in available attentional resources for detection of low battery signals, thus increasing performance in monitoring tasks. However, such conditions were not examined in this study and should be evaluated in future work in order to more clearly explain any important relationship between teleoperation task difficulty and human operator attention allocation.

It was hypothesized that LOD would significantly affect subjective mental workload. This hypothesis did not prove to be true. LOD was not significant in effect on ratings of mental workload as measured using the MCH ($F(2,21) = 1.91$ $p = 0.1730$). All subjects were able to complete the tasks during the trials. Subjects may have felt that mental workload levels during task performance were acceptable (low) because they were confident in their performance and able to successfully complete the task. It may also be that the MCH scale was not sensitive to the LOD manipulations evaluated in this research because it did not capture aspects of workload potentially important to task performance, including frustration or time pressure. A measure of frustration might have revealed how annoyed subjects were during mine neutralization task performance. For example, subjects in the high LOD group might have felt discouraged or irritated when having to navigate for long periods without quickly locating mines (as compared to subjects in the lower LOD groups). Pressure to find the mines quickly and complete the task in a timely manner might also have caused feelings of anxiety for subjects regarding the rate at which they completed each mine neutralization. Further, the MCH Scale is a subjective method requiring subjects to report their perceptions of workload. Subjects may not have been able to accurately report mental workload levels subsequent to task performance. Wickens (1992) states that a cost associated with subjective measures, like the MCH scale, is the uncertainty with which an operator's statement truly reflects the demand for mental resources. Subjects in this investigation may have been basing workload ratings on portions of the test trial (e.g., the mine neutralizations occurring at the end of the task) rather than the entire trial. They may not have been able to accurately recall all mine neutralization experiences in order to provide a workload rating for the whole trial. As well, 32 of the total 48 trials ended with detonation of a mine. The detonation task only required subjects to aim and discharge the shotgun to neutralize a mine, as opposed to manipulating the robot arm and vehicle to pick-up, grasp, remove, and transport a mine to the storage box. It is likely that subjects viewed a mine detonation as easier than a mine excavation. It is possible that subjects were not able to accurately recall all neutralizations and based workload ratings on these last mine neutralizations. Consequently, this may also be a reason that results revealed no LOD effect on MCH scores.

Another note of explanation regarding lack of LOD effects on workload ratings might be that the manipulations of LOD were not directly associated with all portions of the teleoperation task. The change in LOD was associated with the concentration of mines in the VE. Altering the number and spacing of mines did not alter the steps of the neutralization task or the total number of tasks to be completed simultaneously. The LOD manipulation likely only affected the search time, leaving all other task activities the same. Though an increase in LOD may have increased the search time, it did not change how the search task was carried out. Thus, reported workload ratings across groups were not significantly affected.

Because the MCH scale evaluates tasks on the basis of factors like accomplishability and number of errors, it was likely insensitive to other important workload factors like time pressure, frustration, and perceptual activity. For this particular manipulation of LOD, a more sensitive method of workload assessment may have been the NASA Task Load Index (TLX) (Hart & Staveland, 1988). The NASA-TLX evaluates workload along several different dimensions, including mental, physical and temporal demand, as well as performance, frustration, and effort. For example, the "Frustration" component of the NASA-TLX is aimed at assessing how "insecure, discouraged, irritated, and annoyed versus secure, gratified, content, and complacent" a subject felt during a task. Additionally, the mental demand scale is intended to describe how much perceptual activity was required like thinking, deciding, remembering, searching, etc., in addition to any other mental loading. Using this type of detailed and multi-factor scale might have resulted in more information regarding the perceived workload of subjects, and may have revealed more about the effects of LOD on perceived workload in this study.

4.1.2 Teleoperation Task Performance

ANOVA results on the average performance data under various LODs revealed a significant effect ($F(2,21) = 5.89$, $p < 0.05$) of LOD on the logarithmic transform of average performance. Duncan's MR test revealed that subjects performed the task significantly faster ($p < 0.05$) under the lowest LOD (mean = 312.84 sec.) than under the moderate (mean = 504.07 sec.) or high (mean = 525.91 sec.) difficulty levels, which were not significantly different. Figure 7 shows a graph of mean performance across LOD groups.

The effect of LOD on average performance was expected. As the number of mines increased and the distance between mines decreased, subjects located mines more quickly. The decrease in distance between mines increased the likelihood that subjects would drive the rover over a mine. Finding mines quickly likely meant less time on total neutralization, resulting in faster performance for subjects in the low LOD. Additionally, subjects did not appear to adopt any pattern of navigation in searching for mines, such as straight-line motions or concentric circles. Most subjects seemed to drive the rover in a random manner. This lack of strategy may have made it more difficult to locate mines at the high and moderate LOD levels. Failing to adopt a search strategy may have resulted in failure to adequately explore certain areas of the VE. For example, in the demining simulation, mines could be buried near the edges and in the corners of the virtual space. However, subjects often appeared to randomly drive the rover in the middle portion or open spaces of the VE. Failing to systematically search all areas of the virtual space may have meant failing to locate mines in the corners or along the edges, which could have substantially decreased chances of locating other mines spaced relatively far apart under the higher LODs. This may, ultimately, have resulted in subjects in the higher LOD taking much longer to find mines and performing the neutralization task significantly slower.

The lack of a significant difference in performance among the moderate and high LODs may be because the relatively low power value ($1-\beta$) for the F-tests for performance (time-to-mine-neutralization). The power level was approximately 0.50. Keppel (1982) suggests that a "more acceptable" level of power is 0.80 or higher. Thus, this test did not provide high sensitivity for detecting violations of the null hypothesis. The effect size for 24 subjects across the three LOD groups for average performance was 1.09, which corresponds to a medium to large effect (Cohen, 1977).

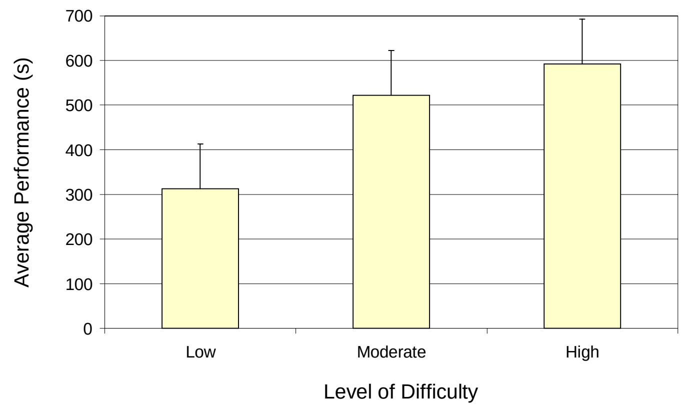


Figure 7. Graph of mean performance across LODs.

4.1.3 Telepresence

ANOVA results on telepresence scores agreed with the hypotheses stated in Chapter III. Results on the logarithmic transformation of PQ indicated a significant effect of LOD on subjective telepresence experiences during virtual task performance ($F_{(2,21)} = 4.69$, $p < 0.05$). Duncan's MR test revealed that as LOD was increased a decrease was observed in reported PQ scores. Subjects in the high (mean = 78.556) LOD group reported significantly lower ($p<0.05$) telepresence than subjects in the low (mean = 101.286) or moderate (mean = 100.316) LOD groups. The means of the low and moderate LOD groups were not significantly different. Figure 8 shows a graph of mean telepresence across LOD groups.

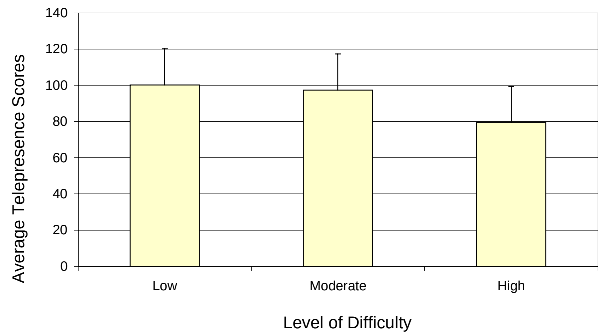


Figure 8. Graph of mean telepresence across LODs.

The PQ scores observed in this study are consistent with scores observed in previous research (Riley & Kaber, 1999). Subjects in low and moderate LOD groups reported high levels of telepresence, with averages of approximately 100 points out of a total of 133 points. The substantially lower telepresence experiences at the high LOD setting may be due to feelings of frustrations caused by an increase in difficulty of locating mines under this condition. The mines under the high LOD were placed in the VE such that each mine was at least 20 feet apart (more than three times the length of the rover). Frustration may have led to subject detachment from the task.

As mentioned earlier, subjects did not appear to develop any strategies for navigation. Aimless driving in the virtual space may have increased the time to locate mines, resulting in a greater portion of time-on-task in the VE without "active" robot control or object manipulation, such as using tools to uncover mines or manipulating the robotic gripper to excavate mines. As well, subjects may have felt unable to control events or affect the VE when mines could not be easily located.

During navigation, there was no auditory feedback unless a mine was detected or the rover collided with other objects. Extended navigation times under the high LOD without auditory feedback may have further contributed to decreased feelings of involvement in, or ability to perform the task.

Since subjects in this study were not experienced operators of teleoperated ordnance disposal systems, it is possible that they were skeptical about the realism of the task when mines were not easily located. This may have influenced their responses to questions regarding the compatibility of the virtual task and environment with a real situation.

Any of the issues mentioned above could have resulted in decreases in subject perceptions of association with the task and environment. The results presented here support the postulations of other researchers (e.g., Sheridan, 1992) that the difficulty/complexity of the virtual or teleoperation task significantly influences telepresence and is a major determinant of presence experiences. Consequently, if telepresence is to be used as a design criterion, how the complexity of the teleoperation task, specifically high LODs, influences subjects perceptions of the task should be considered.

The power value ($1-\beta$) for the F-tests for telepresence was determined to be approximately 0.65. The effect size for 24 subjects across the three LOD groups for telepresence was 1.26, which corresponds to large effects (Cohen, 1977).

4.2 Telepresence Regression Model

4.2.1 Full Model of Telepresence

The regression model used to assess the utility of SA and attention measures for describing telepresence is presented below and incorporates variable names defined earlier in this document.

$$\text{Telepresence} = \beta_0 + \beta_1 \text{LOD} + \beta_2 \text{AVGSA} + \beta_3 \text{ATTNRAT} + \beta_4 \text{IQ} + \varepsilon$$

The ANOVA results indicated that this model adequately describes telepresence experiences ($F(4,42) = 6.444$, $p < 0.05$, $R^2 = 0.3803$). The coefficient of determination indicates that the multiple least squares regression line produced by the model explains approximately 38% of the variability in the PQ scores reported by subjects. The results also indicated that at the α=0.001 level, LOD and IQ were significantly related to PQ scores.

4.2.1.1 Multicollinearity Analysis

As previously stated, there was an expected potential for multicollinearity in the model, as IQ, SA, and attention are hypothesized to be related to one another. Multicollinearity analysis was conducted using the SAS software to reveal any evidence of collinearity among regressors in the full model. Results of the diagnostics showed no evidence to support existence of collinearity among the predictor variables. All tolerance values were close to “1”. All VIF values were below “10” and close to “1”. All condition index (CI) values were outside of the critical range. No variables were removed from the model to account for multicollinearity.

4.2.1.2 Residual Analysis and Influential Diagnostics

Residual analysis and influential diagnostics were conducted on the full regression model. Examination of residual plots revealed no violations of regression assumptions in terms of normality or linearity. However, there was some evidence of non-constant variance. Thus, a log transform was applied to PQ responses to account for assumption violations.

On the basis of influential diagnostics, outliers were removed from the data set. All of the points removed from the data set were identified as outliers based on values reported for the COVRATIO statistic and studentized residuals provided by the SAS regression output. The cutoff values for the COVRATIO statistic in this study were 1.25 for the upper bound and 0.75 for the lower bound. Values exceeding 2.0 in magnitude were considered suspect for the studentized residuals. The cutoff value for the Bonferroni method would have been 3.52. (The Bonferroni cutoff value is referenced from Table C4 in Myers' (1990) text on page 481.) Additionally, information and data for all subject observations was reviewed to provide explanation of suspected influential points. Removal of influential data did not involve exclusion of all data points for any particular subject. All outliers were removed subsequent to transformation of the data set.

Three points were identified and removed from the data set. The characteristics of these outliers are listed in the Table 4.

Table 4. Characteristics of influential points.

