Human Factors and Ergonomics in Manufacturing, Vol. 14 (1) 51–67 (2004)
DOI: 10.1002/hfm.10050

Situation Awareness and Attention Allocation Measures for Quantifying Telepresence Experiences in Teleoperation

Jennifer M. Riley
SA Technologies, Marietta, GA 30066

David B. Kaber
North Carolina State University, Raleigh, NC 27695–7906

John V. Draper
Raven Research, Lenoir City, TN 37771

ABSTRACT

This study assessed the utility of measures of situation awareness (SA) and attention allocation for quantifying telepresence, the sense of being present at a remote site, in a teleoperation task scenario. Attention and SA have been identified as cognitive constructs potentially underlying telepresence. The motivation for this research was to establish an objective measure of telepresence and investigate the relationship between telepresence and teleoperation performance. Twenty-four research participants performed a virtual ordnance disposal task at varying levels of difficulty (LODs). The task involved locating, identifying, and disposing of virtual land mines in an outdoor environment using a simulated remote-control rover with a robotic arm. Performance, SA, and attention allocation were recorded along with subjective assessments of telepresence. Results demonstrated LOD effects on performance and telepresence. Regression analysis revealed LOD and attention to explain significant portions of the variance in telepresence. Results of the study provide further evidence that telepresence may share a relationship with performance, and that cognitive constructs, such as attention and SA, may serve as alternative, objective measures of telepresence. © 2004 Wiley Periodicals, Inc.

1. INTRODUCTION

Many newly developing technologies in teleoperation (remote robot control) and virtual reality (VR) are enhancing human abilities to complete complex control tasks from remote locations. Advanced human–computer interfaces for teleoperation technologies are thought to be useful for facilitating perception of a remote physical world (or, in the case of VR systems, a computer-mediated world) by using displays that provide rich visual, auditory, and haptic sensory information (e.g., Hines et al., 1995). Interface design as part of these systems is often aimed at conveying to users the feeling of being “present” in the remote or simulated environment (Loomis, 1992). It has been speculated that interfaces that facilitate this sense of “presence” also enhance human task performance with teleoperation systems (Bystrom, Barfield, & Hendrix, 1999). If users feel an association with remote task operations (i.e., as if they were personally, physically embedded within the remote site instead of existing there only through the medium of a robotic device), they may achieve more effective performance through the control interface.

The sense of presence is often called telepresence: It is this sense that is the focus of the present study. We will use "presence" and "telepresence" interchangeably throughout the remainder of the article, and define the single phenomena both words refer to in terms of the experiences of human users of teleoperators and VR. Sheridan (1992) experientially defined telepresence as the "sense of being physically present with virtual objects at the teleoperator site" and "feeling like you are actually 'there' at the remote site of operation" (p. 120). Because of its hypothesized benefit to performance, telepresence has been generally accepted as a design criterion for teleoperation systems. Some researchers have produced empirical evidence to suggest a relationship between subjective ratings of telepresence and performance (cf. Kaber, Riley, Zhou, & Draper, 2000; Witmer & Singer, 1998); however, there are still questions regarding the factors that influence telepresence and its relationship with performance and other constructs of human behavior. Research issues of particular importance include the lack of a valid and reliable objective measure of telepresence and evidence of a relationship between objective telepresence and performance. The debate regarding a possible relationship between these two measures is critical because it will ultimately determine the importance of telepresence. If there is a relationship (causal or correlational), then there may be a legitimate need to continue investigating the concept of telepresence and to establish telepresence-based design guidelines for teleoperation systems (e.g., visual interfaces, physical controls, etc.).

Many factors in teleoperation and/or VR are thought to be fundamental to telepresence/presence experiences. For example, technological factors such as stereoscopic visual cues in synthetic environment interfaces (Hendrix & Barfield, 1996), system response latency to control actions at an interface (Ellis, Dorighi, Menges, Adelstein, & Jacoby, 1997), and display type (Deisinger, Cruz-Neira, Reidel, & Symanzik, 1997) are hypothesized to affect a user's sense of association with a remote environment. Psychological/cognitive factors such as user motivation and ability to concentrate (Psotka & Davison, 1993) and susceptibility to immersion (Witmer & Singer, 1994) are suspected to be critical to experiences of telepresence. Task factors such as length of exposure (Stanney et al., 1998), difficulty/complexity (Draper, Kaber, & Usher, 1998; Sheridan, 1992), and level of automation (Sheridan, 1992) also are thought to influence telepresence.

It is the multidimensional nature of the concept that presents a challenge for researchers in terms of developing a valid and reliable objective measure of telepresence as well as relating such a measure to performance. Further, it is unlikely that a single index will be capable of adequately assessing these experiences. On this point, most researchers working in the area of telepresence would agree (Stanney et al., 1998). However, there is currently no agreement on the "best" way to measure telepresence. The only consensus is that a reliable, repeatable, and robust measure is needed (Sheridan, 1992; Stanney et al., 1998). To this end, both subjective and objective measures have been used and proposed.

