

2010

Trust, situation awareness and automation use: exploring the effect of visual information degradations on human perception and performance in human-telerobot

Minglu Wang
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Industrial Engineering Commons](#)

Recommended Citation

Wang, Minglu, "Trust, situation awareness and automation use: exploring the effect of visual information degradations on human perception and performance in human-telerobot" (2010). *Graduate Theses and Dissertations*. 11266.
<https://lib.dr.iastate.edu/etd/11266>

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

**TRUST, situation awareness and automation use: exploring the effect of visual
information degradations on human perception and performance in human-telerobot
system**
by
Minglu Wang

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Industrial Engineering

Program of Study Committee:
Richard T. Stone, Major Professor
Gary Mirka
Stephen Vardeman

Iowa State University

Ames, Iowa

2010

Copyright © Minglu Wang, 2010. All rights reserved.

TABLE OF CONTENTS

LIST OF FIGURES.....	v
LIST OF TABLES.....	vii
ABSTRACT.....	viii
CHAPTER 1 INTRODUCTION.....	1
A. BACKGROUND	1
B. OBJECTIVES	4
C. RESEARCH HYPOTHESES.....	4
D. HUMAN-ROBOT INTERACTION (HRI)	5
E. THESIS ORGANIZATION.....	5
CHAPTER 2 LITERATURE REVIEW.....	7
A. OVERVIEW	7
B. DEFINITIONS OF TERMS	7
1. <i>Levels of automation</i>	7
2. <i>Trust in automation</i>	12
3. <i>Reliability of Automation</i>	19
4. <i>Trust and situation awareness</i>	20
C. LIMITATIONS AND GAPS	22
CHAPTER 3 MATERIALS AND METHODS	25
A. OVERVIEW	25
B. EXPERIMENTAL TASK	26
C. VARIABLES	27

1. <i>Independent Variables</i>	27
2. <i>Dependent Variables</i>	32
D. PARTICIPANTS.....	34
E. MATERIALS	34
1. <i>IowaBot User Interface</i>	34
2. <i>Equipment</i>	35
F. PROCEDURE	37
CHAPTER 4 RESULTS AND DISCUSSION	40
A. OVERVIEW.....	40
B. PERFORMANCE	42
1. <i>Performance and LOAs</i>	42
2. <i>Performance and Visual System Degradations</i>	43
C. TRUST.....	44
1. <i>Trust and LOAs</i>	44
2. <i>Trust and Visual System Degradations</i>	44
D. SITUATION AWARENESS (SA).....	45
1. SA AND LOAs.....	45
2. SA AND VISUAL SYSTEM DEGRADATIONS.....	46
E. MENTAL WORKLOAD	47
F. DISCUSSION	48
1. <i>Level of Automation</i>	48
2. <i>Visual Information Degradations</i>	50

3. <i>Trust and Situation Awareness over time</i>	51
4. <i>Correlation of Responses</i>	57
CHAPTER 5 CONCLUSIONS AND RECOMMENDATION	60
A. CONCLUSION	60
B. RECOMMENDATIONS FOR FUTURE RESEARCH	62
APPENDIX A SART 10D RATING	64
APPENDIX B NASA TLX	65
APPENDIX C TRUST IN AUTOMATION CHECKLIST	66
APPENDIX D PRE-EXPERIMENTAL QUESTIONNAIRE	67
BIBLIOGRAPHY	68
ACKNOWLEDGMENTS	73

LIST OF FIGURES

Figure 2.1 Four stages of a complex human-machine task (Parasuraman et al, 2000)	9
Figure 2.2 Man-Machine Model of Trust taken from Muir (1994).....	16
Figure 2.3 A conceptual model of the dynamic process (Lee & See 2004)	17
Figure 2.4 Subjective rating of trust in automation (Jian, Bisantz, & Drury, 2000)	18
Figure 2.5 Endsley’s model of SA and Mental Model (2000)	21
Figure 3.1 Selected levels of automation in Sheridan’s taxonomy of LOAs (2002).....	28
Figure 3.2 Block diagram representing the internal and external factors of robot system.....	29
Figure 3.3 Success tree for human-telerobot system	31
Figure 3.4 IowaBot User Interface	35
Figure 3.5 Experimental Workspace	36
Figure 3.6 IowaBot Configuration.....	36
Figure 3.7 Welding Room Configuration.....	37
Figure 4.1 Mean Hit Rate of LLA and HLA in permanent error session.....	43
Figure 4.2 Mean Hit Rate under two types of error in LLA.....	44
Figure 4.3 Mean “harmful” rating of trust questionnaire under two types of error in HLA ..	45
Figure 4.4 Mean “Instability” rating of SART under two types of error in HLA	47
Figure 4.5 Mean “Complexity” rating of SART under two types of error in HLA	47
Figure 4.6 Mean “familiar” rating of trust questionnaire by day in LLA.....	52
Figure 4.7 Mean “familiar” rating of trust questionnaire by day in HLA	53
Figure 4.8 Mean “Attentional Supply” level of SA by day in LLA	54
Figure 4.9 Mean “Attentional Supply” level of SA by day in HLA.....	54
Figure 4.10 Mean “Concentration” rating of SART by day in LLA	55

Figure 4.11 Mean “Concentration” rating of SART by day in HLA.....56

Figure 4.12 Mean “Understanding” level of SA by day in LLA.....57

LIST OF TABLES

Table 2.1 Degrees of Automation (After Sheridan, 2002, p. 62)	10
Table 3.1 Experiment Design	25
Table 3.2 Selected levels of automation in Endsley and Kaber's taxonomy of LOAs (1999)..	27
Table 3.3 Failure rates for selected influence elements of robot system on visual quality	30
Table 3.4 SART Dimensions	33
Table 3.5 First Day Experiment.....	39
Table 3.6 Second Day Experiment	39
Table 4.1 Means of Responses in LLA and HLA groups.....	41
Table 4.2 Summary of the analyses of variance by LOAs in two types of error sessions.....	49
Table 4.3 Summary of the analyses of variance by types of error in two LOAs groups.....	50
Table 4.4 Summary of the analyses of variance by day in two LOAs groups	51
Table 4.5 Summary of correlation between trust and SA at combined conditions	58
Table 4.6 Summary of correlation between trust and NASA TLX at combined conditions ...	59

ABSTRACT

Today's military and industry increasingly uses human-robot system to perform complex tasks, such as firefighting. Automated systems that support or even make important decisions require human operators to understand and trust automation in order to rely on it appropriately. This study used a real human-telerobot system performing a firefighting task in an unknown welding room to examine the effects of two different levels of automation associated with intermittent and permanent visual system degradation on human performance, trust in automation, mental workload and situation awareness.

Twenty-four participants were divided into two groups based on the level of automation use. Each participant completed a series of three 30-minutes sessions in which he or she was required to explore the threat targets in an unknown "hazard" welding room. Results indicated a significant difference between low and high level of control in collision rate when permanent error occurred. And in low level automation group the type of error had a significant effect on the collision rate, while it had a significant effect on situation awareness dimensions in both groups. Generally, in the experiment high level of automation had better performance than low level of automation especially if it is more reliable, suggesting that subjects in the high level of automation group could rely on the automatic implementation to perform the task more effectively and more accurately.

CHAPTER 1 INTRODUCTION

A. Background

During rescue operations firefighters often encounter life-dangerous situations. As technology evolves, people are gradually realizing that robots can be designed to eliminate the exposure of firefighters on fire and toxic smoke, and the possibility of getting killed or injured. Characteristic tasks for mobile robot in the context of search and rescue missions are the exploration of unknown regions (Driewer et al., 2005). The remote operator provides guidance and keeps track of the overall situation in a safe place outside the disaster area.

In South Korea, the Hoya Robot company has developed a robot that one day might help save the lives of both victims and firefighters. The Firefighters Assistant Robot can scout burning buildings when conditions may be too dangerous for humans and size up the scene as well as check for victims who may be trapped inside. The unit can operate for up to 30 minutes and withstand temperatures of up to 320 degrees F. The speedy super robot can cover over one foot per second and can be remote controlled from 54 yards away. It's meant to quickly enter a fire scene where it can transmit image and sound and send back data on, temperature, smoke and gas. About 100 remote-controlled robots were sent to Korean fire stations for testing since last year. In UK, West Yorkshire Fire Service and JCB developed a robotic firefighter, Fire Spy, which can go into the heart of fires and remove flammable or dangerous chemicals in order to save human firefighters' lives. It is based on a tough JCB vehicle but has been developed to withstand temperatures of up to 800 degrees centigrade. The operator can see what is happening in the blaze through two cameras, infra-red and standard, which beam back video pictures. At the front is a powerful grabbing arm. In the United States, in 2007 the Virginia Department of Transportation, using state and federal

money, bought four wireless remote-controlled firefighting robots for four towns in the state's Hampton Roads area at the southern end of Chesapeake Bay, primarily for dealing with fires in tunnels that connect the communities. A rural northeastern Pennsylvania city applied for a federal grant to buy an Austrian-made firefighter robot LUF60 to ventilate industrial buildings in cases of accidents and fires. In Texas, Dallas Fire Rescue officials said they also planned to apply for a grant to buy an LUF60 for fires in high-rise buildings and warehouses. (Bixby, 2007)

However, during remote controlling operators' mental workload is high and situation awareness is low when facing unexpected situations which may lead to human errors and thus the task could not be failed. Accordingly, more and more human factor researchers are focusing on finding an effective way to reduce the mental workload and enhance the situation awareness of operators. In particular, they consider the tele-robot as a teammate and try to allocate the work between human and tele-robot.

Developing autonomous robot techniques has been one of the major trends in industry and military which attempt to reduce operator's workload. Rossum's Playhouse (RP1) as a two-dimensional firefighter robot simulator is designed to be a tool for developers working on robot navigation and control logic. Every time the robot receives the distance from its left and right sensors when it is searching and extinguishing the targets, that data will be used to calculate the next movement in order to avoid obstacles and hits. After some experimentation with the simulation and gaining some level of the satisfaction in observed behavior and performance, many developers and researchers have tried to implement real robots with automation.

However, Parasuraman and Wickens (2008) pointed out that when automation is applied to information analysis or decision-making functions, it must be considered in choosing appropriate levels and stages of automation which lead to differential system performance benefits and costs. So the question of “what functions are performed by automation and to what degree” should be answered when designing the human-robot systems.

Since human operators misuse and disuse the automated systems in ship navigation and other operations occur often (Parasuraman & Riley, 1997), more and more researcher have started to analyze the factors of influence, a major contributing factor of designing the human-automation system is trust, a cognitive state that usually influences the actual, behavioral dependence on automation. At the same time, even the best automation can be unreliable and untrustworthy at times. Loss of trust in an automated system acting as part of a human-machine team may have harmful effects on the team’s overall performance.

However, none of these researches has examined the relationship between an operator’s performance in time-critical tasks and the appropriateness of their trust in the different levels of automation (LOAs) in real human-telerobot system with the conditions under which the system’s performance degraded.

The current study investigated the impact of the perception reliability on human performance, trust in automation, mental workload and situation awareness in a human-telerobot system regarding low level of automation and high level of automation control modes. First, we discuss the previous literature on which contemporary trust in automation research builds. The collection of literature in this effort draws primarily on sources from cognitive psychology and human factors domains. Drawing on early theoretical studies, many researches focus on explored mental domain, drawing on disparate areas of study to

form an integrated concept. Additionally, the current study is manipulating human operators' trust in levels of automation.

B. Objectives

This research examined how performance of the real human-telerobot system is affected by a human operator's trust in different levels of automation when the quality of visual information degrades. It also generalizes the findings to future studies of human-telerobot system. Specifically, this study:

- Assessed the human performance and trust in automation regarding the visual information degradation as well as the levels of automation
- Analyzed the effect of varying levels of automation on human-telerobot system
- Evaluated the effect of varying visual information degradations on human-telerobot system

C. Research Hypotheses

Based on the findings of current literature and the rationale described, the following are hypotheses regarding how reliability and level of automation affect trust, performance, situation awareness and mental workload:

- Trust in automation increases when increasing use of automation
- Trust in automation decreases when visual system degrading, especially under permanent error is worse than under intermittent errors
- Human performance increases when used high level of automation
- The degradation of visual system have negative effects on human performance

D. Human-Robot Interaction (HRI)

HRI is a field of study and a discipline that has gained industry and military attention in recent years because it promises to reduce costs and increase performance. In particular human augmentation, it outlines the future of robotics. Although autonomous robotic systems perform remarkably in structured environments, interacted human-robotic systems are superior to any autonomous robotic systems in unstructured environments that demand significant adaptation.

HRI incorporates the study of multiple domains to assess the complex relationship between humans and the robot systems such as human factors engineering, system safety and training. The current research is relevant to all of them by exploring the connection between cognition and performance as it investigates how an individual's performance is affected by trust in automation, by examining the contribution of the relationship between trust and levels of automation to improving human-robot system performance so as to reduce errors and increase safety and by implicating for training regarding trust acquisition and development related to information automation.

In sum, the study of HRI is a multifaceted approach to achieving a thorough understanding of the relationship between human, robot system, and automation.

E. Thesis Organization

Chapter 2 reviews literature on the categorization of level of automation, trust in automation, reliability of automation and the relationship between trust and situation awareness and Chapter 3 describes the research methodology and experiment used to test the hypotheses above. Chapter 4 and 5 present results and analysis, concluding with a discussion of directions for future research regarding trust in automation in human-telerobot system

domain. Appendix A contains SART 10D Rating Sheet used for evaluating the situation awareness of subjects when they controlled or supervised tele-robot in an unknown hazard environment in the experiment. Appendix B is the NASA Task Load Index (TLX) administered to participants, and Appendix C is the questionnaire of trust in automation, and Appendix D shows the Pre-Experimental Questionnaire used at the beginning.

CHAPTER 2 LITERATURE REVIEW

A. Overview

The current study proceeds from a collection of literature regarding trust in automation. We first review the definition in the field of level of automation. Then, we focus on trust in automation, pointing out critical relevant terms from previous studies which are the groundwork for the current study. We discover the importance of the reliability of automation and also identify and discuss the relationship between trust and situation awareness. Finally, we present the limitations and gaps in the literature that motivate the formation of the current work.

B. Definitions of Terms

1. Levels of automation (Human-Automation Interaction)

Automation is popular in critical systems in industry and military and increasingly in everyday life. Parasuraman and Wickens (2008) reviewed that many researches of human performance in automated systems have been conducted over the past 30 years. However, modeling has been and will continue to be framed by the empirical findings of field studies, and will continue to inform the design of automated systems for effective and safe human use. In many systems the physical danger and the required precision, together with the time constants of the systems, combine to make direct physical control inappropriate. Humans assume the role of a supervisory controller, interacting with the system through different levels of manual and automatic control (Sheridan, 2002; Lee & Moray, 1994). With increasing complexity of automation comes increasing variability in performance of human-robot system. Automation now assists in several areas of task performance, from initial

information acquisition to analysis of options, to selecting and implementing a course of action (Sheridan, 2002).

However, even the best automation, which dramatically challenges satisfaction, performance, and safety, can be unreliable and untrustworthy at times. The most advanced automation still requires humans to identify and interpret failures. Automation systems bring the argument of the role of humans in complex systems and even the nature of human cognition (Sheridan, 2002; Schmorrow, Stanney, Wilson, & Young, 2006). Therefore, as technological innovation and promised economic benefits are likely to drive even more automation, there is now an extensive science base of empirical findings on human-automation interaction which designers attempt to make the appropriate trade-offs to determine which functions to automate (Ahlstrom et al., 2005; Hawley et al., 2005; National Aeronautics and Space Administration, 2008).