Subject	LOD	Observation Number
11	High	21
13	Moderate	25
14	Low	28

There were no recorded experimental anomalies, such as an error in measurement or a system failure, to account for the influential points. However, it was observed that some subjects appeared to have navigation problems during some trials. For example, Subject 13 and Subject 14 used viewpoints that were not conducive to the search and uncover tasks. They did not initially appear to be aware of this, causing them to take longer to locate mines and potentially influence their reported PQ scores. Review of subject data revealed that Subjects 13 and 14 produced individual mine neutralization times that were comparatively longer than all other mine neutralizations in a particular trial. Subject 13 completed one mine in about 790 seconds, while all other mine neutralizations in the same trial took on average 350 seconds. Similarly, Subject 14 completed one mine in about 420 seconds. All other mine neutralizations in the same trial took on average 170 seconds.

Also, Subject 11 appeared to have difficulty orienting him/herself in the VE (e.g., in relation to other objects like the storage box or trees). It is possible that this subject had low spatial abilities, which may have affected his/her ability to navigate and find mines, as well as reported PQ scores. Subject 11 produced one mine neutralization time (for a single mine) that was approximately 1139 seconds. All other mine neutralization times for this subject during the same trial were substantially lower (588, 330, and 438 seconds). The unusually long time for a single mine inflated the overall average performance time for the trial. The extended trial time could have affected the reported telepresence score.

4.2.2 Model of Logarithmic Transform of Telepresence Scores

Following these analyses, the regression procedure was conducted along with the forward selection technique on a model of transformed PQ scores. This was done to evaluate the adequacy of the model of transformed variables and identify significant predictor variables for inclusion in a model of telepresence. ANOVA results on the model of the logarithmic transform of PQ are summarized in Table 5.

Table 5. Summary of ANOVA results on Log(PQ) model.

Analysis of Variance on Model Adequacy					
Source	df	Sum of Squares	Mean Square	F-Value	Prob > F
Model	4	0.70255	0.17564	10.623	0.0001
Error	39	0.64484	0.01653		
Corrected Total	43	1.34739			
t-tests on Individual Parameter Estimates					
Variable	Df	Parameter Estimates	Standard Error	T-value for H_0	Prob > T
Intercept	1	4.391981	0.20532802	21.390	0.0001**
LOD	1	-0.117778	0.02394995	-4.918	0.0001**
AVGSA	1	-0.336406	0.18110554	-1.858	0.0708
ATTNRAT	1	0.071746	0.02895285	2.478	0.017*
IQ	1	0.006401	0.00160262	3.994	0.0003**

*- significant at the α=0.05 level
*- significant at the α=0.001 level

The results indicate that the model of Log(PQ) in LOD, AVGSA, ATTNRAT, and IQ adequately describes telepresence experiences ($F(4,39) = 10.623$, $p < 0.05$, $R^2 = 0.5214$). The coefficient of determination of this model reveals that approximately 52% of the variability in the logarithmic transform of PQ is explained by the multiple least squares regression line produced by the model. The results also indicate that at the $\alpha = 0.05$ level, LOD, ATTNRAT, and IQ are significantly related to PQ scores. AVGSA is marginally significant, with a p-value of 0.0708. Table 6 provides an interpretation of the parameter estimates for the Log(PQ) model of telepresence.

4.2.2.1 Interpretation of Model Parameters

Table 6. Summary of interpretation of parameter estimates of Log(PQ).

Variable	Interpretation
Intercept	The PQ score when all other variables have a value of zero will be equal to 4.391981. This value is not of interest, because it is not realistic to expect an observation where all independent variables take on a value of zero.
LOD	PQ scores will decrease by -0.117778 units if LOD increases by one unit and all other regressors remain constant. Thus, as LOD increases, PQ scores (telepresence) will decrease. This result is consistent with the research hypotheses and the trend of results observed through the PROC GLM procedure.
AVGSA	PQ scores will decrease by -0.336406 units if AVGSA increases by one unit and all other regressors remain constant. When AVGSA on the VE task increases (i.e., there is an increase in percent correct responses to queries), PQ decreases. This result is counter to the research hypotheses.
ATTNRAT	PQ scores will increase by 0.071746 units if the ATTNRAT increases by one unit and all other variables remain constant. Increasing the amount of attention allocated to the VE increases experiences of telepresence in the VE. This result is consistent with the research hypotheses.
IQ	PQ scores will increase by 0.006401 units if IQ scores increase by one unit and all other variables remain constant. As subject susceptibility to immersion increases, PQ scores, reported following task performance, increases. This result is consistent with the research hypotheses.

As expected, LOD had a negative relationship with PQ. The model indicates that as LOD increases by a value of one unit, PQ will decrease. These results are consistent with the finding of the PROC GLM procedure and further support the hypothesis that task difficulty is a determinant of telepresence experiences.

The results on average SA are surprising, as it was expected that increases in SA would result in increases in telepresence experiences. The negative sign associated with the SA parameter estimate may be due to the design of the virtual task interface. For example, in assessing SA, subjects were queried regarding current tool use; tools to be used in the future; types of mine; etc. To achieve SA on these elements of the task, subjects often needed to alternate between viewpoints. For example, to clearly view the tool display provided on the rover, subjects had to select a specific viewpoint of the VE (i.e., an egocentric view encompassing a large portion of the rover, the surrounding environment and the tool display). Using the interface controls (i.e., mouse clicking) to toggle between the various viewpoints to develop SA on components of the tasks may have detracted from task involvement. As well, subjects may have experienced a decreased sense of involvement in the task when attempting to maintain global SA (e.g., remembering the current stage of the task, the total number of mines that have been successfully neutralized, etc.). This might have resulted in reductions in subject perceptions of association with the VE, and it may explain the observed decreases in PQ scores with corresponding increases in SA.

The results on this model indicate that as the ratio of attention allocation across the VE and RW increases, telepresence scores increase. Increases in the attention ratio are indicative of increases in the hit-to-signal ratio in the virtual monitoring task (or a decrease in monitoring performance in reality). An increase in the attention ratio was observed as subjects allocated more attention to the VE and performance in the signal detection task increased. The regression model parameter estimate associated with the ratio of attention is consistent with the hypotheses formulated at the outset of this investigation. The model predicts that the better subjects perform in signal detection in the VE, the higher reported telepresence scores will be. It is possible that the VE was successful in engaging subjects in the virtual teleoperation task. Immersion in the task may have contributed to allocation of attention to the VE over the RW, resulting in increased probability of detecting signals in the VE (or decreased probability of detecting signals in reality). These might be the causes of stronger telepresence experiences in the experimental task.

IQ scores were significantly related to PQ scores, with PQ scores increasing as IQ increased. Results indicate that as subject susceptibility to immersion increased the strength of telepresence experiences also increased. This relationship is consistent with previous research (Witmer & Singer, 1994; Riley & Kaber, 1999; Kaber et al., 2000). The positive value of the model coefficient observed for IQ was expected.

4.2.2.2 Forward Selection Technique

Results of the forward selection algorithm indicate that at the $\alpha=0.05$ level of significance, LOD and IQ were selected for inclusion in the final model of telepresence. The selection procedure selects LOD for inclusion in the model in the first step of the algorithm. Immersive tendency (IQ) was subsequently added. The model was found to adequately describe telepresence ($F\ (2,41) = 15.27$, $p < 0.05$, $R^2 = 0.4269$, with the predictors explaining approximately 43% of the variance in telepresence. A summary of the forward selection results is provided in Table 7. A summary of the ANOVA results on the final model is provided in Table 8.

It appears that the absence of SA in the model of telepresence when using the forward selection procedure lead to a decrease in the significance of attention in the model. ATTNRAT was not included in the model of telepresence with this technique, even though it was significant in the full model. To further investigate this result, a regression analysis was conducted on a model of telepresence including only LOD, IQ, and the attention measure. This was done to determine if the absence of AVGSA in the model would have an effect on the adequacy of the model or the parameter estimates. Results of an ANOVA revealed that the model was adequate in terms of predicting telepresence ($p > 0.05$ and $R^2 = 0.4791$). Only LOD and IQ were significantly related to PQ scores. The attention variable was marginally significant in this model with a p-value of 0.0523. No major changes, however, were observed in the parameter estimates of the other regressor variables with AVGSA removed from the model. It may be that in the absence of AVGSA, some of the variance attributed to that regressor was pooled into the error term and resulted in inflation of the F-test denominator, causing the significance level of ATTNRAT to decrease.

Table 7. Summary of results of forward selection technique.

Step	Variable Entered in Model	Partial R^2	Model R^2	C(p)	F-Values	Prob > F
1	LOD	0.2776	0.2776	18.8692	16.1392	0.0002
2	IQ	0.1494	0.4270	8.6970	10.6872	0.0022

Table 8. Summary of ANOVA results on final telepresence model.

Analysis of Variance on Model Adequacy					
Source	df	Sum of Squares	Mean Square	F-Value	Prob > F
Model	2	0.57528889	0.28764444	15.27	0.0001
Error	41	0.77210076	0.01883173		
C Total	43	1.34738965			
F-tests on Idividual Parameter Esitmates					
Variable	df	Parameter Estimates	Standard Error	F-value for H_0	Prob > F
Intercept	1	4.29401353	0.14726545	850.21	0.0001
LOD	1	-0. 11448542	0.02549413	20.17	0.0001
IQ	1	0.00543187	0.00166156	10.69	0.0022

4.2.3 Final Telepresence Model

The final telepresence model can be written in variable terms as:

$$\text{Log(PQ)} = 4.29401353 + -0.11448542\ (\text{LOD}) + 0.00543187\ (\text{IQ})$$

This model can be used to quantify telepresence experiences in terms of LOD and immersive tendencies. The parameter estimates indicate that changes in LOD will have the largest effect on telepresence. An increase in LOD will degrade telepresence experiences. With this in mind, if telepresence is used as a design criterion, designers should consider the complexity/difficulty of the task and devise ways that might counter ill effects on telepresence observed with substantial increases in LOD. For example, for the task used in this study, automating subtasks under the high LOD level, like navigation of the rover in the virtual space or manipulation of robot arm, might decrease overall task difficulty. As well, team performance might help by distributing task responsibilities and possibly lower negative effects on PQ scores.

Changes in IQ scores appear to have a smaller effect on telepresence as predicted by the model. IQ scores are significantly related to PQ scores; however, it appears that susceptibility to becoming engrossed in the VE task and environment must increase substantially to improve telepresence (or, perhaps, to increase the likelihood of experiencing telepresence). This suggests that it may not be very helpful to select operators on the basis of a tendency to immersion. However, it should be noted that subjects in this study reported somewhat low IQ scores, with an average of approximately 83 out of a total of 126, as compared to IQ scores observed in previous research (Riley & Kaber, 1999; Kaber et al., 2000).

4.3 Correlation Analysis

Results of the correlation analysis are presented provided in Table 9. The table includes the Pearson Product-Moment Coefficients for all pairs of response measures on which observations were collected. The correlation coefficients are presented along with p-values and indicate the strength of the linear associations between untransformed responses. Figures provided in this section illustrate only those significant relationships that were of interest to this investigation (see the statistical analysis section in Chapter III). These include: (1) telepresence (PQ scores) and performance (time-to-mine-neutralization); (2) telepresence and workload (MCH ratings); (3) performance and workload; (4) telepresence and immersive tendencies (IQ scores); (5) SA (percent correct responses to queries) and attention (hit-to- signal ratio); and (6) immersive tendencies and attention. Table 10 provides results on partial correlation analyses for all pairs of response measures. The results indicate significant relationships among variables, with individual differences across subjects taken into account. In all correlation data sets and the resulting plots, influential points and outliers were identified and removed. Significance of the correlations remained after removal of data points.

Correlation analyses on IQ scores with other experimental variables were conducted on a trial-by-trial basis. This was done because subjects only provided one response to the IQ. Therefore, only 24 observations on IQ were available, whereas 48 observations were recorded on all other variables. Thus, the 24 observations for IQ were associated with the 24 response observations collected for Trial 1 and 24 observations collected for Trial 2, separately. Results of this analysis are provided in Table 11.

Table 9. Correlation matrix for response measures.