1.1. Subjective Measures of Telepresence

Typically, telepresence is measured using subjective questionnaires and rating techniques (Nash, Edwards, Thompson, & Barfield, 2000). These methods involve reports, either verbal or written, of perceived levels of telepresence experienced in a virtual or remote environment during or after exposure. For example, items used to subjectively assess telepresence may ask, "On a scale from 1 to 10, rate the amount of presence you felt," or they may require subjective comparisons such as, "Which of the two environments produced the greatest amount of presence?" (Stanney et al., 1998). Some of these subjective methods have been validated as reliable measures such as, for example, the Presence Questionnaire (PQ) developed by Witmer and Singer (1994, 1998). The PQ measures telepresence along three major subscales: (a) involved control, (b) naturalness, and (c) interface quality. The items for these subscales include questions about a user's ability to control remote task events and their involvement in the task, their interactions with the environment through the various control mechanisms, and the fidelity of the sensory modalities and synthetic environment richness. Items on the PQ are aimed at capturing the degree to which characteristics of the environment add to, or take away from, telepresence experiences.

Although generally accepted as a valid means of assessing telepresence, subjective methods do suffer from limitations such as poor subject ability to accurately recall and express VR experiences, particularly in posttrial assessments. There also has been concern that items on telepresence questionnaires may be difficult for participants to understand, particularly when they explicitly refer to the term "telepresence" rather than presenting items that discuss system characteristics potentially affecting the sense of telepresence. Related to this is the issue of a lack of standardization of items included in various rating scales. Several questionnaires have been proposed and used (e.g., Ditton, 1997; Lessiter, Freeman, Keogh, & Davidoff, 2000; Lombard et al., 2000; Witmer & Singer, 1994), but the items presented are rarely the same. Consequently, it is difficult to make comparisons of results across research studies using different measures.

1.2. Objective Measures

For these reasons, there has been a call for objective measures of telepresence in the literature (Sheridan, 1992; Welch, 1999). Objective measures usually involve recording study participants' behavioral and/or physiological responses to teleoperation or VR task circumstances. These measures are often preferred over subjective measures, in part because they do not involve subjective introspection on experiences, and they can typically be administered and recorded during an experience versus making a posttrial assessment.

Several methods to objectively measure telepresence have been proposed, including the philosophy of behavioral realism. This involves observing participants and recording physical reactions to startling or unexpected events in a virtual environment (VE). These behaviors are then compared to behavioral responses to similar stimuli in a real environment (Slater & Wilbur, 1997). A potential drawback to this type of measure is that it could compromise virtual or teleoperation task performance due to artificial interruptions and disturbances (generated as part of the technique) that are not relevant to the task. Alternatively, measuring physiological state changes in users during VE exposure may involve recording posture, muscular tension, or cardiovascular function. The idea behind these measures is that people routinely experience physiological responses to stimuli in the real world (e.g., increased heart rate with the intensity of an emotional response), and similar responses should occur within virtual worlds that serve as representations of real scenarios (provided they are designed in a sufficiently realistic manner) (Barfield & Weghorst, 1993).

The previous measures also are limited in their utility for describing telepresence because a person either experiences telepresence or they do not. There are no intermediate degrees of the phenomenon in terms of existing objective measures nor do the measures allow for description of an index of telepresence. As well, some researchers have expressed concern regarding poor correlations between physiological measures and mental constructs, such as telepresence, which they are intended to describe (Prothero, Parker, Furness, & Wells, 1995). There is no strong evidence to suggest that specific physiological responses are associated with telepresence experiences (Prothero et al., 1995).

As a result of these limitations, other research has proposed objectively quantifying telepresence in terms of established mental constructs that may have similar underlying and influential factors (e.g., human perceptual processing, attentional filtering, susceptibility to vigilance decrements; Draper et al., 1998). Situation awareness (SA) and attentional resource allocation have been related to telepresence experiences and suggested as means for objectively describing telepresence in teleoperations (Draper et al., 1998). Draper et al. (1998) presented a structured attentional resource model of telepresence, defining the concept in terms of concentration on task-relevant and distracter information across local and remote (or real and virtual) environments in a teleoperation scenario. They assumed a multiple resource theory of attention and presented telepresence as a mathematical ratio of attentional resources allocated from pools associated with different modalities to local and remote stimuli (Navon & Gopher, 1979; Wickens, 1992; Wickens, Vidulich, & Sandry, 1984). Increases in the allocation of attentional resources to remote task information are hypothesized to cause increases in telepresence. That is, as users allocate more attention to the remote task, beyond that allocated to real world events, they may feel “present” in the remote site.