Parasuraman, Sheridan and Wickens (2000) provided a model of human-automation interaction that addressed what aspects of a task should be supported and how much support should be provided. They discussed 4 general stages of information processing: (a) information acquisition, (b) information analysis, (c) decision making, and (d) action implementation, with each stage having its own LOA scale (see Figure 2.1). Parasuraman (2008) described stages in details that “automation at Stage 1 involves acquisition of multiple sources of information and includes sensory processing, preprocessing of data, and selective attention. Stage 2 involves manipulation of information in working memory and cognitive operations such as integration, diagnosis, and inference, occurring prior to the point of decision. Stage 3 involves decisions based on such cognitive processing. Stage 4 entails an action consistent with the decision choice.”

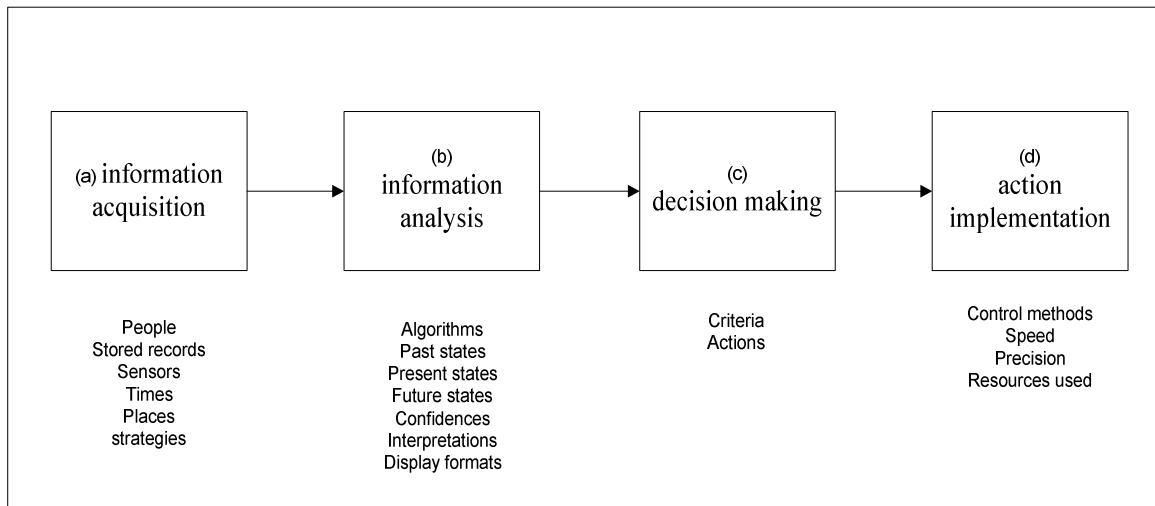


Figure 2.1 Four stages of a complex human-machine task (Parasuraman et al, 2000)

Parasuraman and Wilson (2008) combined Stage 1 and Stage 2 (“information automation”), Stage 3 and Stage 4 (“decision automation”), and distinct one from the other by pointing out the different brain regions and different mental resources each type uses. These types of automation vary greatly regarding the perception ability of the automation when it is not being relied upon. Unlike these decision aids, information automation does not give values to the possible courses of action. Thus, Information automation may promote superior performance than decision automation because the user must continue to generate the values for the different courses of action. Information automation even makes it possible for users to dynamically balance their attention between the information from the automation and the raw data (Wickens, Gempler, & Morpew, 2000).

Thus, to keep the human in the system, Stage 3 and Stage 4 automation are not considered in the present study. Also, given that the difficulty of the visual searching task is in detecting and recognizing targets, it is during the first two stages of information processing (information acquisition and diagnosis) that observers will need help.

In determining how much automation should be provided, Sheridan (2002) discussed the levels of automation at any chosen stage of automation. Levels of automation, which is presented by eight-level scale of degrees, are defined by the degree of control, autonomy, and responsibility shared between the automation and the user (see Table 2.1). At the extreme lowest level, the user has solo control, autonomy and responsibility while this is the case for automation at the extreme highest level. He also suggested that the primary evaluative criteria for determining the level of automation should include consideration of the impact of automation on workload, situation awareness, trust in automation and skill degradation (Hancock & Scallen, 1996; Parasuraman & Riley, 1997; Sheridan, 1992). In addition, further consideration should be made for secondary criteria such as the effects of automation reliability because of its impact on user trust and reliance.

Table 2.1 Degrees of Automation (After Sheridan, 2002, p. 62)

A Scale of Degrees of Automation

1. The computer offers no assistance; the human must do it all.
 2. The computer suggests alternative ways to do the task.
 3. The computer selects one way to do the task AND
 4. ...executes that suggestion if the human approves, OR
 5. ...allows the human a restricted time to veto before automatic execution, OR
 6. ...executes automatically, then necessarily informs the human, OR
 7. ...executes automatically, and then informs the human only if asked.
 8. The computer selects the method, executes the task, and ignores the human.
-

Endsley and Kaber (1999) presented a taxonomy of LOAs developed by allocating to either a human, or a computer, or both, generic control functions including “monitoring,” “generating,” “selecting,” and “implementing” based on the capabilities of each server to perform the functions. These functions were identified for use in developing LOAs by studying an array of dynamic-control tasks including aircraft piloting, tele-operation, complex manufacturing systems control, and process control. They formulated 10 LOAs feasible for use in the context of tele-operations (see Table 2.2). They have been empirically assessed as to their effect on human-machine system performance, and operators’ situation awareness and workload, in a dynamic control task. They have also been studied human performance between normal operations and simulated automation failures and found that human-machine system performance to be enhanced by automation that provided computer aiding in the implementation aspect of the task or allocated the implementation role to the computer. With respect to performance during failure modes, human control was significantly superior when preceded by functioning at LOAs involving the operator in the implementation aspect of the task, as compared to being preceded by higher LOAs. Improved SA and lower levels of overall task demand corresponded with higher LOAs.

Table 2.2 Endsley and Kaber’s (1999) LOAs Taxonomy

Level of Automation	Functions			
	Monitoring	Generating	Selecting	Implementing
Manual Control	Human	Human	Human	Human
Action Support	Human/Computer	Human	Human	Human/Computer
Batch Processing	Human/Computer	Human	Human	Computer
Shared Control	Human/Computer	Human/Computer	Human	Human/Computer
Decision Support	Human/Computer	Human/Computer	Human	Computer
Blended Decision-Making	Human/Computer	Human/Computer	Human/Computer	Computer
Rigid System	Human/Computer	Computer	Human	Computer
Automated Decision Making	Human/Computer	Human/Computer	Computer	Computer
Supervisory Control	Human/Computer	Computer	Computer	Computer
Full Automation	Computer	Computer	Computer	Computer

Since the goal of allocation the functions between human and automation in human-automation system is to maximize the expected value of human-automation system performance of the task. More complex system would be to vary the level of automation according to the momentary situation, known as dynamic allocation (Hancock & Scallen, 1996) or adaptive automation (Parasuraman, Mouloua, & Molloy, 1996). However, these LOAs have not been empirically assessed as to their effect on human operator performance or operators' situation awareness.

2. Trust in automation

Automation use in human-machine system depends on a complex interaction of factors that include workload, cognitive situation awareness, trust in automation, self-confidence, and risk. In particular, Masalonis and Parasuraman (1999) assert that trust is one intervening variable between an automated system and its use. People may or may not use a system because of their trust in it, and their trust in part depends upon their experience using or relying on the system. Parasuraman and Riley (1997) defined three different ways humans improperly use automation to help explain why automation often fails to deliver its promised benefits.

Misuse occurs when operators rely too much on automation, trusting it when it should not be trusted. In these situations, operators might over-trust the automation (Lee & See, 2004), which make them less attentive to judge the information (Skitka, Mosier, & Burdick, 1999, 2000). One important aspect of misuse concerns monitoring failures, in which operators tend to neglect automation breakdowns. Despite taking visual bearings that indicated their true position was perilously off the intended track, the ship's navigation team continued to rely on information provided by the automated navigation equipment that indicated they were on

course (Fahey, 2007). One reason the navigation team failed to identify the problem was that they did not recognize the automation disagreed with what their eyes were seeing; they trusted that the automation was working properly, which was a correct assessment, but they improperly trusted it when their own senses provided contradictory information. Even though automation seems to relieve people of tasks, automation requires more attention to training, interaction design and interface design. Additionally, the likelihood of the monitoring failures is inversely proportional to the likelihood of the failure frequency (Lee, 2008), which provide the evidence to our experimental design.

Disuse happens when operators do not rely enough on automation, ignoring signals and alarms they regard as overly sensitive. Lee (2006) claimed that operators are often slow to accept automation because it threatens their way of life, they have not developed trust in its capability, or the automation lacks the needed functionality. *Abuse* results when designers or managers apply automation incorrectly or without consideration for its effects on human performance. It often occurs because automation designers frequently fail to account for how people adapt to the automation and create automation that has a high degree of authority and autonomy (Starter & Woods, 1994), leading to unanticipated negative consequences.

For example, although one might expect automation to reduce workload and be engaged by operators to mitigate high-workload situations, this is often not the case (Lee, 2008). The reason might be in high-workload situations in which the operator has little confidence in his or her capacity to respond, misuse is more likely than disuse. Many researches focus on the phenomenon to search for the influence factors, and they have proved that trust has emerged as a particularly important factor that influences misuse and disuse and tends to reflect the

capacity of the automation, leading to appropriate use (Lee & Moray, 1992; Muir, 1987; Lee & See, 2004).

A substantial amount of research exists regarding trust in automation, starting with seminal works exploring how human trust automation (Lee & Moray, 1992; Muir, 1994; Muir & Moray, 1996). Muir (1987) explores literature regarding trust between humans and relates it to human-machine interaction. She combined Barber's ideas with those of Rempel, Holmes, and Zanna's (1985) to create a hybrid definition of human-machine trust: trust is the expectation held by a member of a system of the persistence of the natural and moral social orders, and of technically competent performance, and of fiduciary responsibility from another member of the system and is related to objective measures of these qualities.

Lee and Moray (1992, 1994) and Muir and Moray (1996) explored this issue by evaluating human operator's trust in a simulated pump mechanism after it malfunctioned. A total of 60 trials over 3 days were conducted and this included 10 training trials conducted on the first day. They agreed that trust is one important factor that guides operators' interaction with automation and reflects the capabilities of the automation. Subjective rating scales were used to measure operators' trust in and perceptions of the predictability and dependability of the system at the end of each trial. Their work presented the relationship between changes in operators' control strategies and trust in automation and concluded that the allocation function between human manual control and machine automatic control is based on trust in automation and self-confidence in the ability to control the system manually. If operators' confidence in their own ability to control was greater than their trust in human automation, they tended not to use it. When the reverse was true, they tended to use automation.

Since the trend to explore the appropriateness of trust, Lee and Moray (1992) identified performance, process, and purpose as the general three basic dimensions of trust. Performance refers to the current and historical operation of the automation and includes characteristics such as reliability, predictability, and ability, describing what the automation does. Process is the degree to which the automation's algorithms are appropriate for the situation and able to achieve the operator's goals, describing how the automation operates. Purpose refers to the degree to which the automation is being used within the realm of the designer's intent, describing why the automation was developed. Designing interfaces and training to provide operators with information regarding the purpose, process, and performance of automation could enhance the appropriateness of trust.

Muir (1994) made similar distinctions in defining the factors that influence trust in automation. She proposed a model of trust consists of three dimensions of expectations: Persistence, Technical Competence, and Fiduciary Responsibility. Each of these dimensions is crossed with three levels of experience: Predictability, Dependability, and Faith (see Figure 2.2). According to the models above, we can measure an operator's trust through subjective measures due to it is based on operator's judgments and expectations.

Sheridan (2002) noted that trust can be both an effect and a cause. In human-automation terms, repeated use of a system may have the effect of increasing the operator's trust. Additionally, trust may cause further reliance by the human on the automation. Thus, development of a measure of trust is important in order to design automated systems that encourage appropriate use by human.

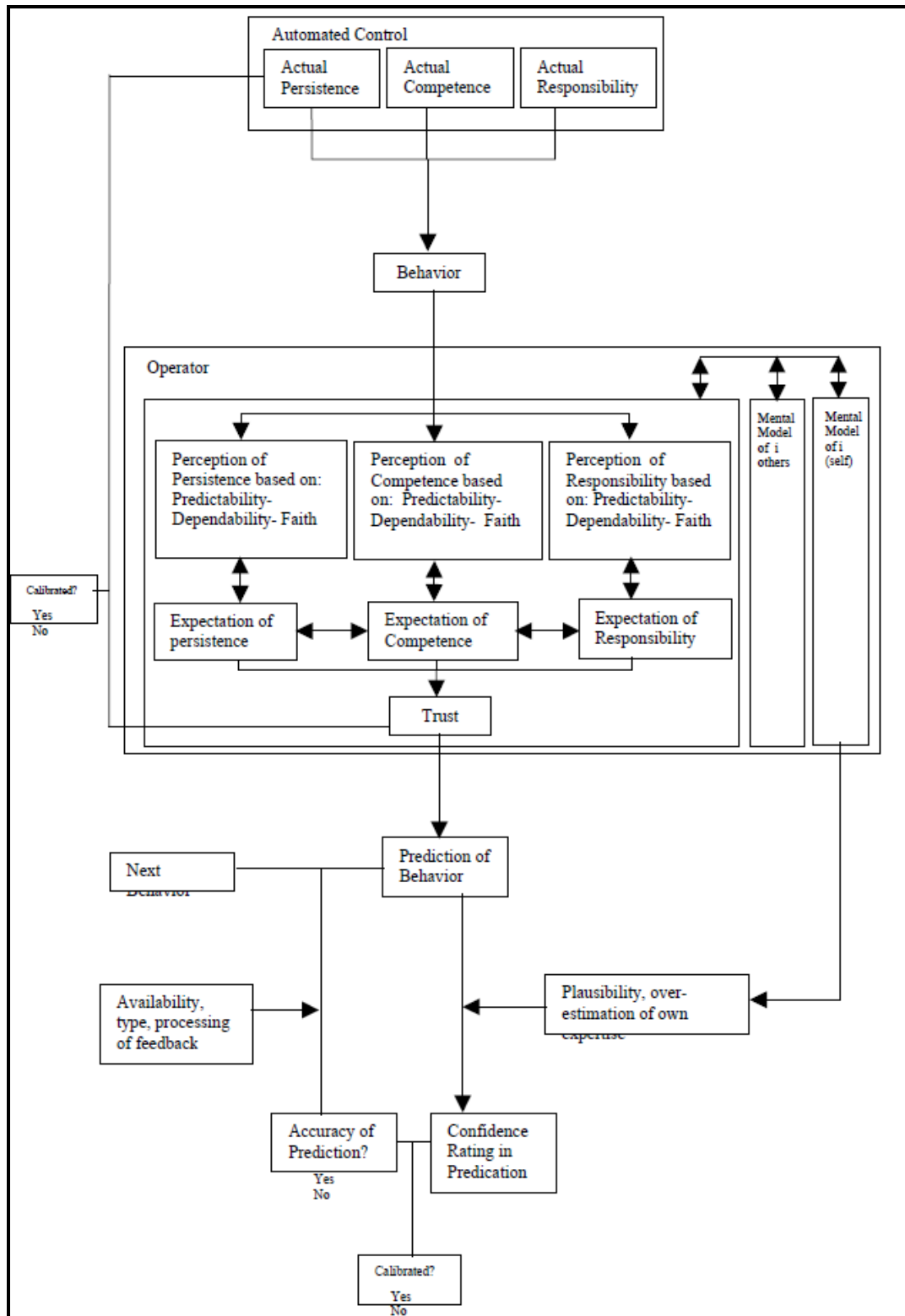


Figure 2.2 Man-Machine Model of Trust taken from Muir (1994)

Lee and See (2004) proposed a conceptual model of the dynamic process to show the trust and its effect on reliance are part of a closed-loop process in which the dynamic interaction with the automation influences trust and trust influences the interaction with the automation, as well as the interaction among appropriateness of trust, the influence of context, the goal-related characteristics of the agent, and the cognitive processes that govern the development and erosion of trust (see Figure 2.3).