Response Measure	Avgperf	Avgsa	Avgsa1	Avgsa2	Avgsa3	MCH	Attnratio	HSVE	HSRW
PQ	r =-0.3270 *p* =0.001	r =0.04031 *p* =0.5798	r =0.05387 *p* =0.4593	r =-0.1487 *p* =0.0400	r =0.03057 *p* =0.6747	r =-0.6247 *p* =0.0001	r =0.1992 *p*=0.7849	r =-0.1695 *p* =0.0190	r =-0.1036 *p* =0.1536
Avgperf		r =-0.2240 *p* =0.0018	r =-0.2677 *p* =0.0002	r =0.0205 *p* =0.7783	r =0.8198 *p* =0.2566	r =0.5692 *p* =0.0001	r =-0.1045 *p* =0.1501	r =-0.1315 *p* =0.0697	r =-0.0805 *p* =0.2679
Avgsa			r =0.6214 *p* =0.0001	r =0.8164 *p*=0.0001	r =-0.0633 *p* =0.3619	r =-0.2423 *p* =0.0007	r =0.1998 *p* =0.0056	r =0.2235 *p* =0.0019	r =-0.1276 *p* =0.0784
Avgsa1				r =0.3197 *p* =0.0001	r =-0.1707 *p* =0.0182	r =-0.1947 *p* =0.0070	r =0.1710 *p* =0.0180	r =0.2015 *p* =0.0052	r =-0.1311 *p* =0.0706
Avgsa2					r =-0.0362 *p* =0.6787	r =-0.0934 *p* =0.1985	r =0.1614 *p* =0.0257	r =0.0888 *p* =0.2217	r =-0.1232 *p* =0.0893
Avgsa3						r =-0.0225 *p* =0.7569	r =0.0090 *p* =0.9007	r =0.0127 *p* =0.5606	r =-0.0336 *p* =0.6443
MCH							r =-0.0706 *p* =0.3317	r =0.0782 *p* =0.3317	r =0.0715 *p* =0.3254
Attnratio								r =0.2837 *p* =0.0001	r =-0.8820 *p* =0.0001
HSVE									r =-0.0552 *p* =0.4480

Table 10. Partial correlation matrix for response measures.

Response Measure	Avgperf	Avgsa	Avgsa1	Avgsa2	Avgsa3	MCH	Attnratio	HSVE	HSRW
PQ	r =-0.3365 *p* =0001	r =0.0376 *p* =0.6067	r =0.0613 *p* =0.4002	r =-0.1479 *p* =0.0416	r =0.0315 *p* =0.6661	r =-0.6338 *p* =0.0001	r =0.0213 *p*=0.7699	r =-0.1717 *p* =0.0178	r =-0.1052 *p* =0.1486
Avgperf		r =-0.1823 *p* =0.0118	r =-0.2235 *p* =0.0019	r =0.0424 *p* =0.5612	r =0.0956 *p* =0.1891	r =0.5574 *p* =0.0001	r =-0.0893 *p* =0.2204	r =-0.1579 *p* =0.0353	r =-0.0984 *p* =0.1784
Avgsa			r =0.5886 *p* =0.0001	r =0.8221 *p*=0.0001	r =-0.0846 *p* =0.2437	r =-0.2133 *p* =0.0031	r =0.1835 *p* =0.0113	r =0.2582 *p* =0.0083	r =-0.1109 *p* =0.1276
Avgsa1				r =0.3030 *p* =0.0001	r =-0.1950 *p* =0.0062	r =-0.1595 *p* =0.0366	r =0.1517 *p* =0.0366	r =0.2416 *p* =0.0008	r =-0.1127 *p* =0.1214
Avgsa2					r =-0.0428 *p* =0.5576	r =-0.0790 *p* =0.2785	r =0.1533 *p* =0.0347	r =0.0996 *p* =0.1712	r =-0.1158 *p* =0.1116
Avgsa3						r =-0.0142 *p* =0.8453	r =0.0039 *p* =0.9569	r =0.0182 *p* =0.8832	r =-0.0291 *p* =0.6896
MCH							r =-0.0584 *p* =0.4230	r =0.0659 *p* =0.3663	r =0.0608 *p* =0.4046
Attnratio								r =0.2943 *p* =0.0001	r =-0.8812 *p* =0.0001
HSVE									r =-0.0629 *p* =0.3885

Table 11. Correlation matrix for IQ with other response measures.

IQ Correlations				
	Trial 1		Trial 2	
Response Measure	Simple	Partial	Simple	Partial
PQ	r =0.0110 *p* =0.9578	r =0.30149 *p* =0.1621	r =0.38950 *p* =0.0662	r =0.39203 *p* =0.0712
Avgperf	r =-0.02346 *p* =0.9133	r =-0.00785 *p* =0.9717	r =0.0376 *p* =0.8892	r =0.04346 *p* =0.8477
Avgsa	r =0.14687 *p* =0.4934	r =0.13084 *p* =0.5518	r =-0.16161 *p* =0.4613	r =-0.18150 *p* =0.4189
Avgsa1	r =-0.01521 *p* =0.9438	r =-0.08316 *p* =0.7060	r =-0.28733 *p* =0.1837	r =-0.28315 *p* =0.2016
Avgsa2	r =-0.00445 *p* =0.9835	r =0.00345 *p* =0.9875	r =-0.18991 *p* =0.3854	r =-0.2096 *p* =0.3491
Avgsa3	r =0.30657 *p* =0.1452	r =0.29678 *p* =0.1691	r =-0.03245 *p* =0.8831	r =-0.04351 *p* =0.8475
MCH	r =0.10239 *p* =0.6340	r =0.11046 *p* =0.6159	r =-0.10236 *p* =0.6421	r =-0.09330 *p* =0.6796
Attnratio	r =-0.34912 *p* =0.0945	r =-0.36129 *p* =0.0903	r =-0.03823 *p* =0.8625	r =-0.04428 *p* =0.8449
HSVE	r =0.02427 *p* =0.9104	r =0.03390 *p* =0.8779	r =-0.04437 *p* =0.8407	r =-0.04523 *p* =0.8416
HSRW	r =0.37537 *p* =0.0707	r =0.38471 *p* =0.0699	r =0.05287 *p* =0.8106	r =0.05739 *p* =0.7797

4.3.1 Telepresence and Teleoperation Task Performance

PQ scores were found to be significantly, negatively related to average time-to-mine-neutralization (r = -0.32720, $p < 0.05$). Similar results were obtained with the partial correlation procedure (r = -0.3365, $p < 0.05$). A graph of telepresence versus performance is shown in Figure 9. It is important to note that the graph and correlation analysis indicate a positive relationship between telepresence and the construct of performance. As the subject performance increased (i.e., as the time to complete neutralization tasks decreased), subject association and involvement with the virtual task increased. This finding is consistent with the research hypotheses and agrees with previous research (Kaber et al., 2000; Riley & Kaber, 1999). These results suggest that telepresence may play a critical role in teleoperation task performance (or vice versa). They offer further support to the contention that telepresence should be examined as a potential design criterion in teleoperation (along with other factors like LOD). Thus, more studies are needed and should be designed to clearly elucidate what the relationship is between telepresence and performance.

4.3.2 Telepresence and Cognitive Workload

Telepresence was significantly, negatively correlated with mental workload ($r = -0.62477$, $p < 0.05$). Similar results were revealed through the partial correlation procedure ($r = -0.6338$, $p < 0.05$). A graph of telepresence versus mental workload is shown in Figure 10. As ratings on the MCH Scale increased, the reported PQ scores decreased. Thus, as subjects' perception of cognitive loading increased, their association with the VE and involvement in the task was degraded. This result is consistent with the research hypotheses and past research (Riley & Kaber, 1999; Kaber et al., 2000). The negative relationship between telepresence and mental workload may have been due to perceived frustrations and time pressure to complete the mine neutralization task quickly. Given the relationship observed between PQ scores and performance, and the relationship observe between PQ and workload, it may be necessary for researchers and system designers to develop teleoperators that alleviate cognitive loading on the user. For example, in the task presented as part of this work, automating the monitoring task in the RW or the VE might free-up user mental resources, and possibly reduce frustrations due to attempts to successfully complete all three tasks simultaneously. Such designs might help to increase perceptions of task accomplishability and serve to increase telepresence experiences in the teleoperation task. They could potentially also lead to improved performance.

4.3.3 Task Performance and Cognitive Workload

As hypothesized, average time-to-mine-neutralization was significantly related to MCH ratings (r = 0.56929, p < 0.05). Partial correlation analysis results were similar (r = 0.5574, *p* <0.05). They indicate that subject performance degraded (i.e., time to complete neutralizations increased) as perceptions of mental workload increased. This finding is consistent with previous research in simulated teleoperation tasks (Kaber et al., 2000). As discussed earlier, automating portions of the task might help to alleviate mental workload and serve to improve teleoperation task performance.

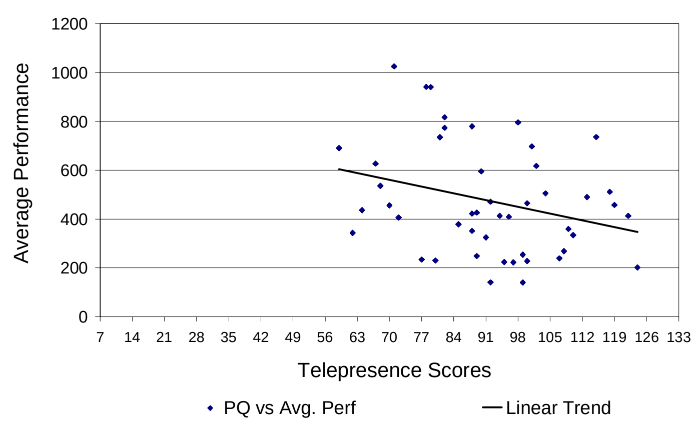


Figure 9. Telepresence scores versus average performance.

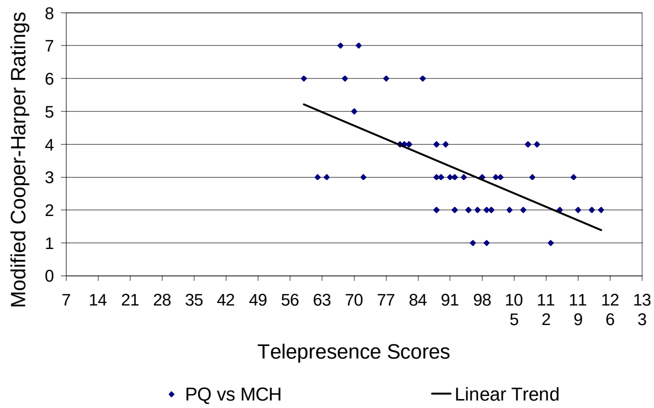


Figure 10. Telepresence scores versus Modified Cooper-Harper ratings.

4.3.4 Telepresence and Immersive Tendencies

The correlation results on telepresence versus immersive tendencies revealed marginal significance for the second experiment trial. The observed trend was consistent with previous research results (Witmer & Singer, 1994; Riley & Kaber, 1999; Kaber et al., 2000). Presence Questionnaire scores and IQ scores were positively related ($r = 0.38950$, $p = 0.0662$). Partial correlation analysis results on this relationship were similar ($r = 0.39203$, $p = 0.0712$). As IQ scores collected prior to task performance increased, PQ scores reported by subjects at the end of the second trial. A graph of immersive tendencies versus telepresence for Trial 2 is shown in Figure 11. This result is important because susceptibility to immersion in a VE could potentially be used as an operator selection criterion in attempts to increase the probability of telepresence experiences. It may be that some persons are more able to “suspend their disbelief” that a VR is reality and view a simulation as a place that they occupy, rather than just a computer generated world (Nash et al., 2000). Barfield & Weghorst (1993) suggest that this is a necessity for telepresence experiences to occur. Selection of persons as operators of teleoperation systems based on IQ scores could potentially increase telepresence experiences, and thus performance on tasks. In light of the regression results (i.e., the parameter estimate associated with IQ), it might be difficult to select participants on the basis of IQ scores, however, and obtain substantial increases in PQ ratings.

The difference between Trial 1 and Trial 2 in terms of the relationship of immersive tendencies and presence scores may be due to the duration of subject exposure to the VE. The correlation results indicate that PQ scores increased for the second trial, resulting in a significant relationship between IQ and PQ for that trial. This result is consistent with findings from previous research (Riley & Kaber, 1999). This research has demonstrated that PQ scores increased with increased exposure to the virtual task. Thus, subjects may have become more engaged in task performance during the second trial because they had become accustomed to the VE and task situation. This would explain the higher observed PQ scores for Trial 2, although the IQ scores remained the same.