Draper et al. (1998) also discussed how SA might be useful for indicating levels of telepresence. Situation awareness has been defined by Endsley (1988) as “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” (p. 97). Like attention, SA can be divided across the local or remote (real or virtual) worlds in teleoperation tasks. Situation awareness can be developed on either the remote or local environments alone or on the two environments jointly. An increase in attention to the remote or local environments for achieving awareness may result in a loss of SA on the alternate environment. In terms of quantifying telepresence, this may mean that as SA on the remote site increases, telepresence also may increase. Conversely, as SA on the local environment increases, telepresence experiences may not be as strong.

Objective measures of SA include the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1988, 1995). This technique involves freezing a task or simulation at random points in time to administer queries concerning both the current and future states of, for example, a teleoperation. The queries as part of the SAGAT are developed on the basis of operator information requirements for successfully completing the given task. Research has shown that the technique of freezing a simulation to pose SAGAT queries to subjects does not substantially negatively affect task performance (Endsley, 1995). Furthermore, the SAGAT helps to avoid biasing in SA data in that participants are not aware of when queries will take place or what questions will be asked of them and therefore cannot prepare in advance for the SAGAT stops.

The primary goal of this research was to determine if established measures of SA and attention allocation, such as the SAGAT (Endsley, 1988, 1995) and signal detection theory (cf. Wickens, 1992), can be used to quantify telepresence in a teleoperation task. A secondary goal of the work was to explore the hypothesized relationships between SA, telepresence, attention, and teleoperation task performance.

2. EXPERIMENT 1

2.1. Teleoperation System

Participants were asked to perform a simulated ordnance disposal task. The task was presented using a high-performance VR system including an Intergraph workstation and Stereographics light shutter glasses. Participants controlled a robotic rover equipped with a manipulator arm and several demining (unearthing and mine disposal) tools in an outdoor VE. The rover tools included (a) an airknife used to remove virtual dirt and uncover mines, (b) a shotgun used to detonate mines that could not be safely removed from the ground, and (c) a gripper to pick-up mines with the manipulator. The simulated vehicle and tools were designed to look and function like telerobots currently used for civilian and military humanitarian demining or minefield remediation, for example, the Enhanced Teleoperated Ordnance Disposal System developed by the Department of Energy (Eisenhauer, Norman, Kochanski, & Foley, 1999). Lemhofer (1999) and Eisenhauer et al. (1999) describe remotely controlled robotic vehicles equipped with manipulators that allow users to observe objects, to use various probes for identifying the objects, and to use tools to destroy complex objects that are identified as bombs. Specifically, the systems are applicable to unexploded ordnance remediation, explosive ordnance disposal (EOD), law enforcement, and antiterrorism. The present simulation could be classified as an EOD application.

2.2. Tasks

The ordnance disposal task required participants to navigate the rover in a very large virtual setting with undulating terrain, trees, ponds, fences, and so on, and to use the demining tools to locate, uncover, identify, and neutralize landmines. Participants used verbal commands and a standard keyboard and mouse controller to drive and manipulate the robotic arm, in general. A “Wizard of Oz” setup (see Figure 1) was used, and an experimenter (the “Wizard”) executed verbal commands via a second keyboard to assist the participants in responding to system auditory warnings, controlling joint rotations of the manipulator, selecting tools, and responding to secondary system-monitoring tasks

Figure 1 Photograph of equipment setup.

(e.g., battery power levels) (The secondary task will be discussed in more detail later in this section.) The mouse controller was used to direct the motion of the rover in the forward, backward, left, and right directions.

During task performance, participants relied on an auditory mine-warning system that operated like a metal detector. Warning signals indicated to users the proximity of a mine to the rover. Once a mine was located, participants used the airknife to uncover it, and then classified it as one-of-two types, "antipersonnel" or "antitank," and determined whether it could be safely removed from the area. Participants were instructed that antipersonnel mines could be safely excavated, and they were mandated to use the gripper tool to pick up and remove the mine by transporting it to a storage bin located in a corner of the VE. When a mine was released into the storage box, the neutralization task was complete. Users were instructed that antitank mines could not be safely removed from the ground because they would explode if picked up using the gripper. Users were required to discharge the shotgun to detonate these mines where they had been virtually buried. A finite number of mines were buried in the virtual space to dictate the level of task difficulty. Participants were required to neutralize (either detonate or excavate) a total of four mines to complete a trial.

Participants completed trials at one of three levels of task difficulty: "low," "moderate," or "high." The level of difficulty (LOD) was manipulated by varying the total number of mines located in the outdoor VE and the distance between them. As the number of mines increased and the spacing between mines decreased, the difficulty decreased. The low, moderate, and high conditions included a total of 20, 15, and 10 mines, respectively. The spacing between mines for these levels corresponded to distances in the VE of approximately one rover length, two rover lengths, and three rover lengths.