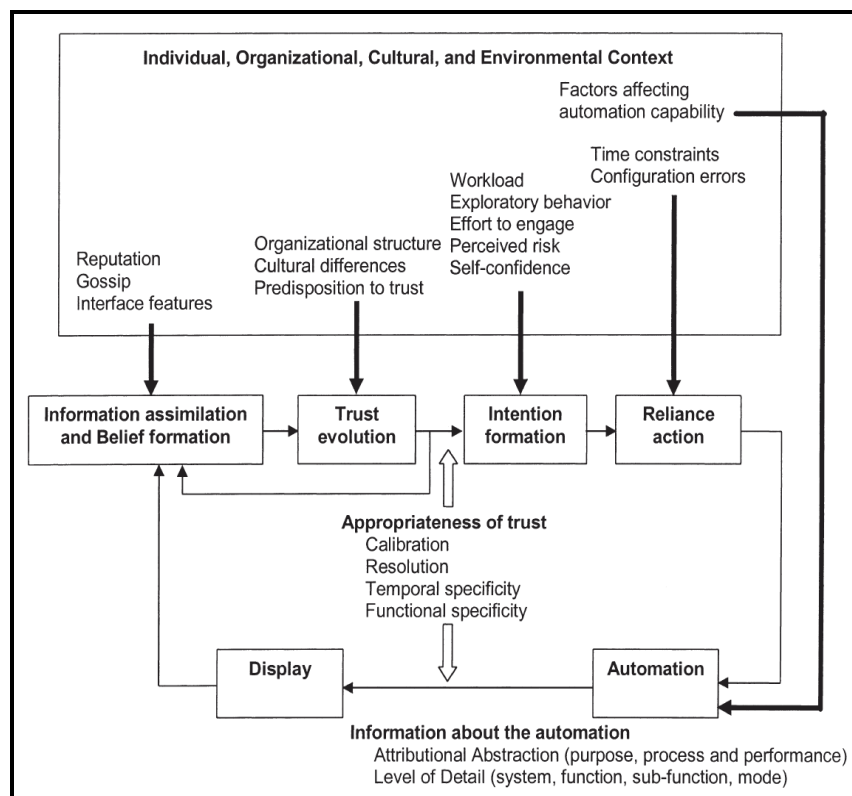


Figure 2.3 A conceptual model of the dynamic process (Lee & See 2004)

Many researchers have used questionnaires to measure subjective feelings of trust (Rempel et al., 1985; Singh et al., 1993; Lee and Moray, 1994; Muir and Moray, 1996), however, these questionnaires have been based on theoretical rather than empirical notions of trust dimensions. Using a subjective scale (Jian, Bisantz, & Drury, 2000) (see Figure 2.4) to measure trust in automated decision aids, Bisantz and Seong (2001) investigated the effect of

source of failure causes on operator trust, similar to our study. The experiment involved a target identification task that required participants to identify targets as enemy or friendly with the assistance of an information automation aid or a decision automation aid. Failure cause was treated as a between-subject, fixed factor, and session was treated as a within-subject, fixed factor. Participants were separated into three groups by what they knew regarding potential automation failures and rated their trust in the automated aid using a seven-point scale anchored at “Not at all” and “Extremely” for each of the following statements.

Figure 2.4 Subjective rating of trust in automation (Jian, Bisantz, & Drury, 2000)

1. The system is deceptive
2. The system behaves in an underhanded manner
3. I am suspicious of the system’s intent, action, or output
4. I am wary of the system
5. The system’s action will have a harmful or injurious outcome
6. I am confident in the system
7. The system provides security
8. The system has integrity
9. The system is dependable
10. The system is reliable
11. I can trust the system
12. I am familiar with the system

The first five questions are negatively framed, while the last seven are positively framed. This distinction allows for testing of different aspects of trust. Responses to the subjective trust questionnaire indicated operator trust declined less in the group who believed the failure source was external to the automated aid, which validated the use of a trust questionnaire that was sensitive to different aspects of trust and to different automation failure conditions.

3. Reliability of Automation

Automation functions could produce erroneous or anomalous outputs due to failures at multiple system levels. For instance, failures in the environment or controlled system of interest, the automation or decision support algorithms, or in the human–computer interface, could contribute to unexpected behavior of human-automation system.

Automation functions imperfectly and failures can be seen to occur intermittently as well as permanently. These failures have an effect on the extent to which users rely on the automated systems and how well they perform manually without its aid. The literature on automation reliability and how it affects operator trust, reliance and performance clearly suggests that the perceived reliability of an automation system relative to manual performance is a critical determinant of the extent to which the aid is relied upon and this is, in turn, reflected in performance; there tends to be greater reliance on automation when it is deemed to be more reliable than manual performance. A critical observation that follows is that users would be able to calibrate their reliance patterns more appropriately if they are able to assess the reliability of the aid more effectively. Lee and Moray (1992)'s research examined the effect of 'transient' and chronic' system errors on development of trust. The results showed that both system performance and the occurrence of errors affect trust. When the system contained 'transient' and chronic' errors, operators' trust and performance dropped and then recovered as they learned to accommodate the errors. In addition, the 'chronic' error led to an increased use of the automatic control. Thus, the reliability of automation system affects the operators' trust in automation and then affects the operators' performance. Within the reliance-compliance framework, loss of trust and dependence when automation errors occur is manifest.

As a result of Muir's experiment (Muir, 1994), workers monitoring automation became complacent when the automation was perceived to perform correctly; and workers spent more time monitoring systems considered to be error prone. She found evidence to suggest that following a perceived error, a person's trust will degrade but will gradually recover over time. Low reliance requires the operator to more closely monitor the raw data, at the expense of concurrent tasks. Her findings have been supported in similar studies (Lee and Moray, 1992). The literature has shown that automation reliability and more importantly, users' perceptions of its reliability are factors critical to make the decision of implementing automation which users trust and depend on it. Perception is defined as the process of attaining awareness or understanding of sensory information. Thus, we selected visual degradation errors which could influence the perception reliability of automation in order to explore how it impedes the operators' trust in automation and performance on visual searching tasks.

4. Trust and situation awareness

Many of studies examined how mental models or shared cognition affects team performance related to complex systems. Rouse et al. (1992) provide a description of mental models, outlining three main functions as they relate to human-system relations. The descriptive function pertains to a person's knowledge of the system's purpose and physical description. The explaining function involves a person's knowledge of the system's operation and its current state. The prediction function relates to a person's ability to form expectations about the system's future state and operations. These components may help support an explanation for appropriate human trust in automation when an individual's mental model is properly developed.

Everyone seems to think that such mental models exist in the human mind, but no one seems to know how to represent them or how to use them. Endsley's research (2000) on situation awareness (SA) contends that a mental model is general while SA is specific to the circumstances one encounters on a minute-to-minute basis. The definition of SA she pointed out is "the perception of the 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. 529). She presented that a person's mental model consists of relatively static components that develop with time and experience, while SA is more dynamic and provides input to the mental model, developing it over time (see Figure 2.5).

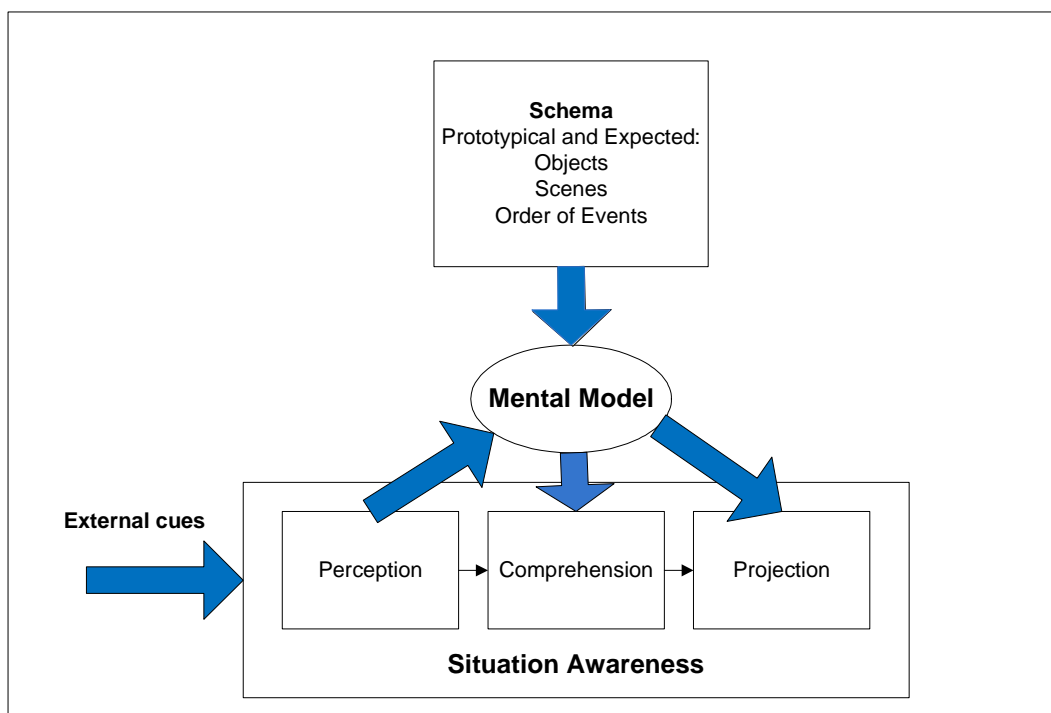


Figure 2.5 Endsley's model of SA and Mental Model (2000)

Adjusting one's mental model with experience may lead to trust that is more accurate. With experience and time, a person adjusts his or her situation awareness based on accumulated information and interactions. Since the operators in the current study were

limited in the time they had to interact with the automation, they could not feasibly develop situation awareness without someone pointing out critical external cues to them. As a result, we made the external cues (walls, furniture, and field layout) in an attempt to make up for the limited exposure participants had with the automation.

C. Limitations and Gaps

Since Lee and Moray introduced faults into pump performance (Lee & Moray, 1992) or faults into either automatic or manual controllers (Lee & Moray, 1994) relating variations in trust in automation and self-confidence to human system performance, there has been a large body of research examining the interrelationships between trust and factors such as automation reliability, error type, task difficulty, and other factors (Bisantz & Seong, 2001; Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001; Wiegmann, Rich, & Zhang, 2001). Muir and Moray (1996) and Lee and Moray (1994) studied issues of human trust in simulated, semi-automated pasteurization plants. In particular, Muir and Moray (1996) altered the quality of the pump systems by introducing either random or constant errors in its ability to maintain a set-point, introduced errors into the pump's display of its pump rate, and the performance of the automated controller in setting and maintaining appropriate settings for the pump. Bisantz and Seong (2001) used a low fidelity simulation of an anti-air warfare task to examine the effect of failure causes. This large body of research has clearly established the importance of trust in the human use of automation. Trust plays a critical role in people's ability to accommodate the cognitive complexity and uncertainty.

However, these researches had not indicated the effect of changing perceptions of reliability on trust in automation. Little research has addressed the challenges of promoting

appropriate trust in the face of a dynamic context that influences its capability. In addition, few studies conducted the experiment using the real human-telerobot system.

Several LOA taxonomies have been proposed in the literature (Endsley, 1987; Endsley & Kaber, 1999; Sheridan, 1992, 2002). Sheridan (2002) developed a LOA taxonomy incorporates issues of what the human should be told by the system, as well as relative sharing of functions determining options, selecting options and implementing. Endsley and Kaber (1999) formulated 10 LOAs feasible for use in the context of tele-operations. Parasuraman et al. (2000) indicated that there was a need for determining experimentally what should and should not be automated, based on cognitive engineering data and other considerations.

However, very little experimental work has been conducted to examine the benefit of applying different LOAs and appropriateness of trust in complex tasks for enhancing specific task performance in an unknown environment, or to examine the effects of LOAs on operators' situation awareness and mental workload during the specific task, such as firefighting in a welding department.

Related to automation error and trust, there is a need to know how varying automation reliability influences operator trust in automation, and if there is a difference in trust of an unreliable low level automation versus high level automation in realistic tasks. There is also a need to explain any differences, or to identify underlying factors. When levels of automation reliability vary, this may pose a different mental demand on human operators. If operators perceive different reliabilities of automation system, they may allocate more attentional resources from an already limited source in order to monitor automation states. Therefore, there may be a negative influence on operator's situation awareness. Furthermore, it may

influence operator situation awareness. Under varying reliability automation, lower reliability conditions require more mental attention, reducing operator perception, comprehension and projection of system states and environment knowledge. Few studies have investigated the impact of automation reliability on situation awareness.

The purpose of the present research is to further examine and compare two different LOAs, specifically in a real human-robot system generalizing results to a real-world application. Further, it was intended to demonstrate the usefulness of LOAs in the context of a tele-operation application. This was accomplished by assessing the impact of LLA and HLA on tele-robot performance under both normal operating conditions and error modes (intermittent error and permanent error), and its effect on operators' situation awareness and subjective workload.

CHAPTER 3 MATERIALS AND METHODS

A. Overview

A firefighter scenario was developed for use with human-telerobot system in the entire experiment. This was done in order to compare trust in automation between low and high level automation for real world use. The experiment consisted of a series of target detection tasks. The targets in the current study were colorful bottles used to present the fire scenes.

This experiment collected quantitative measures of subject performance over the course of each testing session, as well as measurable attitudes and feelings through a pre and post questionnaire. In each testing session, the experiment maintained the same between LLA and HLA groups. Visual system degradation (intermittent and permanent error) was manipulated in each group during Testing II and III sessions, while normal system operation without failure in Testing I session. As a simulated degradation of the visual system, intermittent error presented the degradation occurred every 5 minutes and lasted 1 minutes and permanent error presented the degradation occurred through the entire testing session. These variables were configuration as seen in Table 3.1.

Table 3.1 Experiment Design

Level of Automation	Session	Visual System Degradation
LLA (N=12)	I	None
	II	Intermittent Error
	III	Permanent Error
HLA (N=12)	I	None
	II	Intermittent Error
	III	Permanent Error

Errors causing the visual system degradation was treated as a within- subject, and level of automation was treated as a between-subject. To minimize the impact of short memory of human on the experimental data, the structure of rooms (door position) and the arrangement of rooms (furniture position) changed among three testing sessions.

B. Experimental Task

The experiment conducted in a welding room separated into four rooms which simulated a hazard and dangerous place which firefighter could not get into easily. The operator used the tele-robot to move or put off the threat targets in the field. Subjects were tasked to detect and identify the targets (colorful bottles) by controlling and monitoring the tele-robot via system interface during the searching task.

In LLA group, subject using joystick controlled the tele-robot manually to search the targets in the welding room. Subjects were told the objective of their mission was to find three threat targets in four rooms and to try their best to reduce the collisions during the task. If the object was identified and confirmed by the subject as a target, he/she informed the researcher using a trigger button. Then, subject approached to the target until the distance was less than 50 cm.