4.3.5 Situation Awareness and Attention

Average SA measured in terms of the percent correct responses to all SA queries was positively correlated ($r = 0.19987$, $p < 0.05$) with the ratio of attention allocation across the VE and RW. Partial correlation analysis results were similar ($r = 0.1835$, $p = 0.0113$). The ratio increases as monitoring performance in the VE increases (or RW monitoring performance decreases). Thus, as attention to the VE increased, the percent correct responses to SA queries regarding teleoperation task performance increased. A graph of SA versus attention is provided in Figure 12. This finding agrees with the research hypotheses and statements in other research that describes SA as dependent upon, attention allocation to tasks (Endsley & Jones, 1997). This result is also supported by correlations between average SA and attention to the VE (HSVE) and attention to the RW (HSRW) (see the correlation matrix in Table 9.). The relationship revealed through the correlation analysis indicates that as subjects choose to allocate attention to the VE, they are better able to achieve a picture of the task/system situation for higher SA. Increases in the attention ratio may also be due to involvement of subjects in the teleoperation task or due to attempts to excel in VE monitoring task. Subjects may have been able to better detect signals in the VE, because they were engaged in the teleoperation task and the VE monitoring signal was superimposed over the VE display.

4.3.6 Immersive Tendencies and Attention

Contrary to expectations, IQ scores were marginally, negatively correlated with the attention ratio ($r = -0.34912$, $p < 0.0945$) for the first experimental trial. Similar results were obtained through partial correlation analysis ($r = -0.36129$, $p < 0.0903$). It was hypothesized that as IQ increased, attention to the RW display would decrease, causing an increase in attention ratio. However, as IQ increased, a decrease in attention ratio was observed. A graph of the immersive tendencies and attention relationship for Trial 1 is provided in Figure 13. It is possible that even though subjects were susceptible to immersion, they chose to divide attention across the VE and RW rather than focus primarily on the VE display at the expense of RW monitoring performance, resulting in ability to detect about 47% (on average) of the signals presented in the RW. That is, they may have deemed it detrimental to overall performance (on the three tasks) to focus primarily on one display. Studies have shown that subjects can concurrently attend to non-contiguous locations, as long as new objects do not appear between the attended areas (Hahn & Kramer, 1998). Subjects choosing to divide attention across the two displays may also be because, in this investigation, the monitoring tasks were defined as embedded secondary tasks. Subjects in this experiment were not provided with instructions to emphasize performance in any one task.

The statistical relationship between IQ and the attention ratio observed in the first trial may be attributed to subjects initially allocating attention to the experimental tasks because they were novel experiences. It is possible that subjects reporting low IQ scores were impressed with the simulation at first sight, causing them to attend to the primary task display and detect the low battery signals in the VE. However, after the first trial, the subjects' interest in the tasks may have decreased. It is likely that even though the subjects' IQ scores were the same for Trial 2, they were less compelled to performance by the display and allocated fewer attentional resources for signal detection. Consequently, no significant relationship was observed in the second trial.

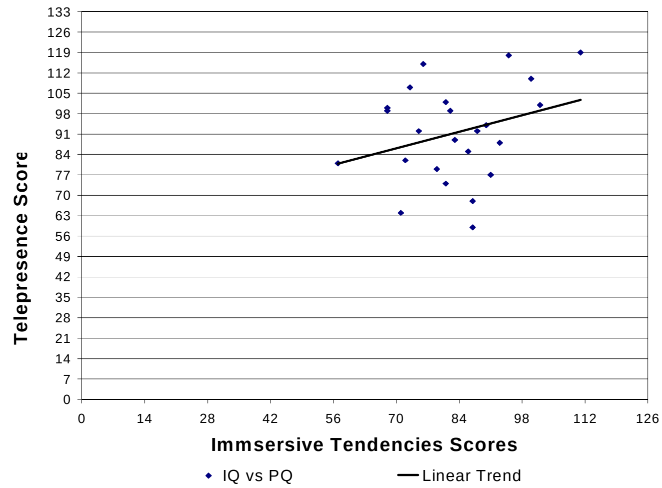


Figure 11. Immersive tendencies versus telepresence, Trial 2.

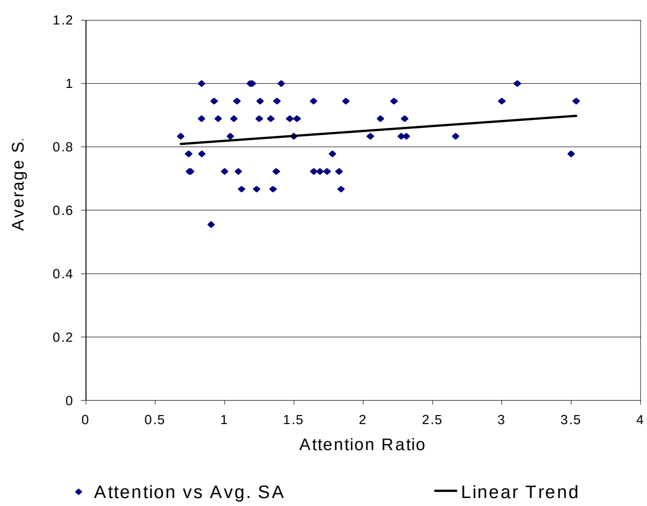


Figure 12. Attention ratio across VE and RW versus overall average SA.

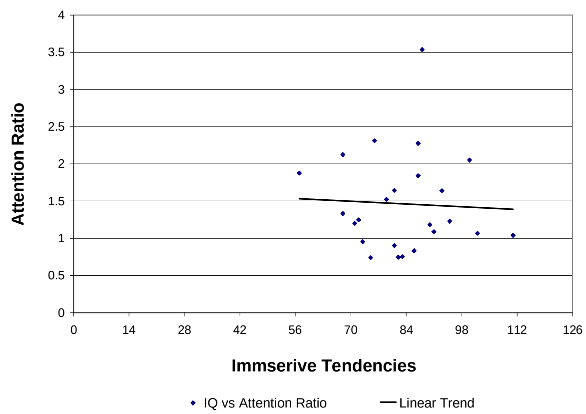


Figure 13. Immersive tendencies versus attention ratio across VE and RW, Trial 1.

CHAPTER V

CONCLUSIONS

This study has provided useful results on the impact of level of task difficulty on human performance and telepresence in a teleoperation task. It has also provided preliminary results on, and insight into, the utility of measures of SA and attention (along with other variables) for explaining telepresence. Previous studies have not investigated a multiple linear regression model of telepresence in an effort to identify new objective measures of the concept based upon existing measures of cognitive constructs. Further, the study has provided insight into relationships between telepresence and other teleoperation task measures. In particular, the study provided further evidence of a relationship between teleoperation task performance and telepresence.

5.1 Limitations of Results

One limitation of this study is that it presents results on a simulated teleoperation task posed to subjects through a computer system. In addition, the task was completed by naive, college-age participants rather than experienced, qualified operators of teleoperation systems. The level of involvement of naive subjects in a simulated task is likely less than that which might be observed in a real-world demining situation with real consequences associated with teleoperator control actions (e.g., loss of expensive equipment due to accidental detonation of mines). It is possible that results on telepresence, performance, workload and other measures recorded on trained operators might follow different patterns than those observed for the naive subject population. However, it should be noted that the subjects used in this research were trained in all tasks to ensure their ability to perform the task. All subjects were also required to successfully complete the training session in order to proceed with the testing procedures.

The final limitation of this study is that the results may be limited to the specific task and environment simulated in VR. Thus, care must be taken in generalizing results to other types of teleoperation tasks. It should be noted, however, that even though the steps of the task were very specific to the neutralization of land mines, the task included activities that are required in other common teleoperation tasks (e.g., motion control of a robotic manipulator, robotic vehicle navigation/environment exploration, etc.).

5.2 Implications and Future Research

Results of this study indicate that LOD is significant in its effect on telepresence and task performance, effectively serving to degrade both during the simulated teleoperation task. Specifically, at the high difficulty level, telepresence and performance levels were significantly less than observed at the low LOD. Sheridan (1992) has suggested three major determinants of telepresence, which are critically influenced by task factors like difficulty: (1) extent of sensory information, (2) ability to affect remote environment, and (3) control of sensors in the remote environment. Draper et al. (1998)

also suggested that complexity of the task affects telepresence experiences. The results observed in this research support the hypothesis that task difficulty is a factor that influences the extent to which users associate themselves with a VE and/or task. It has been speculated that as the difficulty of the task increases, operator concentration on the task might increase and cause telepresence to increase (Kaber & Riley, in press). It is also possible that the opposite trend might occur. That is, as the complexity of the task increases, workload demands increase, causing users to become frustrated and detached, resulting in degradation of telepresence. The latter case was observed in this study. This effect is consistent with previous research (Kaber & Riley, in press) demonstrating that increases in task difficulty, as manipulated through Fitt's Index of Difficulty, serve to significantly degrade task performance.

An interesting note regarding telepresence and performance results across LODs is that the moderate LOD group essentially reported the same telepresence experiences as the low LOD group, but achieved performance levels that matched those of the high LOD group. These results may indicate that the threshold of LOD causing a significant effect on telepresence and performance differs across the two response measures. That is, in this study, the magnitude of the change in LOD that caused significant differences in mine neutralization performance from the low to moderate levels of difficulty was not equivalent to the change in LOD that would have been required to significantly affect PQ scores. Thus, although LOD had a significant effect on both performance and telepresence, that effect was not identical. This finding points to the potential complexity of the relationship between telepresence and performance. As with LOD effects, the effects of other factors, like length of exposure or extent of task automation, may share similar relationships with telepresence and performance and, yet, affect the two in different ways. This may, consequently, contribute to the difficulty in elucidating the true relationship between telepresence and performance in teleoperation tasks.

It was surprising to find that LOD had no effect on SA, attention, or workload in this study. With respect to SA and attention, the range of LOD may not have been substantial enough for effects to be revealed. As for the lack of effect on workload, the MCH scale is subjective in nature and required subjects to rate workload after task performance. It is possible that the subjects were not able to adequately recall and assess the level of mental loading they experienced during the neutralization task at the end of each trial. Further, it may be that the workload scale used here was not sensitive to factors like frustration and time pressure, which serve to affect cognitive loading.

In general, the effects of LOD on telepresence and performance suggest that it should be considered in attempts to further understand telepresence. Level of difficulty in this study was manipulated in terms of mine concentration (or ease of finding mines). The LODS were arbitrarily set. Future research might consider a range of LOD with more levels, perhaps from 5 to 40 mines at 5 point intervals. This could serve to better explain the trend of LOD on performance and telepresence. That is, there might be lower LODs (e.g., more than 20 mines in the VE) that actually serve to degrade telepresence or higher levels of difficulty that increase telepresence. Varying difficulty along different dimensions might also be of interest (e.g., number of tasks to be completed, amount of time to complete tasks, etc.)

Establishing workload levels, at which operators become cognitively incapable of coping with teleoperation tasks, might help researchers to design task elements to decrease workload. For example, research might be conducted to determine if automation of task components could help to moderate operator workload. Further, research could be conducted to determine which aspects of task performance should be automated to observe the largest decrease in operator cognitive loading.

Results of the regression model analysis suggest that SA (marginally) and attention may serve to indicate telepresence experiences in a teleoperation task. However, they do not appear to be the only variables that must be considered when attempting to quantify telepresence. (The LOD and pre-existing subject immersive tendencies were also significant in relationship to telepresence.) The results show that as attention is allocated to the VE (over the RW), telepresence increases. This finding is consistent with the hypothesized relationship between telepresence and monitoring task performance. It may be that operator attention can be guided to the VE in order to increase the possibility of telepresence experiences. Research in this area might consider the design of experiments to determine if there is some minimal amount of attention that must be allocated before telepresence can occur. Also, other work might consider how attention could be directed to the VE in situations where subjects are less engaged.

Development of (or increasing) SA on the VE appeared to negatively affect telepresence, as described by the regression model. This effect was counter to hypotheses. It appears that achieving SA subtracts from telepresence experiences in the current task and experimental setup. That is, as subjects acquire more information on elements of the environment and status of the task, association with the task is decreased. Results of the correlation analyses (see Table 9) indicate that the overall negative effect of average SA on telepresence may be driven by the relationship between telepresence and Level 2 SA. Thus, the comprehension of information on the task may have served to degrade telepresence. Similar correlation results were found for Level 2 SA and performance. It may be possible that operator information gathering and relation of perceived task to overall task performance increased the time subjects needed to complete the neutralization task and subtracted from subject association with the VE. Using mental resources for development and maintenance of Level 2 SA may have resulted in significantly fewer resources available for telepresence. Future research may need to evaluate the design of VE interfaces and controls in terms of the capability to meet operator information requirements for development of SA, and assess how mechanisms for acquiring affect telepresence and performance. Future research in this and any of the areas discussed above is needed to explore teleoperation system designs and task factors in order to establish telepresence-based design guidelines and maximize overall human-system performance.