While completing the teleoperated ordnance disposal task, participants simultaneously performed two secondary monitoring tasks. They were informed that the teleoperation system was equipped with two batteries and that the battery power levels were critical to system operation and control communication. The monitoring tasks required users to observe two graphics displays on battery power levels and to respond to "low battery" signals. One signal was presented on the virtual task display and was superimposed over the image of the teleoperation environment. The second signal was presented on another (portable) computer display in the real world (RW). The signals consisted of a single "light" display that appeared as a green square for normal battery conditions or a green triangle to indicate "low battery" power levels. Both signal-detection tasks appeared identical in features and operating characteristics. The signals on each display were not synchronized. A signal could appear in the RW and not in the VE (and vice versa), or it could appear on both displays. Although users could switch viewpoints in the VE to see a close-up view of the manipulator, a view of a tool display on top of the rover (indicating which demining tool was currently active), or a global view of the environment, the virtual display of the battery signal was present in all views and was consistently located on a display. Participants responded to the low-battery signals in the VE by using verbal commands, and by pressing a key on the keyboard of the portable computer to respond to low-battery signals in the RW. Figure 2 shows a photograph of the workstation and the various displays.

2.3. Subjects

Twenty-four North Carolina State University graduate and undergraduate students were recruited for participation in this study. Twenty-two males and 2 females (range = 19 to 26 years) completed the virtual task. All had 20/20 or corrected-to-normal vision as well

Figure 2 Photograph of the experimental workstation and displays.

as familiarity with personal computers. Participants were randomly assigned to the LOD groups.

2.4. Response Measures

Performance in the demining task was measured as the time-to-mine neutralization, including the search for a mine, the clearing and classification of a mine, and its disposal or detonation. The performance time was recorded in seconds and was averaged across the four mine neutralizations per test trial.

The PQ was used to subjectively assess telepresence experiences during the simulated teleoperation task. As discussed previously, this measure is designed to capture the level of perceived association with a virtual or remote task. Sample items on the PQ include, "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 1 being the lowest rating for the majority of items). A composite score was calculated on the basis of the ratings, and higher scores are considered to indicate increased telepresence (Witmer & Singer, 1994).

The PQ is typically used in conjunction with the Immersive Tendencies Questionnaire (IQ), also developed by Witmer and Singer (1994). The IQ is intended to serve as a measure of susceptibility to, or predisposition toward, immersion in virtual tasks. Like the PQ, each question on the IQ is scored on a 7-point scale, with higher scores indicating an increased tendency toward immersion (Witmer & Singer, 1994). Scores for each item were combined to yield a composite IQ score. The two questionnaires (IQ and PQ) have been found to be a valid measure of telepresence, with reliable results over multiple experiments (Nash et al., 2000; Witmer & Singer, 1998).

Mental workload also was measured during the study using the Modified Cooper-Harper (MCH) scale (Wierwille & Casali, 1983). The MCH evaluates tasks in terms of accomplishability, number of errors, and level of performance, and is designed for application to dual/multiple-task scenarios. It consists of a flow diagram with various descriptor terms regarding task difficulty and operator demand level. The terms are intended to describe a participant's perceived cognitive loading with ratings ranging from 1 (*very* *easy/operator mental effort is minimal*) to 10 (*impossible/instructed task cannot be accomplished reliably*). Mental workload was recorded as a check on the LOD manipulation. In addition, previous research has demonstrated subjective perceptions of mental workload influenced by task-difficulty manipulations to be significantly negatively correlated with subjective ratings of telepresence (PQ scores) (Kaber & Riley, 2000; Kaber et al., 2000; Riley & Kaber, 1999). This work has suggested a potential relationship between cognitive load, in general, and telepresence.

The SAGAT (Endsley, 1995) was used to assess participant SA. It allowed for assessment of SA on three levels as defined by Endsley (1995), including Level 1 SA: perception of elements in the environment, Level 2: comprehension and relation of elements to each other and the overall task goals, and Level 3 SA: projection of future task and environmental states. Example queries for each level of SA for this task are presented in Table 1. Responses to the SA queries were compared to "real" situation data recorded by the experimenter. Situation awareness was measured as the percentage of correct responses to all SAGAT queries during a test trial.

Attention allocation was measured by recording signal-detection performance in the two monitoring tasks. The total number of low-battery signals presented in the VE and RW was recorded along with the total number of signals detected by participants. There was a 20% chance of a low-battery signal during each minute of the teleoperation simulation. Hit-to-signal ratios were calculated and recorded for both the virtual and real forms of the secondary task. In addition, a ratio of attention allocation across the VE and RW displays (i.e., hit-to-signal ratio in the VE/ hit-to-signal ratio in the RW) was computed with the objective of describing telepresence.

TABLE 1. Example SA Queries

Level	Query	Possible Responses
1	What is the current tool type in use?	(a) air gun; (b) gripper; (3) gun; (d) none.
	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.
2	Can the mine currently being neutralized be safely removed from the area?	(a) yes; (b) no; (c) no mine being neutralized.
	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.
3	What will happen if the current mine is picked up?	(a) nothing; (b) explosion; (c) no mine being neutralized.
	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.