In HLA group, subject monitored the tele-robot automatically to search the targets in the welding room. Subjects were told that the while the computer system would automatically search and determine the identity of all the targets, it was possible for sending a message to the computer system to be manipulated by the subject. If the object was identified by the automatic control of tele-robot as a target, the computer system informed the subject to judge whether it was a target. If the subject confirmed the target, tele-robot approached to the

target by itself until the distance was less than 50 cm. Then, tele-robot would enter other rooms to search for the targets until find them all.

C. Variables

1. Independent Variables

This study investigated and examined the effect of two LOAs as an independent variable in performance during both normal operation of tele-robot and simulated failures. As well, participants served as observations of the LOA effect on situation awareness and mental workload.

(1) Levels of Automation

The human-tele-robot system was programmed to allow for the use of two LOAs presented in the taxonomy of Endsley and Kaber's (1999) in Table 3.2 or in the taxonomy of Sheridan (2002) in Figure 3.1. These two levels represented typical low and high level of automation with computer assistance allocated to the human-tele-robot system.

Table 3.2 Selected levels of automation in Endsley and Kaber's taxonomy of LOAs (1999)

Level	Description	Roles			
		Monitoring	Generating	Selecting	Implementing
Low (LLA)	Action Support	Human/Computer	Human	Human	Human
High (HLA)	Shared Control	Human/Computer	Human/Computer	Human	Human/Computer

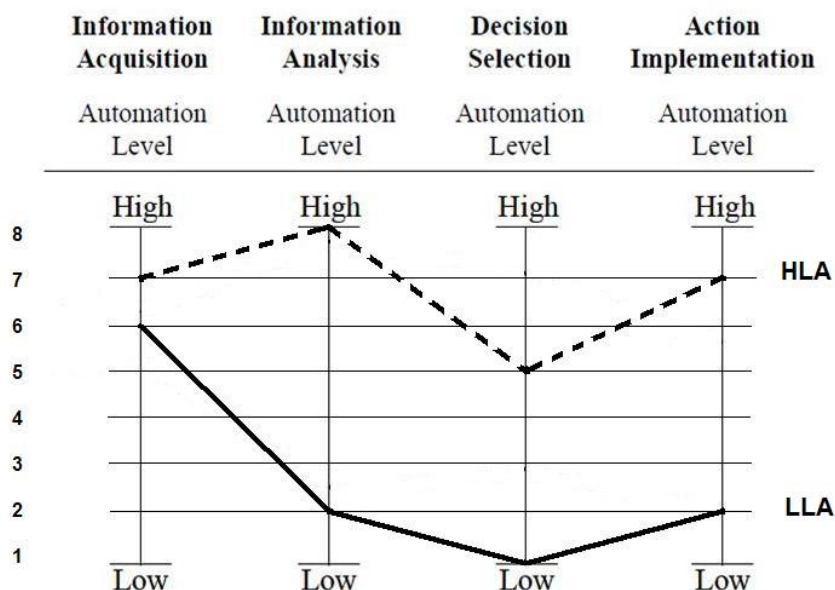


Figure 3.1 Selected levels of automation in Sheridan's taxonomy of LOAs (2002)

(2) The Reliability of Visual Perception (visual system degradation)

As the tele-operation involved mechanism and signals, the issue of reliability and safety became part of the success of its mission and its design. The subject of robot reliability is very complex and there are numerous interlocking variables in evaluating and accomplishing various reliability levels. To many possible failure modes of a human-teleoperator system fall into four principal domains which affect system reliability and its safe operation: system integrity, data integrity, control design, and task requirement (Sturges, 1990). Among all, the data integrity of a tele-operation control system may be lost due to device degradation, time delays in a perfect system, or shifts in the workspace model that go undetected. For our experiment, we focused on perception reliability of human-robot system which includes interface reliability (data integrity domain). Specially, we varied the reliability of visual perception by degrading the quality of visual system which monitor remote conditions and

display them to the operator via camera on the tele-robot as we discussed earlier. To understand the role of the failure in the human-robot system, we used block diagram and success tree methods to analyze the reliability of the entire system.

The probability of nonoccurrence (reliability) of the undesirable robot's output can be estimated from the series block diagram shown in Figure 3.2.

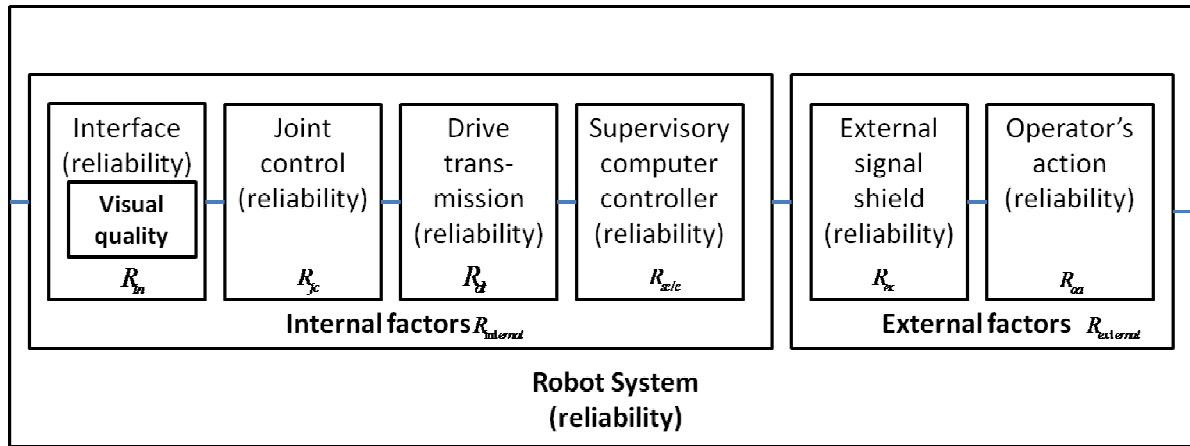


Figure 3.2 Block diagram representing the internal and external factors of robot system

The probability of nonoccurrence (reliability) of the undesirable robot movement is

$$R_{system} = R_{internal} g R_{external} \quad (3.1)$$

Where R_{system} is the probability of nonoccurrence (reliability) of the undesirable robot output, $R_{internal}$ is the reliability of the internal subsystem A, and $R_{external}$ is the reliability of the external subsystem B.

Thus, form (3.1), the probability of occurrence, F_{system} of the undesirable robot output is

$$F_{system} = 1 - R_{system} = 1 - R_{internal} R_{external} \quad (3.2)$$

For internal factors, the reliability of internal subsystem is

$$R_{internal} = R_{in} R_{jc} R_{dt} R_{sc/c} \quad (3.3)$$

Where R_{in} is the interface reliability, R_{jc} is the joint control reliability, R_{dt} is the drive transmission reliability, and $R_{sc/c}$ is the supervisory computer/controller reliability.

Similarly, for external factors the reliability of external subsystem is

$$R_{external} = R_{ex} g R_{oa} \quad (3.4)$$

Where R_{ex} is the reliability external signal shielding, and R_{oa} is the reliability of operator's action with respect to causing robot movement.

Although we simulated the visual quality degradation by programming, we investigated the reliability of visual quality during monitoring task by discussing the influence elements of robot system. Table 3.3 presents failure rates for elements (Dhillon, 1991) which might impact on the visual quality. These influence elements imply that once any one or more elements fail the visual quality information will be affected.

Table 3.3 Failure rates for selected influence elements of robot system on visual quality

No.	Item description	Failure rate (failures per year)
1	Low power transformers (for control and electronic equipment)	0.20
2	Connectors, pin (use: military; use environment; ground, mobile)	0.9636e-4
3	Fiber optic connector (single fiber)	0.876e-3
4	Storage battery (nickel cadmium)	0.0289
5	MOS dynamic RAMS in hermetic packages (use environment: ground, mobile)	0.657e-3

Success tree method is the dual of the fault tree method. In this case, the reliability analyst is concerned with investigating success events instead of fault events, more specifically, nonoccurrence events instead of occurrence events. The incomplete event is represented by a diamond, and may simply be described as a success event whose causes have not been fully developed due either to lack of interest or to lack of information. The AND gate only provides an output if all of its inputs do not occur or fail. On the other hand, the OR gate provides an output, if at least one of its inputs does not occur or fails. The success tree for human-teleoperator system is shown in Figure 3.3.

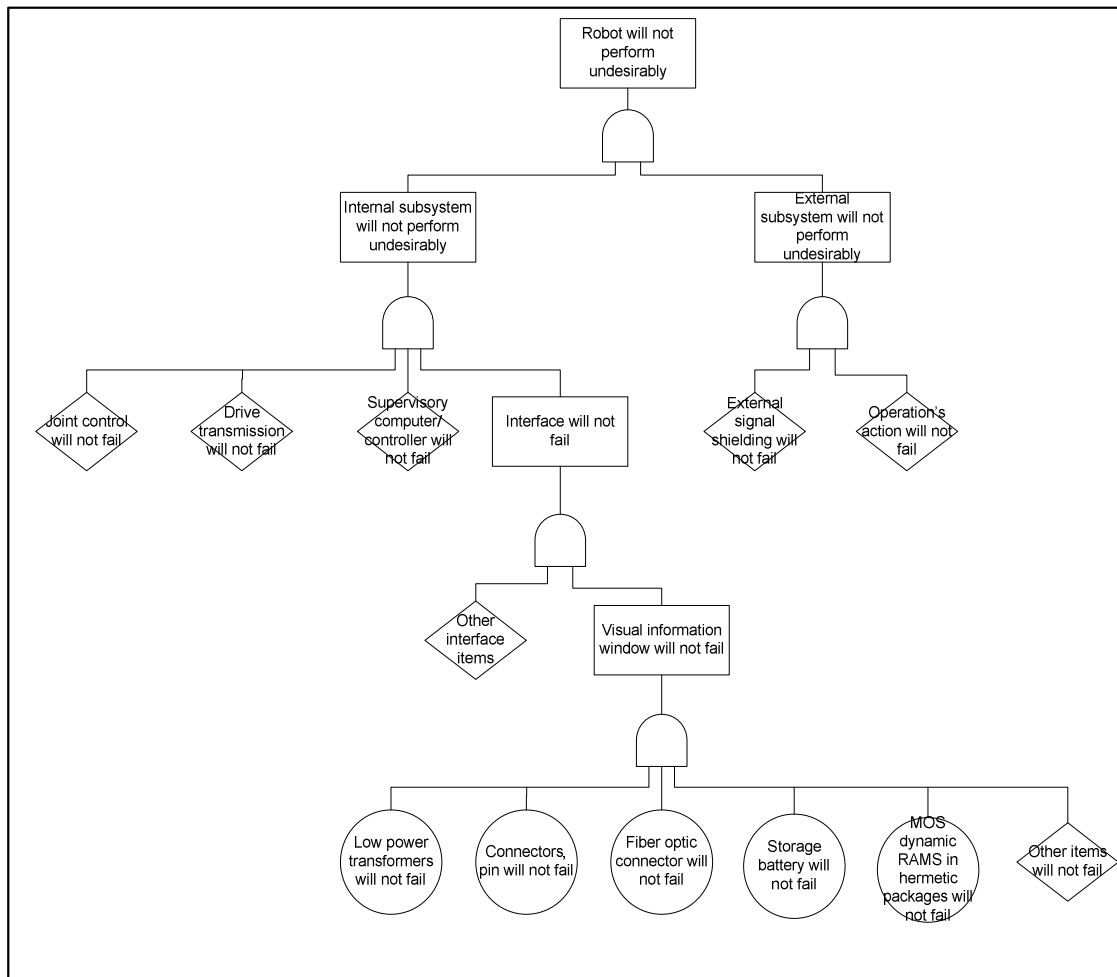


Figure 3.3 Success tree for human-teleoperator system

The performance of human in the loop is one of the greatest sources of mission variation, yet he or she remains indispensable in the human-teleoperator system. In current study, we examined how visual system degradation will impact the human performance, mental workload, situation awareness and trust in automation when using different allocation strategy between human and automation (level of automation). There were two types of simulated visual information degradation error: intermittent error presented the degradation occurred every 5 minutes and lasted 1 minutes and permanent error presented the degradation occurred through the entire 30-minute testing session.

2. Dependent Variables

The dependent variables recorded during the experiment included completion time, the number of wrong locations, hit rate (the number of collisions between robot and obstacles during 30 minutes' task), and the number of target found. Observations were made for normal condition control (Testing I session) and two types of visual system failures control simulated during last two sessions (Testing II and III sessions).

Operators' situation awareness was measured during the study using Situation Awareness Rating Technique (SART; Taylor, 1989) questionnaire regarding the three levels of situation awareness proposed by Endsley (1988). SART is a post-experiment questionnaire and requires the operator to rate 10 dimensions are shown in Appendix A. The questionnaire was posed to participants to rate each dimension on the scale of 1 to 7 after each testing session. Situation Awareness was quantified based on the total scores obtained for each of the three dimensions (see Table 3.4). The formula to calculate the SA metric is "Understanding - (Attention Demand - Attention Supply)". These data served as composites of operator

perception and comprehension of system information, as well as future system state predictions.

Table 3.4 SART Dimensions

Domains	Construct	Definition
Attentional Demand	<i>Instability of situation</i>	Likelihood of situation to change suddenly
	<i>Variability of situation</i>	Number of variables that require attention
	<i>Complexity of situation</i>	Degree of complication of situation
Attentional Supply	<i>Arousal</i>	Degree that one is ready for activity
	<i>Spare mental capacity</i>	Amount of mental ability available for new variables
	<i>Concentration</i>	Degree that one's thoughts are brought to bear on the situation
	<i>Division of attention</i>	Amount of division of attention in the situation
Understanding	<i>Information quantity</i>	Amount of knowledge received and understood
	<i>Information quality</i>	Degree of goodness of value of knowledge communicated
	<i>Familiarity</i>	Degree of acquaintance with situation experience

The NASA-TLX (Hart & Staveland, 1988) was used to subjectively assess the overall workload experienced by operators (see Appendix B). Using the NASA-TLX Windows Version program, participants were required to complete rankings of six subscales: mental demand, physical demand, temporal demand, effort, performance, and frustration after each testing session. Then, participants are repeatedly asked to choose which of a pair of subscales contributes more to their overall workload, until all possible pairs of subscales have been compared. In order to calculate the workload metric, the ratings from the six subscales are combined into a single weighted measure of workload using the number of times a particular subscale was preferred as its weight.

Jian et al. (2000) used a three-phase experiment, in which words related to trust were collected, rated, and clustered, to empirically develop a twelve item trust questionnaire (see Appendix C). This questionnaire incorporates a seven point rating scale in the range from “not at all” to “extremely”. Subjects were requested to rate the degree of agreement or

disagreement of with these twelve trust-related statements. This measure represents the first attempt at empirically generating a scale to measure trust in automation. In the current study, data based on this trust questionnaire were collected to investigate the effects of visual system degradation and levels of automation on the operators' trust and to explore the relationship between trust in automation and allocation of automation.

D. Participants

The Iowa State University Institutional Review Board reviewed and approved the design of this study, satisfying the American Psychological Association criteria for research involving human subjects. We solicited participant through emails and personal contact. All participants indicated informed consent by signing a form notifying them of their rights as participants in experiment.

A total of 24 (21 males and 3 females) students, aged 20 to 30 years, from Iowa State University comprised the participants in this study. Participants either had experience with computer-based games or drive a vehicle often.