REFERENCES

Barfield, W. and Weghorst, S. (1993). The sense of presence within virtual environments: a conceptual framework. In G. Salvendy and M. Smith (Eds.), Human computer interaction: Software and hardware interfaces, 699 – 704. Amsterdam: Elservier.

Battelle, J. and Ridao, P. (1999). Mobile robots in industrial environments. Human Systems Management, 18, 275 – 285.

Biocca, F. and Delaney, B. (1995). Immersive virtual reality technology. In Frank Biocca and M. R. Levy (Eds.). Communication in the age of virtual reality, 57 – 124. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bystrom, K., Barfield, W., and Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. Presence, 8 (2), 241 – 244.

Casali, J. G. and Wierwille, W. W. (1982). A sensitivity/intrusion comparison of mental workload estimation techniques using a flight task emphasizing perceptual piloting activities. In Proceedings of the 1982 IEEE International Conference on Cybernetics and Society, Seattle Washington, October, 1982, 598-602.

Casali, J. G. and Wierwille, W. W. (1983). Communications imposed pilot workload: a comparison of sixteen estimation techniques. In Proceedings of the 2nd Symposium on Aviation Psychology, Ohio State University, Columbus, OH, April, 1983.

Chew, V. (1977). Comparisons among treatment means in an analysis of variance. Agriculture Research Sciences of the United States Department of Agriculture, Washington, DC, 20 – 21 & 33.

Cohen, J. (1977). Statistical power analysis for behavioral sciences (Rev. ed.) New York: Academic Press.

Corker, K. Mishkin, A., and Lyman, J. (1980). Achievement of a sense of operator presence in remote manipulation. (Tech. Report No. 60) UCLA School of Engineering and Applied Science, Biotechnology Laboratory.

Corliss, W. R. and Johnsen, E. G. (1968). Teleoperator controls: An AEC-NASA technology survey (Tech. Report NASA-SP-5070). Washington, DC: National Aeronautics and Space Administration.

Couture, S. A., Hurd, R. L., and Wilhelmsen, K. C. (1998). Robust telerobotics – an integrated system for waste handling, characterization and sorting. In Proceedings of the 7th International Meeting of the American Nuclear Society.

Deisinger, J., Cruz-Neira, C., Reidel, O., and Symanzik, J. (1997). The effect of different viewing devices for the sense of presence and immersion in virtual environments: a comparison of stereoprojections based on monitors, HMDs, and screen. In M. Smith, G. Salvendy, and R. Koubek (Eds.). Design of computing systems: social and ergonomic considerations, pp. 881-884. Amsterdam: Elsevier.

Dillon, C., Keogh, E., Freeman, J. and Davidoff, J. (2000). Aroused and immersed: the psychophysiology of presence. Available on-line at: http://homepages.gold.ac.uk/immediate/immersivetv/P2000-dillon.html.

Ditton, T. (1997). The unintentional blending of direct experience and mediated experience: the role of enhanced versus limited television presentations. Unpublished doctoral dissertation. Temple University, Philadelphia, PA.

Draper, J. B. (1995). Teleoperators for advanced manufacturing: applications and human factors challenges. International Journal of Human Factors in Manufacturing, 5 (1), 53 – 85.

Draper, J. V. and Blair, L. M. (1996). Workload, flow, and telepresence during teleoperation. In Proceedings of the 1994 IEEE International Conference on Robotics and Automation, Minneapolis, Minnesota, April 1996.

Draper, J. V., Kaber, D. B., and Usher, J. M. (1998). Telepresence. Human Factors, 4(3), 354 - 375.

Draper, J. V., Kaber, D. B., and Usher, J. M. (1999). Speculations on the value of presence, CyberPsychology and Behavior, 2 (4), 349 – 361.

Eisenhauer, D. J., Norman, C. O., Kochanski, F. K., and Foley, J. W. (1999). Enhanced teleoperated ordnance disposal system (ETODS) for humanitarian demining. In Proceedings of the 8th International Meeting of the American Nuclear Society, Pittsburgh, PA, April 1999.

Ellis, S. R. (1996). Presence of mind: a reaction to Sheridan's musings on telepresence. Presence: Teleoperators and Virtual Environments, 5(2), 247-259.

Ellis, S. R., Dorighi, N. S., Menges, B. M., Adelstein, B . D., and Jacoby, R. H. (1997). In search of equivalence classes in subjective scales of reality. In M. Smith, G. Salvendy, and R. Koubek (Eds.). Design of computing systems: social and ergonomic considerations, pp. 873-876. Amsterdam: Elsevier.

Endsley, M. R. (1987). SAGAT: A methodology for the measurement of situation awareness (NOR DOC 87-83). Hawthorne, CA: Northrup Corp.

Endsley, M. R. (1988). Situation Global Assessment Technique (SAGAT). In Proceedings of the National Aerospace and Electronics Conference, 789 – 795. New York: IEEE.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. Human Factors, 37(1), 65-84.

Endsley, M. R. and Jones, W. M. (1997). Situation awareness information warfare and information dominance. Tech. Report 97-01.

Endsley, M. R. and Rodgers, M. D. (1994). Situation awareness information requirements analysis for en route air traffic control. In Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society.

Fontaine, G. (1992). The experience of a sense of presence in intercultural and international encounters. Presence: Teleoperators and Virtual Environments, 1(4), 482-490.

Fracker, M. L. (1991). Measures of SA: Review and future directions. (Tech. Report AL-TR-1991-0128). Wright-Patterson Air Force Base, OH: Air Force Systems Command.

Freeman, J. Avons, S. E., Harrison, D., Lodge, N., Pearson, D. (1999). Behavioral realism as a metric of presence. Available on-line at http://homepages.gold.ac.uk/immediate/btpres/BTpres.html.

Hahn, S. and Kramer, A. (1998). Further evidence for the division of attention among non-contiguous locations. Visual Cognition, 5 (1/2), 217-256.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Human mental workload (P. A. Hancock and N. Meshkati (Eds.)). North-Holland: Elsevier Science, 139-183.

Hayter, A. (1996). Probability and statistics for engineers and scientists. PWS Publishing Co., Boston, MA.

Held, R. and Durlach, N. (1987). Telepresence, time delay and adaptation. NASA Conference Publication 10032, chapter 28.

Hendrix, C. and Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. Presence, 5 (3), 274 – 289.

Huber, J. W. and Davies, I. R. L. (1996). Towards specifying the design criteria for the visual front-end of tele-presence systems. Taylor & Francis Inc., 372 – 377.

Hine, B., Hontalas, P., Fong, T., Piguet, L., Nygren, E., and Kline, A. (1995). VEVI: A Virtual Environment Teleoperations Interface for Planetary Exploration. In *Proceedings of the SAE 25th International Conference on Environment Systems*. San Deigo, CA.

James, E. (1890). The principles of Psychology. (Vol 1). New York: Dover.

Jelatis, D. G. (1975). Characteristics and evaluation of master-slave manipulators. In T. B. Sheridan (Ed.), Performance Evaluation of Programmable Robots and Manipulators (Tech. Report NBS Special Publication 459; pp. 141-146). Washington, DC: National Bureau of Standards.

Kaber, D. B. (1997). The effect of level of automation and adaptive automaton on performance in dynamic control environments. (Tech. Work. Doc. ANRCP-NG-ITWD-97-01). Amarillo, TX: Amarillo National Resource Center for Plutonium.

Kaber, D. B., Riley, J. M., Zhou, R., and Draper, J. V. (2000). Effects of visual interface design, control interface type, and control latency on performance, telepresence, and workload in a teleoperation task. In Proceedings of the XIVth Triennial Congress of the International Ergonomics Association / 44th Annual Meeting of the Human Factors and Ergonomics Society, San Diego, CA, August 2000.

Kaber, D. B. and Riley, J. M. (in press). Task difficulty and user motivation effects on performance and telepresence in a teleoperation task. To appear in D. Harris (Ed.) Proceedings of the 3rd International Conference on Engineering Psychology and Cognitive Ergonomics. Aldershot, UK: Ashgate.

Kalawasky, R. S. (1998). VRSART a tool for evaluation of contributory factors associated with presence in spatially immersive environments. Available on-line at: http://sgi-hursk.lut.ac.uk/~avrrc/presence/srsart.html.

Keppel, G. (1982). Design and analysis: a researcher's handbook (2nd ed.). Prentice-Hall, Inc.: Englewood Cliffs, NJ.

Klein, G. (1997). Applied cognitive task analysis (ACTA). Fairborn, OH: Klein.

Lemhofer, A. J. (1999). Robotics in bomb disposal and demining. In Proceedings of the 8th International Meeting of the American Nuclear Society, Pittsburgh, PA, April 1999.

Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2000). Development of a new cross-media presence questionnaire: the ITC sense of presence inventory. Presented the 3rd International Workshop on Presence, Delft, The Netherlands. Available on-line at: http://homepages.gold.ac.uk/immediate/immersive/P2000-lessiter.html.

Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J. and Park, S. (2000). Measuring presence: a literature-based approach to the development of a standardized paper-and-pencil instrument. Presented at the 3rd International Workshop on Presence, Delft, The Netherlands. Available on-line at: http://nimbus.temple.edu/~mlombard/P2000.html.

Lombard, M. and Ditton, T. (1997). At the heart of it all: the concept of presence. JCMC, 3 (2), available on-line at http://www.ascusc.org/jcmc/vol3/issue2/lombard.html.

Loomis, J. M. (1992). Distal attribution as presence. Presence, 1 (1), 113 – 119.

Massimo, M. J. and Sheridan, T. B (1994). Teleoperator performance with varying force and visual feedback. Human Factors, 36(1), 145-157.

Mendenhall, W. and Sincich, T. (1995). <u>Statistics for engineering and sciences</u>, 4th edition, Prentice Hall, Inc., Upper Saddle River, NJ.

Meyers, D. G. (1998). <u>Psychology</u>, 5th edition, 181, Worth Publishers: New York, NY.

Myers, R. H. (1990). <u>Classical and modern regression with applications</u>, 2nd edition, PWS-Kent Publishing Co., Boston, MA.

Nash, E. B., Edwards, G. W., Thompson, J. A., and Barfield, W. (2000). A review of presence and performance in virtual environments. International Journal of Human-Computer Interaction, 12 (1), 1 – 41.

Navon, D. and Gopher, D. (1979). On the economy of the human information processing system. Psychological Review, 86 (3).

Neter J., Wasserman, W., and Kunter, M. H. (1990). Applied linear statistical models (3rd. ed.). Homewood, IL: Irwin.

Norman, C. O., Foley, J. W., Eisenhauer, D. J., and Kochanski, F. K. (1999). The semi-autonomous mine detection system (SAMS) for humanitarian demining. In Proceedings of the 8th International Meeting of the American Nuclear Society, Pittsburgh, PA, April 1999.

Pretlove, J. (1998). Augmenting reality for telerobotics: unifying real virtual worlds. Industrial Robot, 25 (6), 401 – 407.

Prothero, J. D. and Hoffman, H. D. (1995). Widening the field-of-view increases the sense of presence within immersive virtual environments (Human Interface Technology Laboratory Tech. Report R-95-4). Seattle: University of Washington.

Prothero, J. D., Parker, D. E., Furness, T. A., Wells, M. J. (1995). Towards a robust, quantitative measure of presence. In Proceedings of the Conference on Experimental Analysis and Measurement of Situation Awareness, 359 – 366. Available on-line at: http://www.hitl.washington.edu/publications/p-95-8.

Psokta, J. and Davidson, S. (1993). Cognitive factors associate with immersion in virtual environments. Alexandria, VA: Army Research Institute.

Rahimi, M. and Wierwille, W. W. (1982). Evaluation of the sensitivity and intrusion of workload estimation techniques in piloting tasks emphasizing mediational activity. In Proceedings of the 1982 IEEE International Conferences on Cybernetics and Society, Seattle Washington, October, 1982, 593-597.