2.5. Procedures

Initially, participants were familiarized with the study and the VE system and were asked to complete the IQ. This was followed by a three-part training session including instruction and practice on performing the experimental tasks. In the first training period, participants learned the teleoperated ordnance-disposal task, specifically how to navigate the rover using the mouse controller, how to locate and identify mines, and how to use the interface controls to manipulate the robotic arm. The second training period involved practice at completing the teleoperation task while monitoring the secondary-task displays for low-battery signals. All participants completed a third training period, which provided an explanation of the SAGAT queries and how to respond to the queries during test trials. Each person was allowed to practice several mine neutralizations while completing the monitoring tasks and being exposed to SAGAT freezes. Participants not able to successfully complete an ordnance disposal were not allowed to continue in experimental test trials. The total time for the experiment familiarization and training was approximately 2 hr.

Experimental testing occurred on the following day (or within no more than 2 days of the training). The testing protocol began with explanations of how to complete the PQ and the MCH, which was immediately followed by participation in two test trials. During the first 2 min of every trial, no low-battery signals were presented to allow participants to become comfortable in performing the teleoperation task. Following this period, low-battery signals occurred at random. Similarly, no SAGAT queries/freezes were administered during the first 5 min of any trial so that participants could reasonably achieve SA on the task and environment. After the 5-min period, three SAGAT freezes occurred at random times during each trial. Time-to-mine-neutralization was recorded after each of the four mines was neutralized. Telepresence and mental workload data were collected immediately at the close of each trial. The total experimental test time for each subject was approximately 2 hr.

2.6. Hypotheses

Teleoperation performance time and telepresence were hypothesized to be significantly affected by manipulations in LOD, with performance and telepresence degrading with increases in LOD. Based on previous findings (Kaber & Riley, 2000; Riley & Kaber, 1999), increases in the level of task difficulty were expected to result in increases in time to complete the neutralization task and decreases in reported telepresence experiences. Telepresence scores were hypothesized to be significantly positively related to performance and SA on the mine-neutralization task, and negatively related to mental-workload ratings. Telepresence also was expected to be related to performance in the monitoring task shown through the RW display, with increases in attention allocation for signal detection in the RW, corresponding to decreases in perceived association with the VE and elements of the mine-neutralization task.

3. DATA ANALYSIS AND RESULTS

3.1. Effects of Level of Difficulty

A between-subjects experimental design was used with LOD as the independent variable. Repeated measures were recorded on all responses for each participant. Statistical analyses included a one-way analysis of variance (ANOVA) and multiple comparisons using Duncan's multiple range (MR) tests. Though some researchers may object to the use of Duncan's MR tests as being too lenient and possibly leading to the occurrence of large experiment-wise errors, it has been demonstrated to be a sensitive test that is likely to detect true differences among means, if they exist (Chew, 1977, pp. 20–21). Duncan's test also has been reported to be an acceptable and reliable indicator of statistically significant differences when used after an overall F test is found to be significant (Chew, 1977, p. 30), as was the case with the statistical analyses presented here.

There were significant effects of LOD on average performance time, $F(2,21) = 5.89$, $p < 0.02$, and subjective telepresence scores, $F(2,21) = 4.69$, $p < 0.05$. Duncan's MR test revealed that participants performed the task significantly faster ($p < 0.05$) under the lowest LOD (mean = 313 s) than under the moderate (mean = 504 s) or high (mean = 526 s) difficulty levels. The means for the latter two LODs were not significantly different. With respect to telepresence, Duncan's MR test revealed that participants in the high LOD group reported significantly lower ($p < 0.05$) telepresence than those in the low or moderate groups, which were not significantly different (see Figure 3).

Counter to our expectations, results on SA and attention revealed no significant effect ($p > 0.05$) of LOD. Attention and SA were observed as dependent measures in this study. They were not controlled variables as part of the analysis of effects of levels of task difficulty on performance and telepresence. Thus, there was no evaluation of the effects of manipulations of SA and attention on performance or telepresence. However, a regression model, presented in the next section, was intended to assess relationships between these variables (SA and attention) and telepresence in an attempt to establish whether they might serve as better indicators of presence experiences as compared to subjective measures, and so on.

3.2. Regression Analysis

To assess the utility of SA and attention for explaining telepresence experiences, the percent-correct responses to SAGAT queries (SA) and the attention allocation ratio (ATTENTION) were included in a multiple linear regression (MLR) model to predict PQ scores. Because of potential effects of the independent variable on telepresence and the other response measures, LOD also was included as a predictor in this model. Finally, to account for potentially significant individual differences in susceptibility to immersion, the

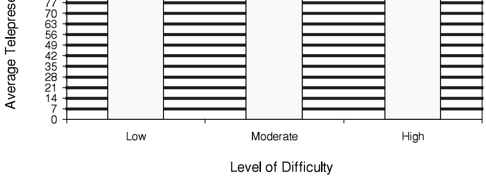

Figure 3 Mean subjective ratings of telepresence across LODs.