E. Materials

1. IowaBot User Interface

The user interface of IowaBot as a control panel provides the information of drive control variables (translate and rotate) though the drive panel, the information of external environment though the visual system and other information of system performance (see Figure 3.4). Simulated visual system degradation by C# programming showed on the user interface to investigate the effect on human-telerobot system.

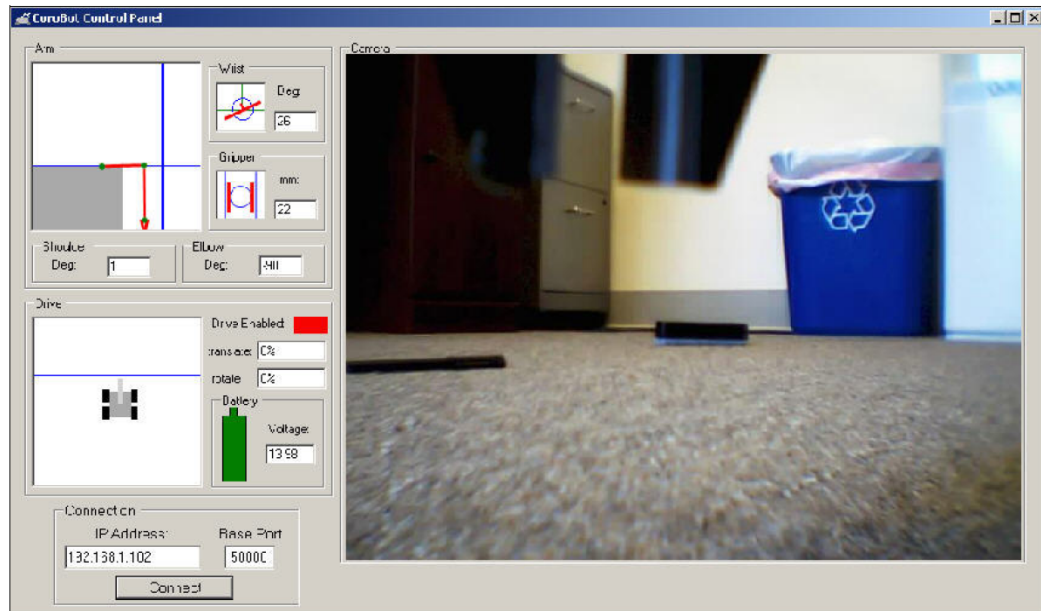


Figure 3.4 IowaBot User Interface

2. Equipment

The workspace (see Figure 3.5) consisted of a chair, a desk surface, quick reference sheets including instructions of operation and a room layout for certain session and a 15-in PC-type laptop computer system connected to a 17-in PC monitor operating under 1600 by 1280 resolution loaded with a Windows XP operating system. The system was deployed the IowaBot user interface program (partially developed by CoroWare, inc.), integrated with a mouse, standard keyboard and joystick controller and used in the study to electronically present self-made VBA program for evaluations.



Figure 3.5 Experimental Workspace

IowaBot (see Figure 3.6) as a remote control robot communicated with the base station (user computer) via wireless radio. It is created as a rugged indoor/outdoor robot that can withstand environmental elements such as dirt, dust, leaf debris, sand, gravel and shallow puddles. The camera is floor mounted for best visibility. The infrared range sensor detects the distance from the front and back of the IowaBot. This information is displayed with a blue line on the CoroBot Control Panel.

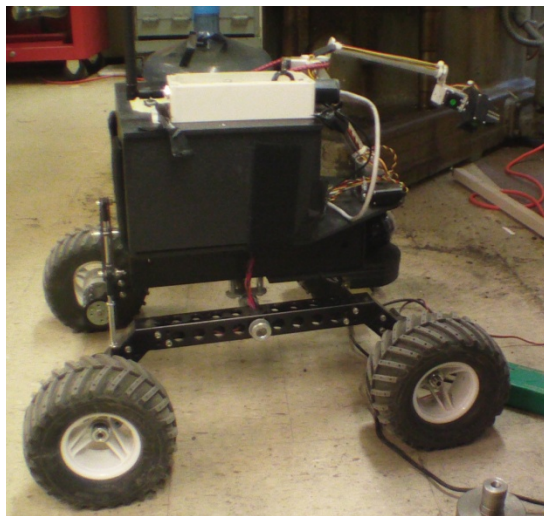


Figure 3.6 IowaBot Configuration

The welding room was furnished with welding machines divided into four rooms by compressed boxes (see Figure 3.7). The compressed boxes were high enough to isolate each sub-room. In addition, the floor of the welding room was dusty because of the welding dust in order to simulate the real environment of fire scenes.



Figure 3.7 Welding Room Configuration

F. Procedure

Participants signed up for a one-hour block of time each day in two successive days. Prior to training, first 10 to 15 minutes participants were asked to review and signed a consent form. After participants made an agreement to participate the experiment, they were asked to fill out a pre-experimental questionnaire which involved the questions about their automated control experience (see Appendix D). They were randomly assigned to one of two experimental groups: one was Low level of Automation (LLA) group; the other was High Level of Automation (HLA) group.

Before operating the system, the operators received an extensive written description of their objectives in controlling the tele-robot, the mission of the tasks, the possibility of faults, and the brief instruction of operation. Then the participants of both LLA group and HLA group were given basic training in how to perform the experimental task of navigating a

robot in an unknown area and exploring the targets during the searching task. Additionally, the participants of HLA were told that the initial mode of robot is high level of automation control but they could shift the control mode between LLA and HLA anytime they wanted. This period of time for both groups involved familiarizing the participants with the concept of Tele-robotics and detailing how to use the joystick controller and control interface system to control the robot. During training no measurements were taken and experimenters acted as a trainer to ensure that the participant properly understood the system they were working with. It lasted for 10 to 15 minutes.

Each operator completed three 30 minutes sessions of testing. On the first day, they completed Testing I Session (see Table 3.5). On the second day, they completed Testing II & III Sessions (see Table 3.6). In Testing I Session, participants conducted a Tele-robotic searching task for exploring targets in an unknown indoor field according to the field structure layout. For the exploring task, they marked the location of the targets in rooms on the layout after they found the targets. Participants will perform the same task they had the day prior. The differences were that the Tele-robotic system error randomly occurred when they performed the tasks and the field structure layout changed each session as well as the location of the targets. In Testing II and III sessions, the intermittent error and the permanent error of video degradation randomly occurred. The purpose of this Testing II & III session was to determine whether the intermittent error or permanent error had an effect on trust, human performance, mental workload, situation awareness and system efficiency in low and high level of automation groups. Also the comparison of results between intermittent error and permanent error was what we concerned.

After each session, the computer displayed a series of questions to establish the operators' subjective feelings about the task and the system. Operators received detailed instruction and explanation to ensure that they had a clear conception of the meanings of their subjective ratings and then evaluated their mental workload, situation awareness and trust in automation during the task. After completed the entire experiment, participants were debriefed and thanked for their time.

Table 3.5 First Day Experiment

First Day Experiment
Pre-questionnaire
10 minutes' training
Testing I:30 minutes' testing session
NASA TLX, Trust in Automation and SART Rating

Table 3.6 Second Day Experiment

Second Day Experiment
Testing II:30 minutes' testing session
NASA TLX, Trust in Automation and SART Rating
Testing III:30 minutes' testing session
NASA TLX, Trust in Automation and SART Rating
Debriefing

CHAPTER 4 RESULTS AND DISCUSSION

A. Overview

This chapter presents an analysis of the data collected during the two-day experiment as described in Chapter 3. The results of this information in relation to the research hypotheses will be compared with the results in the discussion section. The following sections present the results of statistical analysis of the hypotheses under investigation in this experiment.

The analyses presented here were carried out to assess human performance (including task performance), situation awareness, mental workload and trust in automation. Task completion time, the number of wrong location, the number of target found, hit rate were used as indicators of human performance. Trust score was assessed, as well as situation awareness and mental workload.

The experimental data were divided into two subsets for analysis: under LLA control and under HLA control (see Table 4.1). The data sets of each LOA group (Low/High) were analyzed through one-way analyses of variance with two types of errors (II: with Intermittent Errors, and III: with Permanent Error) as a within-subject variable. A check for normal distributed was performed on the dependent variables to ensure that all assumptions of the ANOVA were upheld. This allowed for an examination of the effects of the various errors on trust in automation, human performance, situation awareness and mental workload under LOAs. All the data sets of both groups were combined together and analyzed through ANOVAs with LOA as a between-subject variable in each testing session. The comparison between low and high level automation control reflecting on the dependent variables was showed which LOA would produce superior performance.

Table 4.1 Means of Responses in LLA and HLA groups

		LLA			HLA		
		No degradation	Intermittent Error	Permanent Error	No degradation	Intermittent Error	Permanent Error
Workload	NASA TLX	53.81 (17.56)	45.72 (17.03)	51.75 (19.28)	46.33 (15.46)	47.33 (18.34)	44.08 (19.86)
Situation Awareness	SART	5.75 (2.62)	6.50 (2.03)	6.10 (2.69)	6.02 (2.23)	5.40 (1.91)	6.19 (1.85)
	Demand L1 of SA	3.64 (1.26)	2.94 (1.59)	3.61 (1.70)	3.33 (0.94)	3.44 (0.83)	2.92 (1.01)
	Supply L2 of SA	5.21 (0.80)	4.52 (0.81)	4.48 (0.82)	4.77 (0.91)	4.06 (0.80)	3.92 (1.06)
	Understanding L3 of SA	4.39 (1.61)	5.11 (1.45)	5.39 (1.54)	4.58 (1.29)	4.78 (1.40)	5.08 (1.30)
Performance	Hit rate	6.42 (3.26)	3.67 (2.35)	5.08 (1.78)	2.17 (3.76)	3.25 (4.33)	1.42 (2.68)
	Completion time /min	28.86 (8.84)	18.05 (8.96)	17.93 (6.15)	26.44 (5.34)	24.25 (6.42)	22.79 (6.27)
	# target found	2.75 (0.45)	2.92 (0.29)	3.00 (0.00)	2.75 (0.45)	2.75 (0.45)	2.75 (0.62)
	# wrong location	0.67 (0.98)	0.33 (0.49)	0.67 (0.89)	0.67 (0.98)	0.17 (0.39)	0.42 (0.67)
Trust in Automation	deceptive	2.08 (1.08)	2.08 (1.31)	2.08 (1.56)	2.42 (1.68)	2.17 (0.94)	2.17 (1.27)
	underhanded	2.92 (1.31)	3.17 (1.95)	2.58 (1.93)	2.08 (0.79)	2.42 (1.44)	2.50 (1.57)
	suspicious	1.75 (1.14)	1.50 (0.67)	2.00 (1.65)	1.92 (1.83)	1.58 (0.51)	2.25 (1.42)
	wary	2.00 (1.54)	2.67 (1.72)	2.58 (1.83)	2.33 (1.56)	2.58 (1.68)	2.83 (1.90)
	harmful	1.83 (1.27)	1.42 (0.51)	1.42 (0.51)	1.08 (0.29)	1.50 (0.80)	1.83 (1.11)
	confident	5.00 (1.48)	6.00 (0.60)	5.83 (0.94)	5.17 (1.64)	5.00 (1.71)	4.92 (1.93)
	security	5.00 (1.54)	5.50 (1.38)	5.42 (1.31)	4.92 (1.93)	4.58 (1.73)	4.83 (2.04)
	integrity	4.83 (1.11)	5.25 (1.29)	5.42 (1.31)	5.17 (1.19)	4.75 (1.14)	4.75 (2.01)
	dependable	4.75 (1.36)	5.25 (1.29)	5.33 (1.44)	5.00 (1.76)	4.75 (1.36)	5.00 (1.71)
	reliable	4.67 (1.50)	5.33 (0.78)	5.58 (0.90)	5.00 (1.54)	5.00 (1.48)	5.00 (1.68)
	trust	5.25 (1.48)	5.58 (0.79)	5.67 (0.98)	5.08 (1.62)	5.25 (1.54)	4.92 (1.62)
	familiar	4.08 (1.51)	5.58 (1.68)	5.75 (1.54)	4.08 (1.98)	5.58 (1.24)	6.00 (1.28)

B. Performance

1. Performance and LOAs

At the Chapter 1, it was hypothesized that HLA would produce superior performance than the LLA control due to the allocation strategies. An analysis of variance was conducted on task completion time, the number of wrong location, the number of target found and hit rate regarding level of automation in types of error. Results in testing session with intermittent errors revealed that there was no significant difference in human performance between low and high level of automation ($p > 0.05$).

In testing session with permanent error also there was no significant difference in task completion time, the number of wrong location and the number of target found between two levels of automation ($p > 0.05$). LOA did not significantly affect completion time, the number of target found and the number of wrong locations during various errors. However, there was a significant effect of LOA in permanent session on hit rate which is the number of collisions between robot and obstacles during 30 minutes' task, $F_{0.05,1,22} = 15.59, p = 0.0007^*$. It demonstrated that HLA control group produced significantly lower number of collision between robot and obstacles during 30-minute task than LLA control group did (see Figure 4.1).

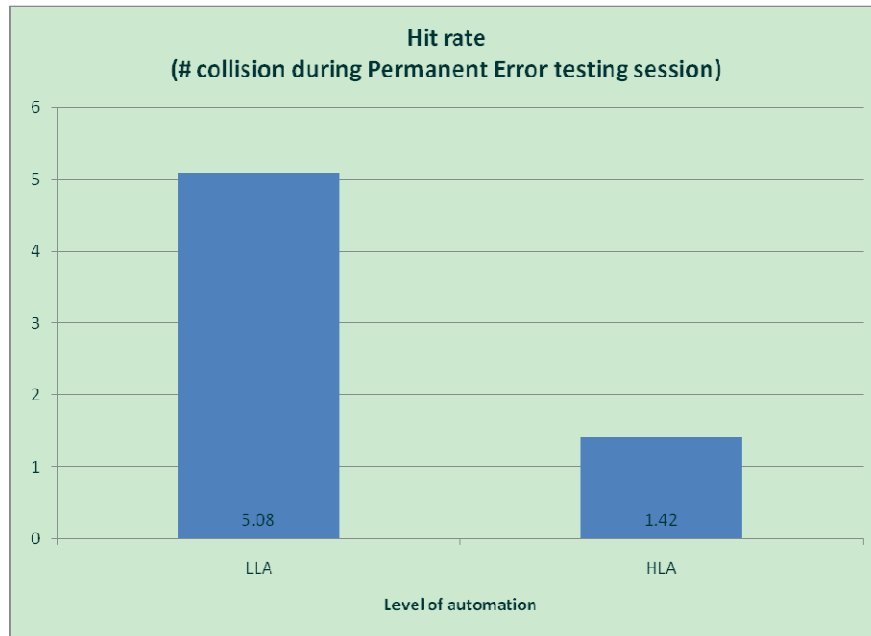


Figure 4.1 Mean Hit Rate of LLA and HLA in permanent error session

2. Performance and Visual System Degradations

Under LLA control: Analyses of variance were conducted on data recorded during the two types of visual degradation errors sessions in which participants were required to perform direct tele-operation as LLA. Results revealed a significant effect of the type of error on hit rate, $F_{0.05,1,11} = 7.59, p = 0.0187^*$. Figure 4.2 shows a increasing of the response as LLA control under permanent error compared to the control under intermittent errors. The effect of type of error on this response demonstrated that permanent error had a worse effect on human performance than intermittent errors in low level of automation. There was no significant effect of the type of error on other human performance ($p > 0.05$).