Riley, J. M. and Kaber, D. B. (1999). The effects of visual display type and navigational aid on performance, presence, and workload in virtual reality training of telerover navigation. In Proceedings of Human Factors and Ergonomics Society 43rd Annual Meeting-1999, Houston, TX, 1251-1255.

Salas, E., Prince, C., Barker, D. P., and Shrestha, L. (1995). Situation awareness in team performance: implications for measurement and training. Human Factors, 37 (1), 123 – 136.

Sanders, M. and McCormick, E. J. (1993). Human factors in engineering design. McGraw-Hill, Inc.

Sarter, N. and Woods, D. D. (1991). Situation awareness: a critical but ill-defined phenomenon. International Journal of Aviation Psychology, 1, 45 – 47.

SAS/STAT ® (1990). User's Guide (Version 6, 4th ed.) Cary, NC: SAS Instutute.

Schloerb, D. W. (1995). A qualitative measure of telepresence. Presence, 4 (1), 64 – 80.

Schubert, T, Friedman, F., and Regenbrecht, H. (1999). Decomposing the sense of presence: factor analytic insights. Presented at the 2nd International Workshop on Presence, University of Essex, Colchester, UK.

Sekuler, R. and Blake, R. (1994). Perception. McGraw-Hill, Inc.: New York.

Sheridan, T. B. (1991). Telerobotics, automation, and human supervisory control. Cambridge: MIT Press.

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. Presence, 1(1), 120-125.

Slater, M. and Steed, A. (2000). A virtual presence counter. Presence, 9(5), 413-434.

Slater, M. and Usoh, M. (1993). Representations systems, perceptual position, and presence in immersive virtual environments. Presence, 2 (3), 221 – 233.

Slater, M. and Wilbur, S. (1997). A framework for immersive visual environment (FIVE): Speculations on the role of presence in virtual environments. *Presence*, 6(6), 603-617.

Smith, K. and Hancock, P. A. (1995). Situation awareness is adaptive is adaptive, externally directed consciousness. Human Factors, 37 (1), 137 – 148.

Smith, T. J. and Smith, K. U. (1989). The human factors of workstation telepresence. In S. Griffin (Ed.), Third annual workshop on space operations automation and robotics (Tech. Report NASA Conference Publication 3059, pp. 235-250). Houston, TX: NASA, Johnson Space Flight Center.

Stanney, K. M., Mourant, R. R., et al. (1998). Human Factors Issues in virtual Environments. Presence: Teleoperators & Virtual Environments, 7(4), 327-.

Stanney, K. M, Salvendy, G., et al. (1998). Aftereffects and sense of presence in virtual environments: formulation of a research and development agenda. International Journal of Human-Computer Interaction, 10 (2), 135-187.

Stassen, H. G. and Smets, G. J. F. (1995). Telemanipulation and telepresence. Man-Machine Systems, 13 – 23.

Stone, R. (1992). Virtual reality and telepresence. Robotica, 10, 461 – 467.

Tachi, S. (1999). Telexistence and R-Cubed. Industrial Robot, 26 (3).

Taylor, R. M. (1990). Situational awareness rating techniqu (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (AGARD-CP-478), 3/1 – 3/17. NATO Advisory Group for Aerospace Research and Development.

Tcherneshoff, M. (1998). In Proceedings of the 7th International Meeting of the American Nuclear Society.

Vidulich, M. A. (1989). The use of judment matrices in subjective workload assessment: the subjective workload domincance (SWORD) technique. In Proceedings of the Human Factors Society, 33rd Annual Meeting, 1406-1410.

Wierwille, W. W., and Casali, J. G. (1983). A validated rating scale for global mental workload measure applications. In Proceedings of the Human Factors Society, 27th Annual Meeting, 129 - 133.

Welch, R. B. (1999). How can we determine if the sense of presence affects task performance? Presence, 8 (5), 574 – 577.

Welch, R. B., Blackman, T. T., Liu, A., Mellers, B. A., and Stark. L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on subjective sense of presence. Presence: Teleoperators and Virtual Environments, 5 (3), 263-273.

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), Attention and Performance VIII. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C. D., Sandry, D., and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. Human Factors, 25, 227-248.

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

Witmer, B. G. and Singer, M. J. (1994). Measuring Immersion in Virtual Environments (Tech. Report 1014). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. Presence, 7 (3) 225 – 240.

APPENDIX A
TABLE OF ACRONYMS

Literature Acronym	**Description**
VE(s)	Virtual Environment(s)
VR	Virtual Reality
RW	Real World
PQ	Presence Questionnaire
IQ	Immersive Tendencies Questionnaire
SA	Situation Awareness
LOD	Level of Difficulty
MCH	Modified Cooper-Harper Scale
SAGAT	Situation Awareness Global Assessment Technique
Statistical Acronym	**Description**
LOD	Level of Difficulty
PQ	Telepresence Scores
AVGPERF	Average Performance
AVGSA	Average Situation Awareness
AVGSA1	Average Situation Awareness 1
AVGSA2	Average Situation Awareness 2
AVGSA3	Average Situation Awareness 3
MCH	Mental Workload Ratings
ATTNRAT	HSVE/HSRW
HSVE	Hit-to-Signal ratio in Virtual Environment
HSRW	Hit-to-Signal ratio in Real World
IQ	Immersive Tendencies Scores

APPENDIX B
LIST OF VERBAL COMMANDS

List of Verbal Commands

SIGNAL ON	To turn auditory signal on.
SIGNAL OFF	To turn auditory signal off.
MENU ON	To display the 2-D HUD menu.
MENU OFF	To remove the 2-D HUD menu.
UPPER ARM UP	To rotate upper arm upward.
UPPER ARM DOWN	To rotate upper arm downward.
LOWER ARM UP	To rotate the lower arm upward.
LOWER ARM DOWN	To rotate the lower arm downward.
SHOULDER RIGHT	To rotate the shoulder to the right.
SHOULDER LEFT	To rotated the shoulder to the left.
WRIST UP	To rotate the wrist upward.
WRIST DOWN	To rotate the wrist downward.
ROTATE WRIST	To rotate the wrist counterclockwise.
STOP	To stop rotation of any component.
SELECT AIRKNIFE	To select the airknife tool.
SELECT SHOTGUN	To select the shotgun tool.
SELECT GRIPPER	To select the gripper tool.
BATTERY	To respond to low battery signals in VE.

APPENDIX C
SUBJECT INSTRUCTIONS

Subject Instructions

Introduction

Thank you for volunteering to participate in this experiment. The study is being conducted to examine the relationships between teleoperation task performance, telepresence, situation awareness, attention, and mental workload. The experimental task will require you to use a standard 2-D mouse and conventional keyboard to control the motion of a virtual mobile robot (or rover) in a virtual outdoor minefield. 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 uncover and detonate or remove mines from the environment. The environment will be presented to you through a 21-inch computer monitor. During the experiment, you will complete an extensive training session and two test trials. The training will occur today and the experimental tests will occur tomorrow.

Overview of Procedures

The procedures we will follow during the experiment include: (1) a brief equipment familiarization and instruction period; (2) completion of an "immersive tendencies" questionnaire to assess susceptibility to feeling a part of the virtual environment during testing; (3) an extensive training session including three periods to ensure that you are able to complete the experimental tasks; (4) explanation of the Presence Questionnaire which is intended to assess your association with the virtual task and environment during performance; (5) explanation of the Modified Cooper-Harper Scale which is intended to evaluate the level of mental workload you experience during testing; (6) performance in the virtual environment navigation and mine neutralization task; (7) completion of the PQ; (8) completion of the MCH scale; (9) a short break; (10) a second trial followed by completion of the PQ and MCH scale; (11) 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 4 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 the read the form. This form will not be associated with any of the other survey forms used in this experiment.

[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.

[Complete all payment forms for participation. Be sure to record the time that the subject started participation.]

<u>Instructions</u>

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 cause control of the robotic arm. Keys as part of the number keypad will be used to cause rotation of joints of the arm.

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

Pre-Testing Survey

Before we begin the training, you will participate in the immersive tendencies questionnaire that was mentioned during the summary of the experimental procedures.

[Give the IQ to the subject.]

This survey seeks to determine how well or easily you can immerse yourself in different environmental situations. It establishes the degree to which you are susceptible to feeling a part of a virtual environment. I will read the instructions at the top of the form regarding its completion.

[Read statements at top of questionnaire.]
[Allow time to complete survey.]

At this time, I would like you to remove your watch or timepiece from your person and place it in a pocket. Due to the nature of the experiment, keeping your attention focused on completing the training session is important. I also ask that you refrain from talking during the training periods. If you have any difficulties, however, please do not hesitate to bring them to my attention and I will assist you.

Training Session

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 keyboard and mouse to complete the experimental tasks.

The first training period will last approximately 30 minutes. This training period will provide instruction and practice in how to complete the teleoperated mine neutralization task. Your goal during the teleoperation task will be to locate mines in the virtual outdoor space. To do so, you will need to control the motion and direction of the rover, while listening for an auditory signal indicating the proximity of a mine to the rover. The auditory signal is a “bell ringing” sound.

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 need to 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, attempting to locate a mine using the auditory detection signal.]

During the task, you may turn the auditory signal on or off. A key on the keyboard will be pressed to enable or disable the signal. A controller will press keys on the keyboard to toggle the auditory signal. To initiate this action you will be required to use verbal commands to which the controller will respond. The verbal commands are as follows:
SIGNAL ON – to turn to auditory signal on.
SINGAL OFF – to turn the auditory signal off.
You may want to turn the signal off while you are attempting to neutralize a mine and turn it back on when you begin to search for mines.

[Demonstrate use of the keys and commands to the subject for toggling the auditory cue. Refer to command list.]

Three viewpoints will be available to you in the virtual environment. One viewpoint will be from a position several feet above the terrain and several feet behind the rover. This viewpoint will give you a view of the entire rover and some of the surrounding environment. The other two viewpoints will be from virtual cameras mounted on the rover. One viewpoint is from a camera that is located on a stand 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. The other camera is located on the lower arm of the robotic manipulator. 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. You may toggle between the three viewpoints by using the right mouse button on the mouse controller.

[Demonstrate use of the mouse button for toggling viewpoints.]

When a mine is located, you will need to use tools on the robot arm to uncover and neutralize the mine. Some mines will need to be picked-up and removed from the ground. Others will need to be detonated because of their potential for exploding when handled. You will need to manipulate the robot arm to position the tools for use. They keyboard will be used to position the robot arm. Keys on the number keypad are associated with rotational motion of components of the arm. Review the “key map” that to identify the specific motions linked to the various keys.

[Show subject the “key map”. Select the camera stand view so the robot arm is visible. Explain the motion caused by using each key, while the keys on the number keypad are pressed by the subject.]

The 4 and 6 keys cause the robot shoulder (and its other parts) to rotate from left to right. Pressing the 4 key rotates the shoulder to the left. Pressing the 6 key rotates the shoulder to the right.

The 8 and 2 keys rotate the upper arm up and down. Pressing the 8 key rotates the upper arm up. Pressing the 2 key rotates the upper arm down.

The 7 and 1 keys rotate the lower arm up and down. Pressing the 7 key rotates the lower arm up. Pressing the 1 key rotates the lower arm down.

The wrist may also be rotated up and down. Pressing the 9 key rotates the wrist up. Pressing the 3 key rotates the wrist down.

The wrist may be rotated in the counterclockwise direction. Pressing the “+“ key on the number keypad rotates the wrist.

The controller will use the keyboard and the appropriate keys to cause motion of the robotic manipulator. You will be required to use verbal commands to direct the motion of the manipulator and position it in the environment. The controller will respond to each of your verbal commands to cause the manipulator arm to move in the desired direction. 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.
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.
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.
WRIST UP - this command will be used to rotate the wrist upward.
WRIST DOWN - this command will be used to rotate the wrist downward.
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. Allow a short practice with the commands (3 mins.)]

Once the mine is located and the robot arm is in position, you will have to use the tools on the arm to neutralize the mine. Three tools are available to you: (1) an airknife that can be used to uncover the mines by removing dirt; (2) a shotgun that can be used to detonate the mines; and (3) a magnetic gripper that can be used to pick-up, carry and release mines. The tools must be selected by using the keyboard. Pressing the 1 key of the main keyboard will select the airknife. Pressing the 2 key will select the shotgun. Pressing the 3 key will select the gripper. A small visual display is located on the rover. This display will indicate which tool is in use. You can view this display by using the camera located on the stand of the rover.