IQ score was included as a model parameter. In general, the inclusion of these variables in the predictive model of telepresence was aimed at avoiding biases in the parameter estimates for SA and attention that might have occurred if variance due to LOD and immersive tendencies were falsely attributed to SA or attention.

Although the effect of LOD on subjective ratings of telepresence did not correspond with the influence (or lack of influence) of LOD on average SA and attention allocation in the ANOVA analyses, it was considered important to structure a predictive model of telepresence on the basis of SA and attention behavior for the reason that one would not expect SA or attention to completely explain telepresence if they are all unique phenomenon. It could not be expected that the results of, for example, the SAGAT queries would perfectly predict the outcomes of the PQ given the differences in the nature of the instruments and the focus of the items/questions as part of each. However, it is possible that SA or attentional resource allocation would explain aspects of telepresence, or perceptions of telepresence.

With this in mind, the regression model used to assess the utility of the SA and attention allocation measures for describing telepresence was structured as:

$$\text{Telepresence} = \beta_0 + \beta_1\,\text{LOD} + \beta_2\,\text{SA} + \beta_3\,\text{ATTENTION} + \beta_4\,\text{IQ} + \epsilon$$

Because of potential correlations between variables in the model (e.g., IQ, SA, and ATTENTION), multicollinearity analysis was conducted to ensure that there were no strong interrelationships among the regressors. No evidence of multicollinearity was revealed through diagnostic procedures. Three multicollinearity procedures were used in conjunction to test for relationships among the regressors, including the variance inflation factor (VIF), the condition index (CI), and tolerance values. All VIF values were less than “10” and close to “1,” all CI values were outside of the critical range (30–100), and all tolerance values were close to “1,” indicating no evidence of multicollinearity (SAS/STAT, 1990).

Residual analysis and influential diagnostics were used to account for any assumption violations and to investigate potentially influential points. Examination of residual plots revealed no violations of regression assumptions in terms of linearity. However, there was some evidence of nonnormality and nonconstant variance (e.g., significant Shaprio–Wilks tests, atypical normal probability, and residual plots). Consequently, a log transform was applied to PQ responses. Outliers were removed subsequent to transformation of the dataset and were identified on the basis of statistical measures including covariance ratio values (<0.75 or 1.25), studentized residual (Those 2.0 in magnitude were considered suspect.), and anecdotal observations during the data collection (e.g., participant problems with navigation). Three points were identified as outliers as a result of considering all of these diagnostics in conjunction.

Table 2 summarizes the results of the regression model on the log transform of PQ. In general, results indicated that the model in LOD, SA, ATTENTION, and IQ adequately described telepresence experiences, $F(4,39) = 10.623$, $p = 0.0001$. Approximately 52% of the variability in log (PQ) scores was explained by this model. The results also indicated that at the $\alpha = 0.05$ level, LOD, attention, and IQ scores were significantly related to PQ scores. However, the significance of SA (percent-correct responses to SAGAT queries) did not meet the conventional alpha criterion, with $p = 0.0708$.

TABLE 2. Summary of Regression Analysis Results on Log(PQ) Model

t-tests on Individual Parameter Estimates

Variable	*df*	Parameter Estimates	t Value for H_0	Prob > T
Intercept	1	4.391981	21.390	$p < .0001$**
LOD	1	−0.117778	−4.918	$p < .0001$**
AVGSA	1	−0.336406	−1.858	$p > .05$
ATTNRAT	1	0.071746	2.478	$p < .02$*
IQ	1	0.006401	3.994	$p < .001$**

*significant at the $\alpha = 0.05$ level.
*significant at the $\alpha = 0.001$ level.

3.3. Correlation Analysis

In addition to the previous analyses, Pearson product–moment correlation coefficients were computed to assess any relationships among the various response measures. Results of the correlation analyses revealed significant relationships between telepresence and performance and workload. PQ scores were significantly negatively related to average time-to-mine-neutralization ($r = -0.33, p < 0.05$) (see Figure 4). Telepresence was significantly negatively correlated with mental workload ($r = -0.62, p < 0.05$).

4. DISCUSSION

4.1. Level of Difficulty Effects on Performance and Telepresence

The LOD effect on performance was anticipated. The increased number of mines under the low-level difficulty condition and the decreased spacing among mines increased the likelihood of driving over and locating a mine. Quickly locating mines led to a decrease in overall neutralization times. Participants performing under higher levels of difficulty appeared to abandon any type of search strategy when they could not locate mines quickly. They often drove the rover aimlessly in the virtual space, failing to adequately explore areas along the perimeter of the environment. Ultimately, this meant failing to locate

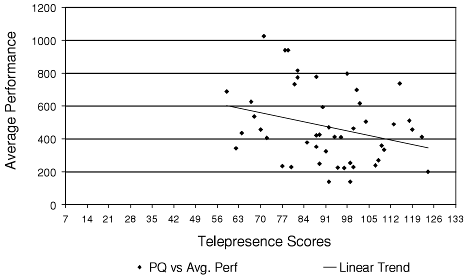

Figure 4 Telepresence scores versus average performance.

mines and resulted in significantly longer search and overall neutralization times for participants in these groups.