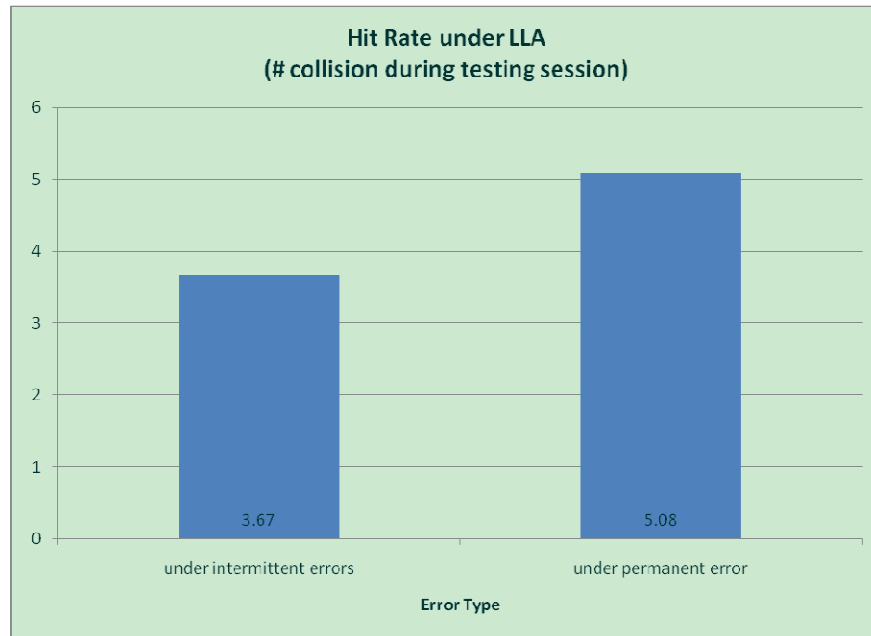


Figure 4.2 Mean Hit Rate under two types of error in LLA

Under HLA control: Due to automatic execution the tasks, there was no significant difference human performance between two types of visual system degradation error according to the ANOVAs analyses ($p > 0.05$).

C. Trust

1. Trust and LOAs

According to the analyses of variance, there was no significant difference between two levels of automation in each trust element rating in trust in automation questionnaire in both intermittent errors and permanent error sessions ($p > 0.05$).

2. Trust and Visual System Degradations

Under LLA control: There was no significant effect of the type of visual information degradation error on each trust element rating in trust in automation questionnaire under low level of automation control ($p > 0.05$).

Under HLA control: Results revealed that there was a significant effect of the type of visual system degradation error on “harmful” rating which is one element of trust questionnaire stated as “The system’s action will have a harmful or injurious outcome” $F_{0.05,1,11} = 5.50, p = 0.0388^*$. Subjects in high level automation group thought the system with permanent visual degradation error had a significant larger possibility of having harmful outcome than the system with intermittent errors (see Figure 4.3).

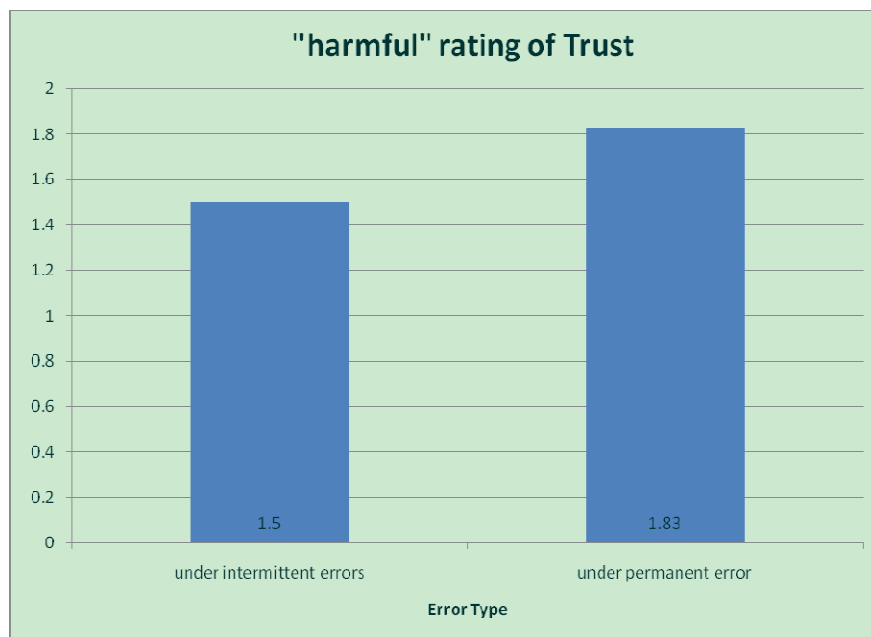


Figure 4.3 Mean “harmful” rating of trust questionnaire under two types of error in HLA

D. Situation Awareness (SA)

1. SA and LOAs

According to the analyses of variance, there was no significant difference between two levels of automation in overall score of situation awareness and its ratings in both intermittent errors and permanent error sessions ($p > 0.05$).

2. SA and Visual System Degradations

Under LLA control: Analyses of variance were conducted on the situation awareness ratings by two types of error in LLA. There was no significant effect of the type of visual information degradation error on overall score of situation awareness and its ratings ($p > 0.05$).

Under HLA control: Results indicated that the effect of the type of visual information degradation error was not significant in influencing the overall situation awareness score computed by the ten different dimensions ($p > 0.05$). However, compared two types of visual degradation error there were significant differences in “Instability of situation” rating $F_{0.05,1,11} = 11.47, p = 0.0061^*$ and “Complexity of situation” rating $F_{0.05,1,11} = 9.14, p = 0.0116^*$ of SART which defined at “Attentional Demand” level of situation awareness as the likeliness of the situation to change suddenly and the degree of complication of the situation. The rating questions asked were “how changeable is the situation” and “how complicated is the situation”. If subjects thought the situation was stable or simple, then they tended to choose a low score of the each rating; otherwise, they chose a high score. In the experiment, subjects felt that the high level automation control under intermittent errors was more complex and more unstable than the control under permanent error (see Figure 4.4 and 4.5).

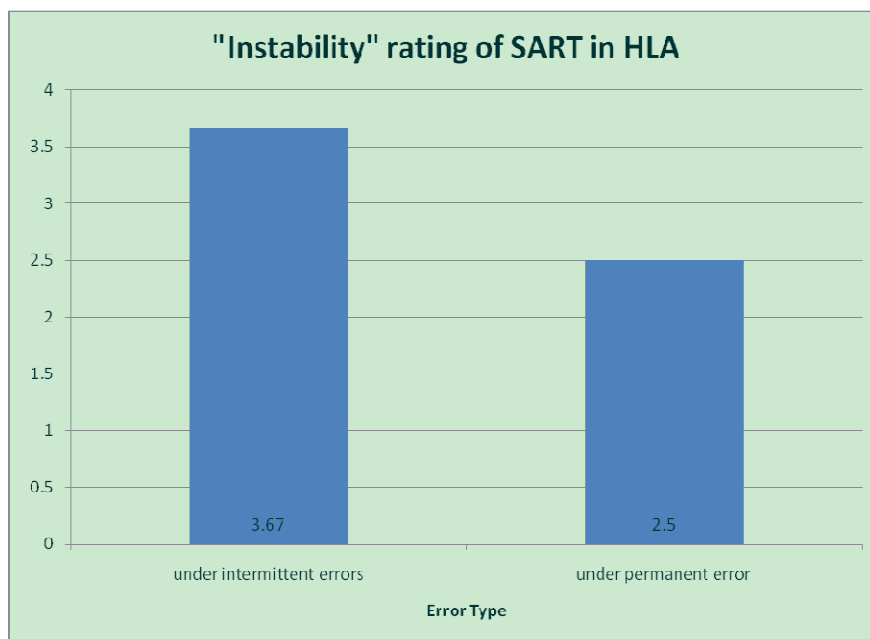


Figure 4.4 Mean "Instability" rating of SART under two types of error in HLA

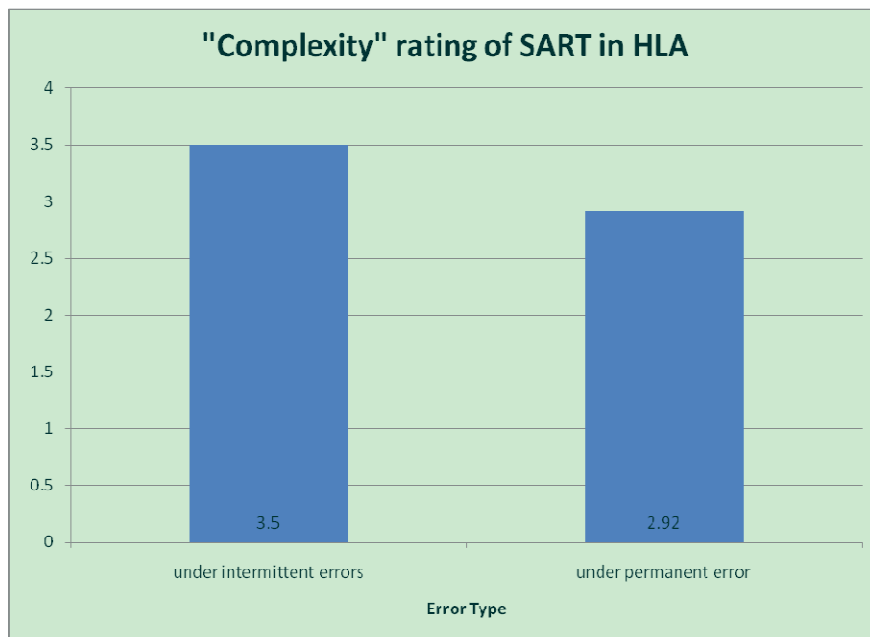


Figure 4.5 Mean "Complexity" rating of SART under two types of error in HLA

E. Mental Workload

Similar to situation awareness, there were no significant differences in overall workload scores of NASA TLX by LOAs, and two types of error with both $p > 0.05$. The former

finding is consistent with the results of Endsley and Kaber's research (1999) that "Action Support" was no significant different from "Shared Control" in workload.

F. Discussion

The purpose of the study was to examine how different levels of automation of varying perception reliabilities affected human performance, trust in automation, situation awareness and mental workload in searching and exploring tasks using tele-robot. Specially, the study was concerned with compared to control under normal condition how the use of LLA and HLA automation affected operators' abilities to perform the task when intermittent errors and permanent error occurred as visual system degradation during tasks. The following sections will describe the results in relation to the questions of interest and the manipulated variables with the correlation analyses. The explanations are offered for the interaction between perception reliability and levels of automation in terms of how they affect human performance, trustiness in automation, situation awareness and mental workload.

1. Level of Automation

According to the results showed in Table 4.2, hit rate was significant different between low and high level automation group in permanent error session showing that high level automation tended to have lower number of collision during entire testing session. This finding reveals the benefit of combination human decision making and computer processing motion control over the tele-robot. Direct remote motion control using joystick controller required human involvement in the implementation aspect of the searching task which needed the motion path control so that produced the lower performance. This can be attributed in part to the difficulty participants had in controlling the tele-robot using the joystick controller. They were required to mentally map translate and rotate parameters from

the hand-controller to the real movement of the tele-robot according the information on the user interface, especially the drive performance panel and camera visual system. It appeared to be a cognitive consumption of subjects to keep track of and isolate all different movements during performance. And unavoidably, for a real world tele-operation, time delay and external noise more or less impact on the human-telerobot system which may make high cognitive consumption issue more serious. The high level automation control involved the human decision making combined with computer-generated alternatives assistant to generate an optimal plan. Then the computer implemented the plans. This combination of human decision making with computer processing, in the context of the tele-operation, served to significantly benefit performance accuracy.

Table 4.2 Summary of the analyses of variance by LOAs in two types of error sessions

	Dependent Variables	F Ratio	Prob>F
Intermittent Error	NASA TLX	0.05	0.8256
	SART	1.87	0.1851
	Hit Rate	0.09	0.7723
	Completion Time	3.8	0.0642
	# targets found	1.16	0.2936
	# wrong location	0.85	0.3676
	Dependent Variables	F Ratio	Prob>F
Permanent Error	NASA TLX	0.92	0.3477
	SART	0.01	0.92
	Hit Rate	15.59	0.0007*
	Completion Time	3.68	0.068
	# targets found	1.94	0.1775
	# wrong location	0.61	0.4441

*There was a significant effect of LOA on the response

2. Visual Information Degradations

Table 4.3 Summary of the analyses of variance by types of error in two LOAs groups

	Dependent Variables	F Ratio	Prob>F
Low Level of Automation	NASA TLX	1.7	0.2189
	SART	0.63	0.4432
	harmful of Trust	0	1
	Hit Rate	7.59	0.0187*
	Completion Time	0.0037	0.9529
	# targets found	1	0.3388
	# wrong location	3.14	0.1039
	Instability of SA	1.1	0.3172
	Complexity of SA	2.13	0.1725
	Dependent Variables	F Ratio	Prob>F
High Level of Automation	NASA TLX	1.23	0.2914
	SART	2.8	0.1224
	harmful of Trust	5.5	0.0388*
	Hit Rate	3.37	0.0936
	Completion Time	1.71	0.2175
	# targets found	0	1
	# wrong location	0.32	0.275
	Instability of SA	11.47	0.0061*
	Complexity of SA	9.14	0.0116*

*There was a significant effect of type of error on the response

In low level automation group, hit rate was significant higher when permanent error occurred than the occurrence of intermittent errors (see Table 4.3 low level of automation). Subjects tended to have a worse performance when visual information degradation got worse, in the experiment as the frequency of error occurrence increased.

In high level automation group, there was no difference in human performance between two types of visual information degradation error revealing that system with automatic execution operated task consistently even under the unexpected situations. The finding demonstrates tele-robot automatic execution assists human operator to gain efficiency and accuracy of the performance. However, compared to intermittent error session, subjects felt

that the system may have more harmful outcome in permanent error session due to the worse monitoring condition. These findings could be explained by that during a certain period of time without direct tele-operation high level of automation with malfunctions could make subjects consider about the worse outcome of the system due to the slow recovery of trust (Endsley & Kaber, 1999). In terms of situation awareness, compared two types of visual degradation error subjects thought permanent error was more stable and more straightforward than the intermittent error. The finding supports the distinction of the two different types of visual degradation error with different degrees of mental demand (see Table 4.3 high level of automation).

3. Trust and Situation Awareness over time

Table 4.4 Summary of the analyses of variance by day in two LOAs groups

	Dependent Variables	F Ratio	Prob>F
Low Level of Automation	NASA TLX	1.28	0.2816
	SART	1.63	0.2285
	confident of Trust	4.71	0.0527
	familiar of Trust	11	0.0069*
	Concentration of SA	8.19	0.0155*
	Attentional Demand L1 of SA	4.74	0.0522
	Attentional Supply L2 of SA	9.9576	0.0092*
	Understanding L3 of SA	5.09	0.0453*
	Dependent Variables	F Ratio	Prob>F
High Level of Automation	NASA TLX	0.04	0.8407
	SART	1.59	0.233
	confident of Trust	0.14	0.7126
	familiar of Trust	15.63	0.0023*
	Concentration of SA	6.6	0.0261*
	Attentional Demand L1 of SA	0.15	0.71
	Attentional Supply L2 of SA	10.22	0.0085*
	Understanding L3 of SA	0.46	0.5101

*There was a significant effect of day (time) on the response

Analyses of variance were conducted on data of trust ratings recorded during first 30-minute testing on the first day and on the second day in both LLA and HLA group data sets (see Table 4.4). Results indicated that in LLA group there was a significant effect of day on “familiar” rating of trust in trust questionnaire which stated as “I am familiar with the system” $F_{0.05,1,11} = 11.00, p = 0.0069^*$, as well as in HLA group $F_{0.05,1,11} = 15.63, p = 0.0023^*$. Figure 4.6 and Figure 4.7 show you the comparisons of the rating in LLA group and HLA group. It revealed that no matter using low level automation or high level of automation, subjects were more familiar with the system on the second day testing compared to the first day testing. The finding reveals that when it is reliable increasing use of automation may increase the familiarity of operators with the system showing the benefit of training.