You will select the appropriate tool by using verbal commands. The controller will respond to each command to cause the tools to be selected. The commands that you will use are as follows:
SELECT AIRKNIFE – to cause the airknife tool to be selected.
SELECT SHOTGUN – to cause the shotgun to be selected.
SELECT GRIPPER – to cause the gripper to be selected.

[Demonstrate use of the keys and commands to the subject and point out the visual display on the rover.]

When a tool is selected, the left mouse button can be pressed to activate the tool. For example, if the gripper is selected and the left mouse button is pressed, the magnetic gripper will be charged or de-charged. When the gripper is charged the mines can be picked up. When the gripper is de-charged, the gripper will release the mines. Similarly, if the shotgun is selected, pressing the left mouse button will cause the gun to shoot. If the airknife is selected, pressing the left mouse button will case the tool to release a burst of air.

[Demonstrate use of tools with mouse and tools selection at the keyboard.]

As discussed earlier, you will be required to remove some mines and detonate others. There are two broad categories of mines: anti-personnel and anti-tank. The anti-personnel (or AP) mines can and should be removed from the ground. That is, you will be required to uncover the mines with the airknife, identify whether it is and AP mine, and use the gripper tool to pick the mine up.

The mines will then have to be moved to a storage container located in the environment.

[Show the subject the storage box.]

You should be able to recognize the AP mines by there shape. These are the AP mines.

[Show the subject the AP mines and discuss the physical characteristics.]

These mines should not be detonated. They should be picked up and moved to the storage are.

The anti-tank (or AT) mines should not be handled using the robot gripper. These mines must be detonated by shooting them with the shotgun. You will be able to recognize the AT mines by shape. These are the AT mines.

[Show the subject the AT mines and discuss the physical characteristics.]

These mines should be detonated. They should not be picked up with the gripper.

A 2-D text menu is available to you as a memory aid. The menu will provide information on which mouse buttons are to be pressed for activating tools and alternating between viewpoints. You may toggle the menu on and off by using verbal commands. The controller will respond to each command. The commands for toggling the 2D menu display are as follows:
MENU ON – to display the menu.
MENU OFF – to remove the menu.

[Demonstrate the use of the keyboard and commands for toggling the menu. Check to see if the subject has any questions.]

You will now be allowed to train in the teleoperation task. Take this time to practice navigation of the rover for locating mines and manipulating the arm and tools for neutralizing mines in the environment. A command menu will be available to you as a memory aid during training and during experimental task performance. The command menu will provide information on all verbal commands to be used for directing motion of the manipulator arm, selecting tools, and activating the 2D menu on the display.

The simulation is intended to be a somewhat realistic representation of a mission with an enhanced teleoperated ordnance disposal unit. With this in mind, you should expect the search and neutralization tasks to be as difficult as you might expect them to be in reality. If at any time during the training session you begin to feel uncomfortable or fatigued, please inform me immediately. We will begin the training now.

[Allow subject to practice in neutralization of one mine at a time. Be sure that he/she has had the opportunity to practice neutralizing at least one mine of each type. Total training time should be about 30 minutes.]

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

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

[After break, begin second training period.]

We will now begin the second training period. This time is provided to allow you the opportunity to practice in two monitoring tasks that are associated with operation of the rover. The rover is equipped with two batteries. The power levels of these batteries are critical to the operation of the rover and must be monitored to prevent a system power loss and failure. One battery signal is associated with the power level of the rover itself. If the battery is used at low power levels motion of the rover and functioning of the tools may be compromised. This signal will be presented in the virtual environment and will be superimposed over the environment on the display. The signal will be located in the lower left corner of the display.

[Show signal to subject.]

The signal has two distinct states. For a normal battery power level, the signal will be a green square. This means that the battery power level is adequate for operation of the rover. For a low battery, the signal will be a green triangle. This signal indicates that the battery power level is low. To maintain normal functioning of the rover you will need to correct for low battery signals by using a verbal command to which the controller will respond by pressing keys as part of the number keypad. The verbal command to be used is as follows:

BATTERY – to respond to low battery power signals.

This command will allow you to change to a new battery for powering the rover, and the signal will return to its normal state (a green square). The battery command is also included on your verbal command sheet.

[Demonstrate use of the key and verbal commands for responding to low battery power of the rover battery.]

The second battery signal is associated with the power level of the rover's communications system. If this battery is used at a low power level, communication between the rover and the mouse and keyboard control may be compromised. This signal will be presented in the real world on a separate display screen.

The communications signal also has two distinct states. For a normal battery power level, the signal will be a green square. This means that the battery power level is adequate for maintaining communication between the controls and the rover. For a low battery, the signal will be a green triangle. This signal indicates that the battery power level is low. You will control all responses to the communication battery. To maintain normal communications with the rover you will need to correct for low battery signals by pressing the "b" key on the keyboard. This action will allow you to change to a new battery, and the signal will return to its normal state (a green square).

[Show the "b" key for responding to the communications battery.]

You will now be allowed to practice in the monitoring tasks separately. This will ensure that you are able to perform each task during task performance.

[Allow time for subject to practice each monitoring task. Approximately 7-min should be provided for each monitoring scenario. For rover battery, allow subject to use the verbal command.]

During task performance you will be required to perform the monitoring tasks and the teleoperated demining task simultaneously. You will now be allowed to practice in multi-task performance to ensure that you are able to complete the tasks during experimental testing. The steps of the task are listed for you and displayed on the task flow chart.

[Show the subject the task flow chart and explain the portions of the task. Explain the terms detection, identification, excavation, detonation, neutralization, etc.]

Please take this time to practice completing the teleoperation task while also monitoring form low battery signals.

[Allow time for subject to practice multi-task performance. Approximately 14 minutes should be provided.]

You have completed the second training period. Please take a short break before we begin the third and final training period.

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

You will now complete the third training period. As discussed earlier, an objective of this study is to assess the utility of situation awareness for describing telepresence experiences in a teleoperation task. Situation awareness can described as perception of elements in an environment, including their current and future status. Situation awareness will be measured during task performance. The measure of situation awareness to be used for this study will require you to answer questions regarding the state of the task and environment at specific points during performance. This training period is provided to ensure that you are aware of how the situation awareness measure is administered and how to respond to the questions.

The Situation Awareness Global Assessment Technique (SAGAT) will be used to measure situation awareness. This technique involves stopping task performance at random points in time to administer questions concerning both the present and future states of the task. These queries will be presented to you through a paper and pencil survey. You will be allowed time to read and answer all questions. You will respond to the questions by circling a response. During a SAGAT stop, I will as you to stop task performance and turn away from the display. You will be required to stop performance immediately. I will then present the questions to you and ask you to answer all questions. You will then be allowed time to respond to the questions. Once you have completed the survey, I will ask you to turn back to the display and return to task performance. You will then be required to continue with performance of the task immediately. Several SAGAT stops may occur during task performance. You will not be informed of the timing of the stops or the questions to be asked at each stop.

This forms shows all of the questions that may potentially be presented to you during performance of the task.

[Show the subject all the questions. Explain terms detection, identification, excavation, detonation, neutralization, etc.]

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

You will now be allowed to practice multi-task performance incorporating exposure to the SAGAT freezes and queries.

[Allow subject to train in performance with SAGAT. Follow SAGAT testing protocol for training. Approximately 20-min should be allowed for practice.]

You have completed all training requirements for participation in the experiment. Do you have any questions regarding the training and/or task performance? I ask that you return tomorrow to complete the experimental test trials.

[Remind the subject of his/her scheduled time for participation. Return personal belongings to subject. Help subject complete payment forms for participation in training.]

Experimental Testing

Thank you for returning to participate in the experimental testing. Today you will complete two test trials in the task that you trained on yesterday. The experimental procedures for today are as follows: (1) explanation of post trial surveys; (2) participation in Trial 1; (3) completion of the PQ; (4) completion of the MCH scale; (5) a short break; (6) participation in Trial 2; (7) completion of post-trial questionnaires; (8) and a debriefing of the study.

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

Before we begin with your first test trial, I will explain the surveys to be administered following the trials. The PQ will be completed directly subsequent to each trial. This questionnaire is intended to capture the degree to which you felt a part of the teleoperation task and environment. I will show you the survey and read the instructions on the survey so that you will know how to respond to the questions with a rating following experimental trials.

[Show the subject the PQ. Read statement at the top of a copy of the PQ.]

You will also be asked to complete a workload rating following each experimental trial. I will explain how to rate your mental demand level using the flow diagram and descriptor terms so that you will know how to assign an appropriate rating following each trial.

Start at the bottom of the flow diagram at the “operator decisions” box. Read the question in the first diamond. If the answer to the question is “No”, then follow the diagram to the appropriate rating box. If the answer to the question is “Yes”, follow the diagram to the next diamond. Read the question in the diamond. If the answer to the question is “No”, follow the flow diagram to the difficulty and operator demand level descriptor terms. Read the statements and phrases in each box and determine which describes the level of difficulty and mental demand you experienced in the immediate past task performance. Assign the rating that corresponds to the box that you have chosen. If the answer to the question in the diamond is “Yes”, follow the flow diagram to the next diamond. Read the question in the diamond and proceed in a similar manner as explained for the previous diamond. If you are unable to process a mine, then you must rate the task as “Impossible”.

[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. Recall that some mines will require you to use tools for removing them from the area, while others will require use of tools for detonation. You should neutralize a total of four mines in the environment.

[Remind subjects of which mines are to be removed and which are to be detonated.]

Remember that you will have to attend to the monitoring tasks at the same time you complete the teleoperation task. Recall that there are two critical signals to be observed; one for the rover motion and control and one for the rover communication system. You will need to attend to both signals for the duration of task performance and respond if either is not in a normal functioning state (i.e., the green square). Recall that you will use verbal commands to respond to the rover battery and press the “b” key to respond to the communications battery.

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

At this time, I would like you to remove your watch or timepiece from your person and place it in a pocket. Due to the nature of the experiment, keeping your attention focused on completing the task is important. I also ask that you refrain from talking during the training periods. If you have any difficulties, however, please do not hesitate to bring them to my attention and I will assist you. Please take a short break while I prepare the system for your test trial.

[Set up the trial.]

1. **Erase all old data files (atten_data, mine_nuet on portable and PC.**
2. **Start both simulations.**
3. **Record the subject number, level of difficulty, etc.**
4. **Make note of the SA query stop times and be sure to have the SA questions ready.**

You may now return to the experimental set-up. You will now start the trial.

[Ask subject if they are ready to proceed. Say to the subject, "Please press the "g" key and begin task performance. Begin the trial by pressing the "g" key on the controller keyboard. Observe subject performance, record neutralization times with the "n" key, and administer SA queries using the "SA Query Instructions".]

SAGAT Query Instructions

1. **Review the random stop list to determine when the query stops should occur. This list contains random stop times and will provide times query stop can take place. You should have a stop time for Query 1, Query2 and Query 3 before the start of the trial.**
2. **When you start the trial, allow the simulation to run for 5 minutes without a query stop.**
3. **After 5 minutes, start your stopwatch.**
4. **When the stopwatch reaches the random stop time for the first query, stop the watch and say to the subject:**

"Please stop task performance and press the "S" key on your keyboard. Turn away from the display. "

5. **Press the "S" key on the controller keyboard.**
6. **When the subject has turned away from the display, give the subject a SA query sheet.**
7. **Allow the subject time to complete all questions.**
8. **Complete the experimenter query sheet with the correct responses.**
9. **Collect the SA query sheet from the subject.**

10. Reset the stopwatch for the next query stop (Q2 or Q3). Say to the subject:

"Please turn back to the display. Press the "S" key on your keyboard and immediately return to task performance."

11. Press the "S" key on the controller keyboard and start the stopwatch.
12. Allow 1 minute to pass without a query.
13. Then stop the watch at the next random stop time.

[When subject has completed a total of four mines, stop task perforance. Say to the subject, "Please press the "q" key and stop task performance." Press the "q" key on the controller keyboard to stop task performance and record data.]

[Following completion of the trial administer the PQ.]

I would now like you to complete the PQ that we discussed in covering the experimental procedures and instructions. This survey is intended to capture the degree to which you felt a part of the virtual environment. I will read you the instructions on the survey. Please respond to each question with a rating.

[Give subject the MCH scale.]

Now I would like you to take time to complete the workload assessment survey. You will subjectively rate the level of mental workload using the flow diagram and verbal descriptors.