The PQ scores observed in this study were consistent with scores observed in previous research (Riley & Kaber, 1999). Lower telepresence scores under the high LOD setting may have been due to participants' feelings of frustration in not being able to locate mines quickly. Users experiencing frustration with the task or interface might have felt unable to control events in the VE, and thus disengaged from the teleoperation task. This is a typical negative emotional response to the synthetic environment, and may have produced lower telepresence scores. Under the high LOD, mines were placed in the VE such that a distance greater than three times the length of the rover separated them. The difficulty in locating mines caused longer navigation and search periods, which increased the amount of time users spent interacting with the VE in a relatively passive manner (navigating) as compared to actively controlling the robot arm and tools or manipulating objects during mine neutralization. The lack of active control of the robot for extended periods of time may also have caused users to disengage from the task. This is a typical reaction to low task demands. Either of these experiences could have led to detachment from the task and lower PQ scores in the high LOD group. However, the implications of the two conditions may be different. A negative emotional reaction may have an effect on future willingness to engage in the virtual task whereas satisfaction with low task demands may not.

4.2. Situation Awareness and Attention Allocation

The lack of an LOD effect on SA and attention allocation across the two environments (VE and RW) was not expected. As previously discussed, the difficulty level in the demining task was manipulated by varying the total number and spacing of mines virtually buried in the area. This type of manipulation mostly affected the time required to navigate the rover and search for mines in the VE; however, it did not alter the information requirements, or the specific steps, to complete the task. Thus, there was no change in the elements to be perceived or how they were comprehended to project future states of the task with changes in LOD. The low correlation between the SA requirements of the task captured in the SAGAT queries and the basis for the LOD manipulations most likely caused the lack of a significant LOD effect on SA.

With respect to LOD effects on ATTENTION, the range of LODs (or differences between LODs) may not have been substantial enough to affect participants' ability to detect low-battery signals across the two environments. In general, users were able to detect on average 64 and 47% of the signals presented through the VE and RW displays, respectively. This may not have been the case for a larger range of LODs. For example, at very low difficulty levels (e.g., 30–40 closely spaced mines), participants might experience decrements in monitoring task performance due to a need to frequently attend to subtasks in the demining scenario. That is, the increased probability of locating mines would lead to more active control—manipulation of the robot arm, selection and use of the demining tools, or identification and transportation of mines—reducing attentional resources available for signal detection in the RW display. Similarly, at very high LODs (e.g., five mines), participants might experience increases in monitoring performance in the VE or RW displays due to long periods of search and navigation. In either case, there would be larger differences in the percentage of signals detected across the two displays. Such extreme LOD conditions were not examined in this study and should be considered in future work to more clearly elucidate any important relationship between teleoperation task difficulty and human operator attention behavior.

4.3. Implications of Regression Analysis

The parameter estimates of the regression model indicated that LOD shared a negative relationship with telepresence and had the strongest effect on PQ scores. These results are consistent with those of the ANOVA procedure on the LOD model and further support the finding that task difficulty is a determinant of telepresence experiences, even when accounting for operator SA, attention behavior, and immersive tendencies.

It was expected that increases in SA would result in increases in telepresence experiences; however, the parameter estimate for SA indicated a negative relationship among the two concepts. This result may be explained by the design of the virtual interface. In assessing SA, participants were queried regarding current tool use, tools to be used in the future, the types of mines being neutralized, proximity of the gripper to a mine, and so on. To achieve SA on these elements, participants often needed to alternate between viewpoints using the mouse controller. For example, to clearly view the tool display as part of the rover, users 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). It is possible that using the interface controls, such as mouse buttons, to toggle between the various viewpoints to develop SA on components of the tasks may have detracted from task involvement. Some also may have experienced a decrease in task involvement when attempting to maintain global SA when, for example, attempting to remember the total number of mines eliminated or the next stage of the task, and so on. This may have reduced attentional resources available for experiencing telepresence and may explain the observed decrease in PQ scores with increases in SA.

Parameter estimates for ATTENTION and IQ indicated positive relationships with telepresence. As the ratio of attention allocation across the task environments (VE and RW) increased, telepresence scores increased. Thus, the model predicted that as participants paid more attention to signal detection in the VE, the sense of telepresence increased. The VE engaged those in the virtual teleoperation task, particularly at the lower LOD, when disposing and detonating mines. Immersion in the task may have contributed to allocation of attention to the VE over the RW and increased the probability of detecting signals in the VE (or a decreased probability of detecting signals in reality). This, in turn, might have led to stronger telepresence experiences in the VE. For those with greater susceptibility to immersion (higher IQ scores), the strength of telepresence experiences was greater. This result was expected and is consistent with previous research (Kaber et al., 2000).