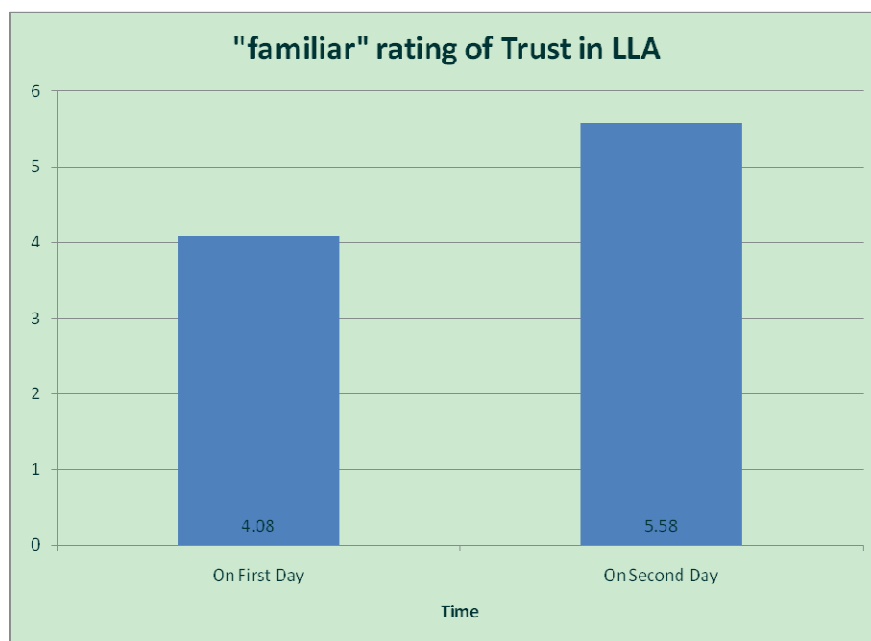


Figure 4.6 Mean “familiar” rating of trust questionnaire by day in LLA

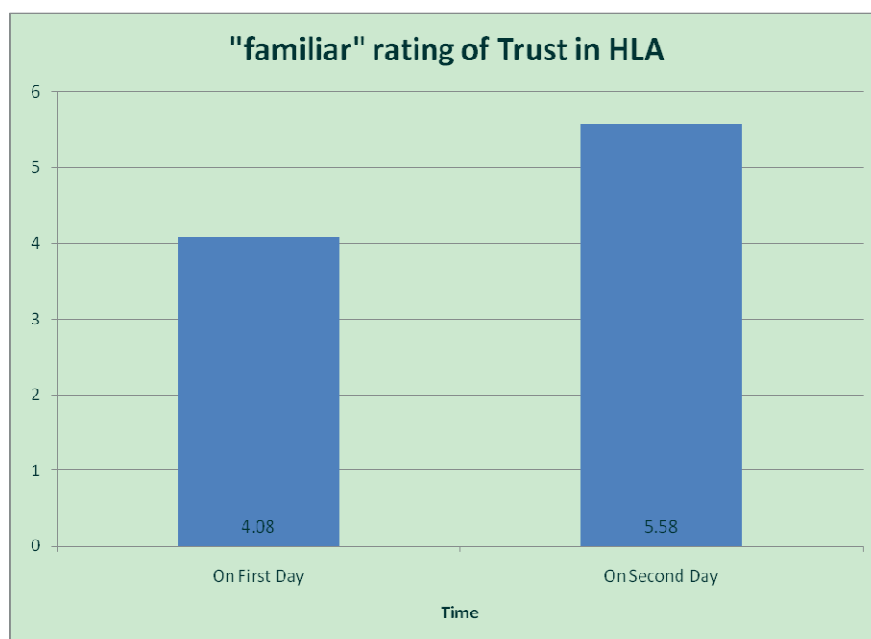


Figure 4.7 Mean “familiar” rating of trust questionnaire by day in HLA

Analyses of variance were conducted on data of SART ratings recorded during first 30-minute testing on the first day with no degradation and on the second day with degradations in both LLA and HLA group data sets. Combine ratings to three levels of situation awareness, results indicated that in LLA group there was a significant effect of day on “Attentional Supply” level of situation awareness which involved how much mental resource supplied to complete the task $F_{0.05,1,11} = 9.96, p = 0.0092^*$, as well as in HLA group $F_{0.05,1,11} = 10.22, p = 0.0085^*$ (see Figure 4.8, 4.9). This would be considered a lack of level 2 situation awareness which is as the comprehension of the significance of perceived information (Endsley, 1995). It revealed that on the second day when visual degradation error occurred, no matter using low level automation or high level of automation, subjects could not supply more attention to the situation compared to the first day testing.

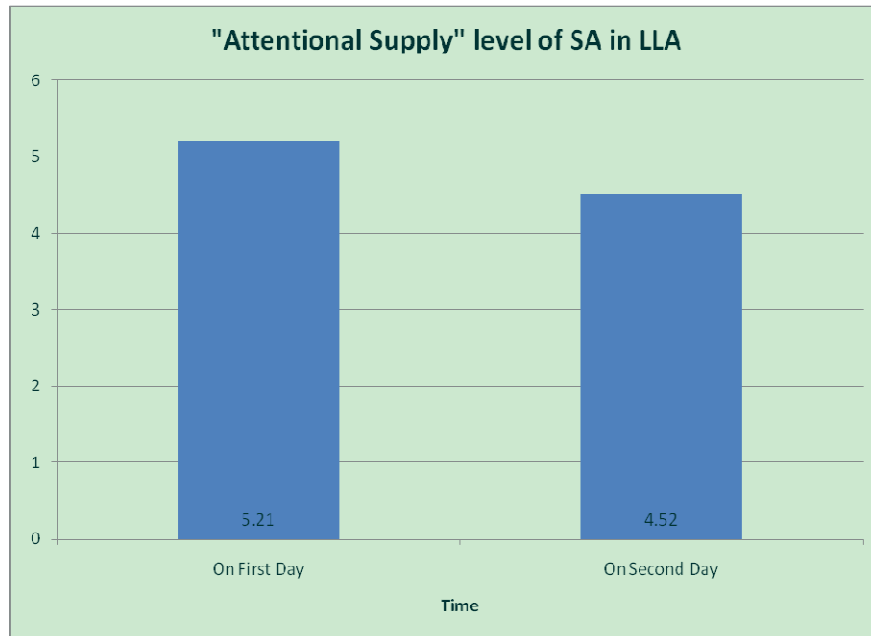


Figure 4.8 Mean “Attentional Supply” level of SA by day in LLA

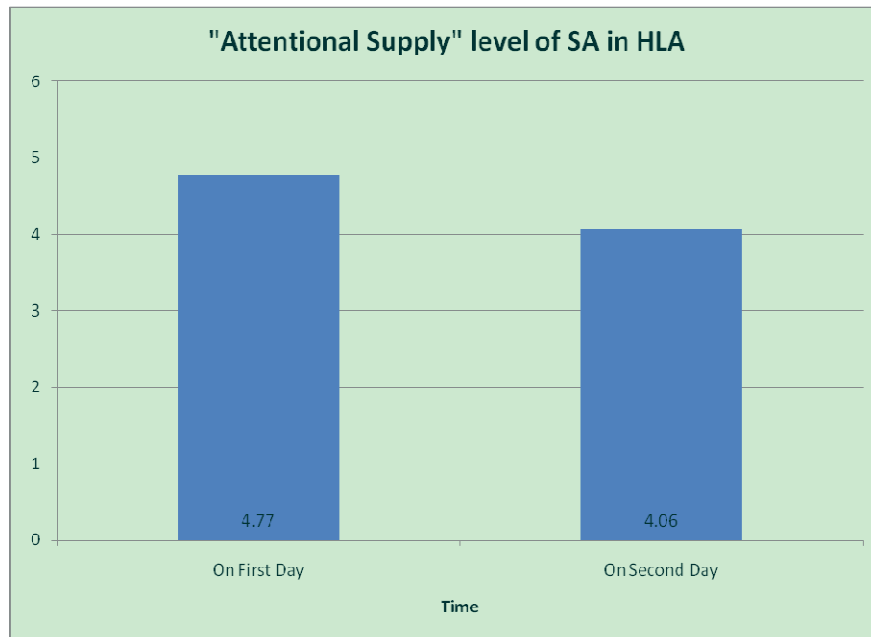


Figure 4.9 Mean “Attentional Supply” level of SA by day in HLA

Its “Concentration” rating defined as the degree that one’s thoughts are brought to bear on the situation was significantly affected by the occurrence of error in low level of

automation group $F_{0.05,1,11} = 8.19, p = 0.0155^*$ and high level automation group $F_{0.05,1,11} = 6.6, p = 0.0261^*$ (see Figure 4.10, 4.11). The rating question asked was “how much are you concentrating on the situation”: if subject brought all his or her thought to bear the situation then he or she tended to choose a high score of the rating; otherwise, he or she chose a low score. In the experiment, subjects had a lower concentration on the situation on the second day first 30-minute testing session with degradations compared to the first day testing session with no degradation. This finding reveals the importance of the reliability of automation due to the degradation of available attention supplied to the situation, particularly mental resource supplied to concentrate on the situation. The situation with error occurrence strips the operators of the resource of concentration on the task.

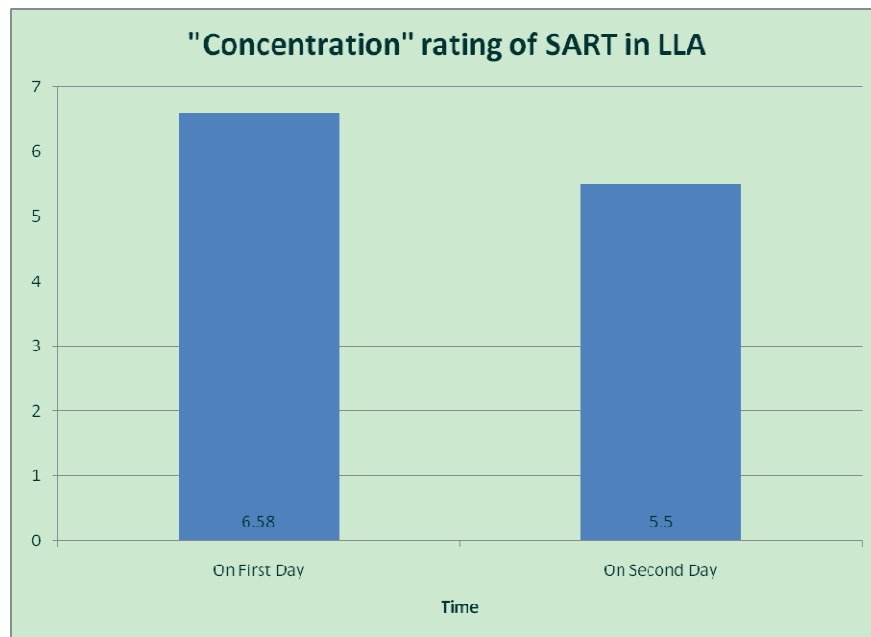


Figure 4.10 Mean “Concentration” rating of SART by day in LLA

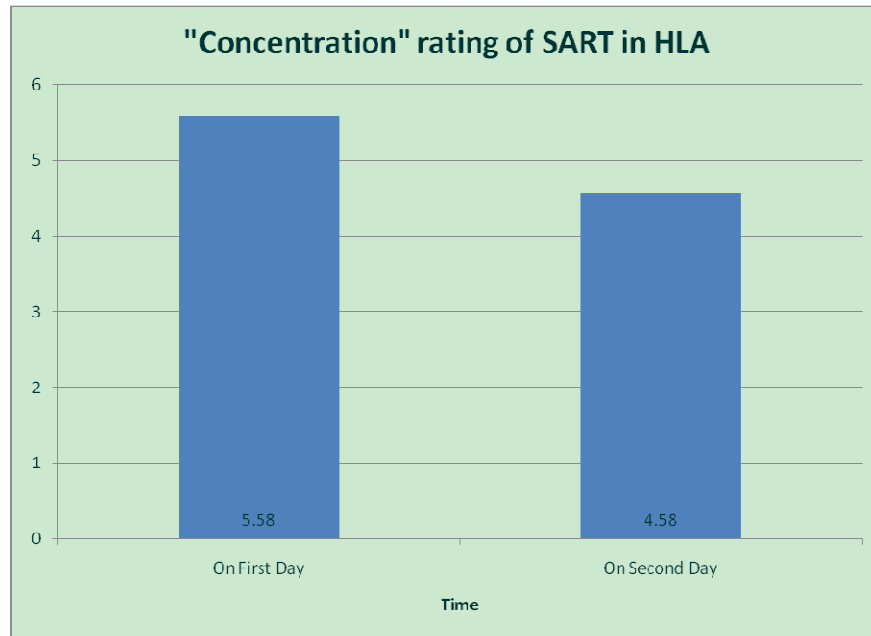


Figure 4.11 Mean “Concentration” rating of SART by day in HLA

Results also indicated that in LLA group there was a significant effect of day on “Understanding” level of situation awareness which involved the quantity and the quality of the information gained from the situation and the degree of how familiar with the situation $F_{0.05,1,11} = 5.09, p = 0.0453^*$ (see Figure 4.12), but there was no significant difference in “Understanding” level of SA by day in HLA group ($p > 0.05$). Subjects accumulated the knowledge and the information by direct manual motion control over time so that they understood the situation better than before. Using high level automation control which execute the task by computer strips operators of the capability to personally practice and perceive the characteristics of robot motion so that although robot execution over time, they could not further understand the system by monitoring.

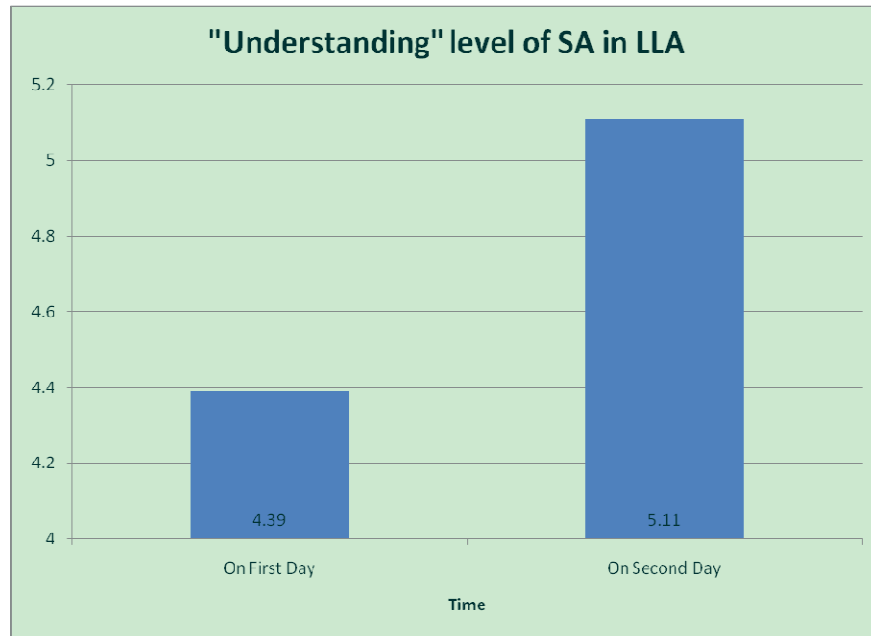


Figure 4.12 Mean “Understanding” level of SA by day in LLA

4. Correlation of Responses

Correlation of Trust with Performance: Correlation analyses were conducted on the performance and trust in automation response measures. Under low level of automation, results revealed a significant negative correlation, $r = -0.6563$, $p = 0.0204$, between the hit rate and one of the positively trust related statements ratings – “I can trust the system” with permanent error. Under high level of automation, results revealed significant negative correlations between the number of wrong location operator marked during exploring task and two of the positively trust related statements ratings – “I am confident in the system” and “I am familiar with the system” ($r = -0.7225$, $p = 0.0107$ and $r = -0.6285$, $p = 0.0286$). That is, the variation of human performance may associate with the change of trust in automation due to the opposite trends.