[Allow time to complete the MCH.]

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

[Allow subject a 5-min break.]

You will now complete your second experimental trial. The method of navigating the rover and manipulating the robot arm is the same. The task will be the same. Following this trial, as in the first test, you will complete the PQ and the MCH scale.

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

Closing

At this time, you have completed all of the procedures required as part of the experiment. I will now assist you in completing the payment form for your compensation for participation in the experiment. This form will not be associated with any of the other forms used in the experiment.

[Work with the subject to determine the total time and payment for participation.]

You can expect to receive payment for the experiment from the university within a minimum of four weeks. You will need to report the income you derive from your participation to the IRS, as it is taxable. Thank you for your time and participation in the study.

[Make sure all forms have been completed and returned.]

APPENDIX D
PRESENCE QUESTIONNAIRE

Presence Questionnaire

Characterize your experience in the environment, by marking an "x" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when marking your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change you answer.

<u>WITH REGARD TO THE EXPERIENCED ENVIRONMENT:</u>

1. How much were you able to control events?

|___|___|___|___|___|___|___|

NOT AT ALL SOMEWHAT COMPLETELY

2. How responsive was the environment to actions that you initiated (or performed)?

|___|___|___|___|___|___|___|

NOT RESPONSIVE MODERATELY RESPONSIVE COMPLETELY RESPONSIVE

3. How natural did your interactions with the environment seem?

|___|___|___|___|___|___|___|

EXTREMELY ARTIFICAL BORDERLINE COMPLETELY NATURAL

4. How much did the visual aspects of the environment involve you?

|___|___|___|___|___|___|___|

NOT AT ALL SOMEWHAT COMPLETELY

5. How natural was the mechanism which controlled movement through the environment?

|___|___|___|___|___|___|___|

ETREMELY ARTIFICIAL BORDERLINE COMPLETELY NATURAL

6. How compelling was your sense of objects moving through space?

NOT AT ALL	MODERATELY COMPELLING	VERY COMPELLING

7. How much did your experiences in the virtual environment seem consistent with your real world experiences (e.g., building navigation)?

NOT CONSISTENT	MODERATELY CONSISTENT	VERY CONSISTENT

8. Were you able to anticipate what would happen next in response to the actions that you performed?

NOT AT ALL	SOMEWHAT	COMPLETELY

9. How completely were you able to actively survey or search the environment with your vision?

NOT AT ALL	SOMEWHAT	COMPLETELY

10. How compelling was your sense of moving around inside the virtual environment?

NOT COMPELLING	MODERATELY COMPELLING	VERY COMPELLING

11. How closely were you able to examine objects?

NOT AT ALL	PRETTY CLOSELY	VERY CLOSELY

12. How well could you examine objects from multiple positions?

NOT AT ALL	SOMEWHAT	EXTENSIVELY

13. How involved were you in the virtual environment experience?

NOT INVOLVED	MILDLY INVOLVED	COMPLETELY INVOLVED

14. How much delay did you experience between your actions and the expected outcomes?

NO DELAYS	MODERATE DELAYS	LONG DELAYS

15. How quickly did you adjust to the virtual environment?

NOT AT ALL	SLOWLY	LESS THAN ONE MINUTE

16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

NOT PROFICIENT	REASONABLY PROFICIENT	VERY PROFICIENT

17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

NOT AT ALL	INTERFERED SOMEWHAT	PREVENTED TASK PER- FORMANCE

18. How much did the control devices interfere with the performance of assigned tasks or with other activities?

NOT AT ALL	INTERFERED SOMEWHAT	INTEREFERED GREATLY

19. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

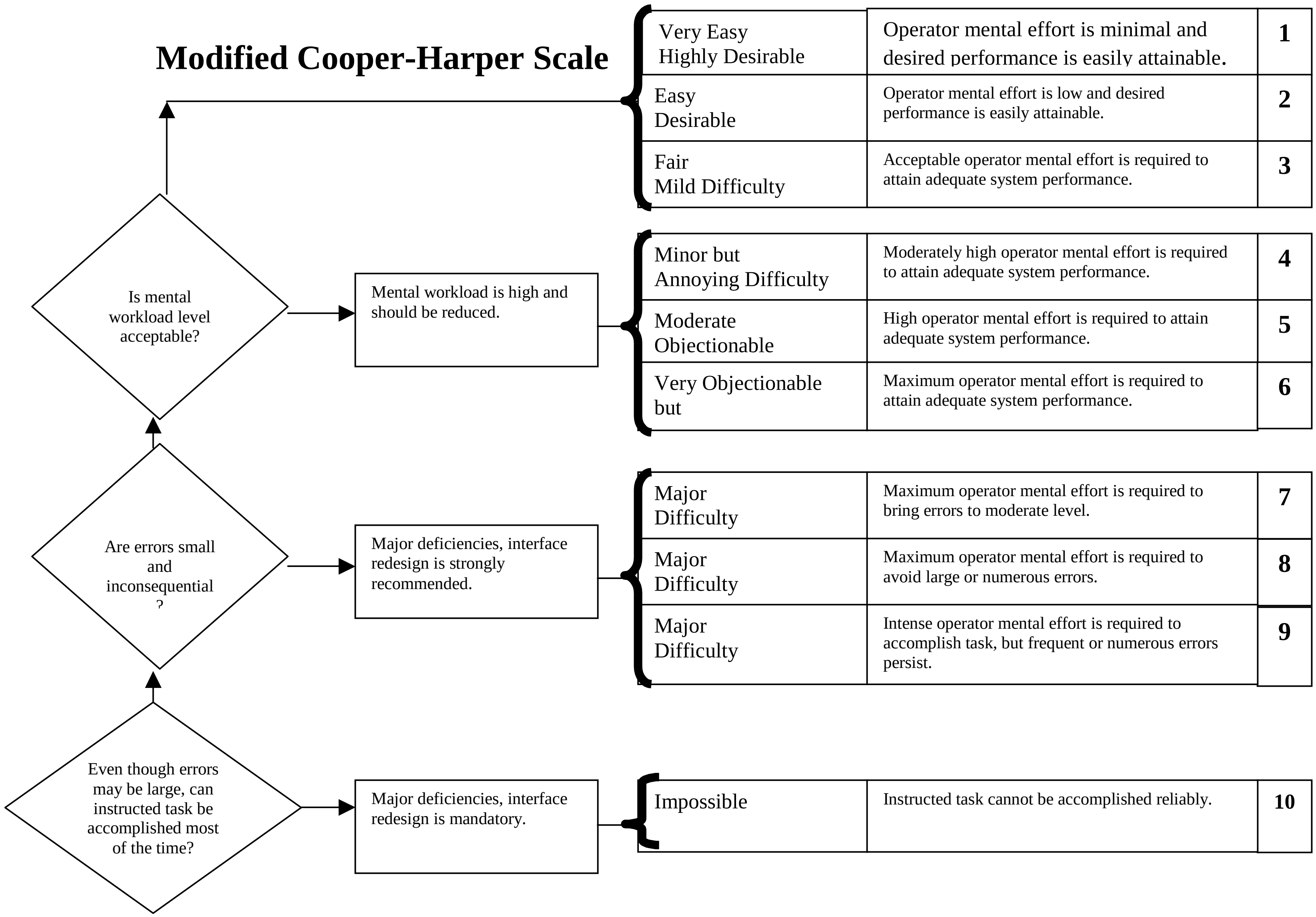
NOT AT ALL			SOMEWHAT			COMPLETELY

Do not write below this line. Experimenter use only.

Subject Number: ______________ LOD : ________________

APPENDIX E
MODIFIED COOPER-HARPER SCALE

Modified Cooper-Harper Scale



APPENDIX F
SITUATION AWARENESS QUERIES

Queries on SA, potential responses, and the Level of SA targeted by each query.

SAGAT QUERY	POSSIBLE RESPONSES	LEVEL OF SA
What is the current tool type in use?	(a) air gun; (b) gripper; (3) gun; (d) none.	1
What step in the task is currently being completed?	(a) detection of mine; (b) uncovering mine; (c) identification of mine; (d) excavation; (e) detonation; (f) transportation of mine.	1
What type of mine is currently being neutralized?	(a) Anti-Personnel; (b) Anti-Tank; (c) none.	1
Is the auditory signal currently activated?	(a) yes; (b) no.	1
Is there a mine currently in the telerobot gripper?	(a) yes; (b) no.	1
How many mines have been neutralized so far (current mine plus completely neutralized mines)?	(a) 0; (b) 1; (c) 2; (d) 3; (e) 4	1
Is there a mine in close proximity to the rover?	(a) yes; (b) no.	2
Can the mine currently being neutralized be safely removed from the area?	(a) yes; (b) no; (c) no mine being neutralized.	2
Are the rover and telerobot arm currently close enough to a mine to pick it up?	(a) yes; (b) no; (c) no mine being neutralized.	2
How many mines have been detonated (including the current mine, if it is to be detonated)?	(a) 0; (b) 1; (c) 2; (d) 3; (e) 4	2
How many mines have been excavated (including the current mine, if it is to be removed from the area)?	(a) 0; (b) 1; (c) 2; (d) 3; (e) 4	2
How many mines will have been neutralized once the current mine is complete?	(a) 1; (b) 2; (c) 3; (d) 4	2
What tool will you need to use next (after the current tool)?	(a) air gun; (b) gripper; (3) gun.	3
How many more mines will you have to neutralize after the current mine is completed?	(a) 0; (b) 1; (c) 2; (d) 3, (e) 4.	3
What will happen if the current mine is picked-up?	(a) nothing (b) explosion; (c) no mine being neutralized.	3
What task step will need to be executed next (after successful completion of the current step)?	(a) detection of mine; (b) uncovering mine; (c) identification of mine; (d) excavation; (e) detonation; (f) transportation of mine.	3

APPENDIX G

IMMERSIVE TENDENCIES QUESTIONNAIRE

Immersive Tendencies Questionnaire

Indicate your preferred answer by marking an “x” in the appropriate box of the seven-point scale. Please consider the entire scale when marking you responses, as intermediate levels may apply. For example, if your response is once or twice, the second box from the left should be marked. If your response is many times but not extremely often, then the sixth (or second box from the right) should be marked.

1. Do you easily become deeply involved in movies or TV dramas?

NEVER OCCASIONALLY OFTEN

2. Do you ever become so involved in a television program or book that people have problems getting your attention?

NEVER OCCASIONALLY OFTEN

3. How mentally alert do you feel at the present time?

NOT ALLERT MODERATELY FULLY
ALERT

4. Do you ever become so involved in a movie that you are not aware of things happening around you?

NEVER OCCASIONALLY OFTEN

5. How frequently do you find yourself closely identifying with the characters in a story line?

NEVER OCCASIONALLY OFTEN

6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?

NEVER OCCASIONALLY OFTEN

7. What kinds of books do you read most frequently? (CIRCLE ONE ITEM ONLY!)

Spy novels	Fantasies	Science fiction
Adventure novels	Romance novels	Historical novels
Westerns	Mysteries	Other fiction
Biographies	Autobiographies	Other non-fiction

8. How physically fit do you feel today?

NOT FIT MODERATELY FIT EXTREMELY FIT

9. How good are you at blocking out external distractions when you are involved in something?

NOT VERY SOMEWHAT VERY
GOOD
GOOD GOOD

10. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

NEVER OCCASIONALLY OFTEN

11. Do you ever become so involved in a daydream that you are not aware of things happening around you?

NEVER OCCASIONALLY OFTEN

12. Do you ever have dreams that are so real that you feel disoriented when you awake?

NEVER OCCASIONALLY OFTEN

13. When playing sports, do you become so involved in the game that you lose track of time?

NEVER | | | OCCASIONALLY | | | OFTEN

14. How well do you concentrate on enjoyable activities?

NOT AT ALL | | | MODERATELY WELL | | | VERY WELL

15. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

NEVER | | | OCCASIONALLY | | | OFTEN

16. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

NEVER | | | OCCASIONALLY | | | OFTEN

17. Have you ever gotten scared by something happening on a TV show or in a movie?

NEVER | | | OCCASIONALLY | | | OFTEN

18. Have you ever remained apprehensive or fearful long after watching a scary movie?

NEVER | | | OCCASIONALLY | | | OFTEN

19. Do you ever become so involved in doing something that you lose all track of time?

NEVER | | | OCCASIONALLY | | | OFTEN

Do not write below this line. Experimenter use only.

Subject Number: ____________ LOD: ____________________