4.4. Implications of Correlation Analysis

The results of the correlation analysis indicated a positive relationship between telepresence and the construct of performance. Thus, as performance increased (time to neutralize mines decreased), association with the VE, as reported using the PQ, was stronger (see Figure 4). This finding is consistent with previous research (Kaber et al., 2000; Riley & Kaber, 1999) and indicates that telepresence may be critical to task performance (or vice versa). In general, the results support the argument for further examination of telepresence as a potential design criterion in teleoperation (along with other factors such as level of task difficulty). More studies are needed and should be designed to clearly elucidate the relationship between objective measures of telepresence and performance.

As participants' perceptions of cognitive load increased, their association with the VE and involvement in the task was degraded. This result also is consistent with previous research (Kaber et al., 2000). The negative relationship between telepresence and mental workload may be attributed to user frustration and perceived time pressure in completing the mine neutralization task. Under the higher LOD conditions, some participants became frustrated when they were not able to locate mines, which may have led to detachment from the task as well as an increased sense of workload.

Given the relationships among telepresence, performance, and workload, it may be necessary for researchers and designers of teleoperators and virtual-task environments to develop systems that alleviate user cognitive loading. For example, automating components of the task might "free up" user mental resources, possibly reducing frustrations associated with multitasking and ultimately increasing telepresence and performance.

5. CONCLUSIONS

This study has provided potentially useful information on the impact of level of task difficulty on human performance and telepresence in a teleoperation task. It also provided preliminary results on, and insight into, the utility of measures of SA and attention (along with other variables) for explaining perceptions of telepresence. The results of the regression analysis revealed that SA and attention share important relationships with telepresence, and might be useful for quantifying telepresence experiences in human use of teleoperation and VR systems. Previous studies have not investigated a model of telepresence based on existing measures of cognitive constructs in an effort to identify objective indicators of telepresence. More research should be conducted in this direction to clearly establish any relationships between SA, attention, and telepresence and, in general, to provide telepresence with a stronger pedigree in accepted aspects of cognition. In particular, future studies may need to evaluate system-design characteristics to determine the best designs for supporting SA while fostering telepresence experiences.

This research has provided additional information of relationships between subjective ratings of telepresence and teleoperation task measures, including performance and workload. The correlation results on telepresence and performance presented here further support the contention that telepresence is a concept deserving consideration in the design of teleoperation systems and VE interfaces. The correlation results on telepresence, workload, and performance suggest a need for research that identifies levels of workload that are not detrimental to telepresence and performance.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF) under Grant IIS-0196342. Work on this article by Jennifer Riley was sponsored by the U.S. Army Research Laboratory (ARL) under the Advanced Decision Architectures Collaborative Technology Alliance (Cooperative Agreement DAAD19–01–2-0009). The authors thank Mohamed Sheiknainar and Laura Hughes for their contributions to this research in the development of the VE simulation and data collection. The opinions presented are those of the authors and do not necessarily reflect the views of the NSF or the ARL.

REFERENCES

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

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

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

Deisinger, J., Cruz-Neira, C., Reidel, O., & 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, & R. Koubek (Eds.), Design of computing systems: Social and ergonomic considerations (pp. 881–884). Amsterdam: Elsevier.

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.

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

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

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

Endsley, M.R. (1988). Situation global assessment technique (SAGAT). Proceedings of the National Aerospace and Electronics Conference (pp. 789–795). New York: IEEE.

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

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

Hines, B., Hontalas, P., Fong, T., Piguet, L., Nygren, E., & Kline, A. (1995, July). VEVI: A virtual environment teleoperations interface for planetary exploration. Proceedings of the SAE 25th International Conference on Environment Systems. San Diego.

Kaber, D.B., & Riley, J.M. (2000, October). Task difficulty and user motivation effects on performance and telepresence in a teleoperation task. Proceedings of the 3rd International Conference on Engineering Psychology and Cognitive Ergonomics. Aldershot, UK: Ashgate.

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

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

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2000, March). Development of a new cross-media presence questionnaire: The ITC sense of presence inventory. Paper presented at the 3rd International Workshop on Presence, Delft, The Netherlands. Available: http://homepages.gold.ac.uk/immediate/immersivetv/publications2.htm

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

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

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

Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. Psychological Review, 86(3), 214–255.

Prothero, J.D., Parker, D.E., Furness, T.A., & Wells, M.J. (1995). Towards a robust, quantitative measure of presence. Proceedings of the Conference on Experimental Analysis and Measurement of Situation Awareness, 359–366.

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

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

SAS/STAT. (1990). User's guide (Version 6, 4th ed.). Cary, NC: SAS Institute.

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

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

Stanney, K.M., Salvendy, G., Deisigner, J., DiZio, P., Ellis, S., Ellison, E., 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.

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

Wickens, C.D., Vidulich, M., & Sandry, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks. Human Factors, 26, 533–543.

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

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

Witmer, B.G., & Singer, M.J. (1994). Measuring immersion in virtual environments (Tech. Report No. 1014). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

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