Correlation of Trust with Situation Awareness: Correlation analyses were conducted on the trust in automation and SART response measures. Table 4.5 shows the significant correlation between trust and situation awareness overall score as well as levels of SA in different levels of automation with various type of error revealing the association of trust elements with situation awareness and its levels. Endsley (2000) concluded the decline in SA to a number of automation-related factors, including increased monitoring demands and subsequent vigilance decrements, complacency caused by overreliance on automation, system complexity, poor interface design, and a lack of trust in the automation.

Table 4.5 Summary of correlation between trust and SA at combined conditions

LOA	Session	Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
Low	intermittent error	Overall score of SA	familiar	0.6194	12	0.0706	0.8804	0.0317
		Demand L1 of SA	deceptive	0.7995	12	0.417	0.9415	0.0018
	permanent error	Overall score of SA	familiar	0.612	12	0.0587	0.8776	0.0344
		Demand L1 of SA	deceptive	0.6291	12	0.0864	0.8839	0.0284
High	intermittent error	Understanding L3 of SA	dependable	0.7008	12	0.2123	0.9091	0.0111
		Understanding L3 of SA	reliable	0.6145	12	0.0627	0.8786	0.0335
		Understanding L3 of SA	trust system	0.6015	12	0.0422	0.8738	0.0385
	permanent error	Understanding L3 of SA	dependable	0.773	12	0.3578	0.933	0.0032
		Understanding L3 of SA	reliable	0.7496	12	0.3083	0.9254	0.005
		Understanding L3 of SA	trust system	0.7061	12	0.2222	0.9109	0.0103

Correlation of Trust with Mental Workload: Correlation analyses were conducted on the trust in automation and NASA TLX response measures. Table 4.6 shows the significant correlation between trust and workload overall score in different levels of automation with permanent error. Results revealed that operators' workload may associate with trust elements. In particular, there were significant negative correlations of workload with positively trust

related statements ratings, while there were significant positive correlations of workload with negatively trust related statements ratings. This finding is supported by Brown and Galster's work (2004) that pilot trust in the automation was higher when the workload level was low and that there was a significant difference in confidence ratings between low and high workload levels when the automation was unreliable.

Table 4.6 Summary of correlation between trust and NASA TLX at combined conditions

LOA	Session	Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
Low	permanent error	NASA TLX	trust system	-0.6833	12	-0.9031	-0.18	0.0143
		NASA TLX	dependable	-0.6374	12	-0.8869	-0.1001	0.0258
		NASA TLX	confident	-0.5895	12	-0.8693	-0.0236	0.0437
High	permanent error	NASA TLX	underhanded	0.5927	12	0.0285	0.8705	0.0423
		NASA TLX	wary	0.5919	12	0.0273	0.8702	0.0426

CHAPTER 5 CONCLUSIONS AND RECOMMENDATION

A. Conclusion

The study demonstrated that high level of automation described in this experiment enhance human performance through computer implementing actions especially if it is highly reliable. Its ability appeared to relieve operators of mental demand in searching, detecting and exploring the location of targets, allowing them to develop more complete and accurate knowledge of environment states. It is possible that the use of high level automation in complex controlling tasks, involving unexpected or real hazardous conditions, have different affects on operator performance and SA compared to conditions with environmental stress. Future work should look at the different effect of LOAs under real firefighting condition.

The experiment revealed when using direct manual motion control as a low level automation permanent error as an accumulated small error has a worse effect on operator performance compared to the intermittent errors. It is required high level automation involves to reduce the mental demand of operator.

The current study identified the reliability of high level automation (computer automatic execution) as influencing factors in the linkages of SA levels. Further validation of the relationship and identification of other influencing factors may lead to the model being used for predictive purposes in future systems design. For example, it could be used to predict SA levels in controlling with new forms of high level automation sharing characteristics with current forms of automation in order to determine whether add the assistant of manual control or other aids.

Under high level of automation that provided computer guidance to subjects, operators appeared to become doubtful of the automated control of tele-robot when error gets worse and worse. Automation was generally expected to be more reliable and make fewer errors than a human in the tele-robot task. However, when participants experienced automation errors or inefficiency in control, their trust in automation declined more sharply than trust in automation with manual motion control. Therefore, although there was evidence that high level of automation benefited human performance; users were not able to capitalize on its full potential, possibly because its behavior was less transparent to users. This effect was attributed to maintain operator involvement in the system control loop during operations. This finding is in agreement with recent research by Endsley and Kaber (1999) and Endsley and Kiris (1995) who have all noted difficulties in performance when humans are acting with the assistance. Finally, the study also demonstrated that operators had better understanding of the driving environment when they had higher trust in automation. The resulting knowledge of trust in automation should be applied or considered to the development of future technologies, or the training to the use of automation in human-telerobot system.

In general, results from this experiment confirm many of the findings of previous research (Endsley & Kaber, 1999, Kaber et al., 2006) through a realistic task by using a real human-telerobot system boosting meaningfulness of the results to the design of human-telerobot system and general dynamic robot control systems. The study affirms that trust increases when increasing use of automation in both low and high level of automation system presenting by the increasing familiarity of operator with the system; while operators are most likely to distrust the high level of trust when serious error occurred (high frequency error). On the other hand, when human operators must take control in the event of an automation

failure they are affected by trust in automation. Results suggest that training is necessary to be done to help operators to understand the levels of automation in order to calibrate their trust in the automation appropriately.

There are some limitations of this study that should be noted with respect to using the results as a basis for designing or making decision the use of automation. First, care must be taken in generalization of these results, as a specific task type (searching, detecting and exploring threat targets in firefighting) was investigated. Another limitation of the experiment was the order of presentation of reliability conditions of automation. Due to time constraint we did not randomly assigned the no degradation control to each level of automation. Operator trust in automation was investigated by randomly assigning two types of visual information degradation error in the second day testing sessions. However, no degradation control as a baseline may be important to measure the difference from error occurs. As well, the controlled technological limitations of the computer in the human-telerobot system had an impact on the data.

B. Recommendations for Future Research

This paper has presented a comparison between defined LLA and HLA. The need exists for further research into how human-telerobot system performance is affected by LOAs, perception reliability and trust in automation. This type of research needs to be improved by using high technology of computer programming to make tele-robot implementation more smoothly in high level of automation and to make human-telerobot system more dependable and more integrate such as the interface with the functions of giving direction and mapping the route in real-time. Further investigations are needed to assess whether adaptive automation such as HLA in the experiment can be used to other LOAs over time to achieve

improvements in performance. On the basis of this study, directions of future research include investigating trust in automation and SA in tele-operation task with other LOAs and further describing the relationships among the various elements of trust, levels of SA, and operator performance.

From an experimental design perspective, 3-day separate testing with fixed difficulty of room structures and target locations provide a controlled environment that is ideal for examining issues of perception reliability with three conditions: normal, intermittent errors and permanent error on trust in automation, situation awareness, mental workload and human performance in different LOAs systems. The random ordering of automation reliability conditions might serve to provide clear statistical conclusions on the effects of type of visual degradation error on operator SA and performance.

Future research aimed at incrementally advancing the present study includes introducing additional response measures, for example, eye tracking. Participants visualized the user interface which displayed the control and visual information and the hard copy of the suburb map during experiment trials. It appeared that drivers adopted different visual scanning strategies to balance performance in multitasking (i.e., searching and exploring). Eye tracking data could provide more evidence on the relationships among trust in automation, SA, workload and performance by detailing what drivers attend to, when. A challenging and worthwhile direction of future study involves applying real-world tasks and experienced operators by considering the operators' stress as well as his/her professional knowledge, skill and experience.

APPENDIX A SART 10D RATING

SART 10D Rating Sheet

Participant No.:

Instability of Situation

Low	How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straightforward (low)?	High
Low	Complexity of Situation How complicated is the situation? Is it complex with many interrelated components (high) or is it simple and straightforward (low)?	High
Low	Variability of Situation How many variables are changing in the situation? Are there a large number of factors varying (high) or are there very few variables changing (low)?	High
Low	Arousal How aroused are you in the situation? Are you alert and ready for activity (high) or do you have a low degree of alertness (low)?	High
Low	Concentration of Attention How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?	High
Low	Division of Attention How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (high) or focused on only one (low)?	High
Low	Spare Mental Capacity How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (high) or nothing to spare at all (low)?	High
Low	Information Quantity How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?	High
Low	Information Quality How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it a new situation (low)?	High
Low	Familiarity with situation How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation (low)?	High
0		7

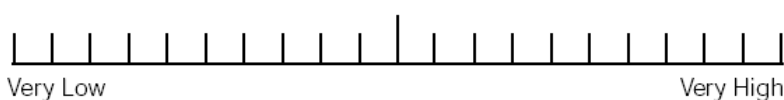
APPENDIX B NASA TLX

NASA Task Load Index

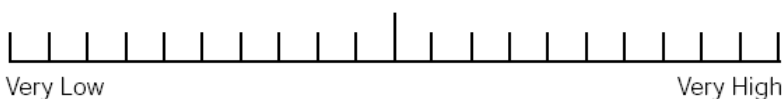
Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

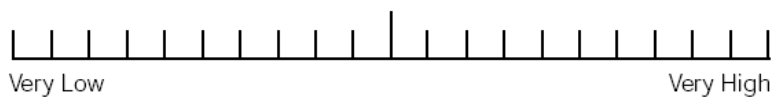
Mental Demand How mentally demanding was the task?



Physical Demand How physically demanding was the task?



Temporal Demand How hurried or rushed was the pace of the task?



Performance How successful were you in accomplishing what you were asked to do?



Effort How hard did you have to work to accomplish your level of performance?



Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?



APPENDIX C TRUST IN AUTOMATION CHECKLIST

Checklist for Trust between people and automation (Jian et al., 2000)

Below is a list of statement for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine.

Please mark an 'x' on each line at the point which best describes your feeling or your impression.

(Note: 'not at all' = 1; 'extremely' = 7)

1.	The system is deceptive	1	2	3	4	5	6	7
2.	The system behaves in an underhanded manner	1	2	3	4	5	6	7
3.	I am suspicious of the system's intent, action, or outputs	1	2	3	4	5	6	7
4.	I am wary of the system	1	2	3	4	5	6	7
5.	The system's actions will have a harmful or injurious outcome	1	2	3	4	5	6	7
6.	I am confident in the system	1	2	3	4	5	6	7
7.	The system provides security	1	2	3	4	5	6	7
8.	The system has integrity	1	2	3	4	5	6	7
9.	The system is dependable	1	2	3	4	5	6	7
10.	The system is reliable	1	2	3	4	5	6	7
11.	I can trust the system	1	2	3	4	5	6	7
12.	I am familiar with the system	1	2	3	4	5	6	7

APPENDIX D PRE-EXPERIMENTAL QUESTIONNAIRE

Pre-Experimental Questionnaire

Do you play video games often?	yes	/	no
If yes, how often?			
Do you play with remote control cars often?	yes	/	no
Do you have a remote control car at home?	yes	/	no
Do you use a joystick controller often?	yes	/	no
Do you drive a vehicle every day?	yes	/	no
Have you ever done a Tele-operational experiment?	yes	/	no
Are you interested in the experiment?	yes	/	no

BIBLIOGRAPHY

- Ahlstrom, V., Longo, M., & Truitt, T. (2005). *Human factors design guide* (DOT/FAA/CT-02-11). Atlantic City, NJ: Federal Aviation Administration.
- Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, 28(2), 85–97
- Bixby, L.. "Firefighter safety spurs interest in Robots". *Fire Apparatus & Emergency Equipment*. August 2007 volume12(8).
- Brown, R. D. & Galster, S. M. Effects of reliable and unreliable automation on subjective measures of mental workload, situation awareness, trust, and confidence in dynamic flight task. *In Proceedings of the 48 th Annual Meeting of the Human Factors and Ergonomics Society*, New Orleans; LA, pp 147-151.
- Conway, A.R. and Engle, R.W., 1994. Working memory and retrieval: A resource dependent inhibition model. *Journal of Experimental Psychology General*, 4, 354–373.
- Driewer, F., H. Baier, et al. (2005). Robot-human rescue teams: a user requirements analysis. *Advanced Robotics* 19(8): 819-838.
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, 13, 147–164.
- Endsley, M. 1987, The application of human factors to the development of expert systems for advanced cockpits, *Proceedings of the Human Factors Society 31st Annual Meeting* (Santa Monica, CA: Human Factors Society), 1388 - 1392.
- Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement, *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97–101). Santa Monica, CA: Human Factors and Ergonomics Society.
- Endsley, M. R. and E. O. Kiris (1995). "THE OUT-OF-THE-LOOP PERFORMANCE PROBLEM AND LEVEL OF CONTROL IN AUTOMATION." *Human Factors* 37(2): 381-394.
- Endsley, M.R., & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42, 462– 492.
- Endsley, M. R. (2000). Situation models: An avenue to the modeling of mental models. *Proceedings of the 14th Triennial Congress of the International Ergonomics*

Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, , 14(44)

- Dhillon, B. S., *Robot Reliability and Safety*, Springer-Verlag New York, Inc., Secaucus, NJ, 1991
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human- Computer Studies*, 58, 697.
- Fahey, L. J. (2007). *Arleigh burke grounding INS plot*. Unpublished manuscript.
- Galster, S. M., Bolia, R. S., Roe, M. M., & Parasuraman, R. (2001). Effects of automated cueing on decision implementation in a visual search task. *In Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 321–325). Santa Monica, CA: Human Factors and Ergonomics Society.
- Halprin, S., Johnson, E., & Thornburry, J. (1973). Cognitive reliability in manned systems. *IEEE Transactions on Reliability*, R-22, 165-169.
- Hancock, P. A., & Scallen, S. F. (1996, October). The future of function allocation. *Ergonomics in Design* (October), 24–29.
- Hancock, P.A. and Warm, J.S., 1989. A dynamic model of stress and sustained attention. *Human Factors*, 31, 519–537.
- 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. Hancock, N. Meshkati (Eds.) pp.139-183. Amsterdam, The Netherlands: North Holland B.V.
- Hawley, J. K., Mares, A. L., & Giammanco, C.A. (2005). *The human side of automation: Lessons for air defense command and control* (Tech. Rep. ARL-TR-368). Aberdeen Proving Ground, MD: Army Research Laboratory, Human Research and Engineering Directorate.
- Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53-71.
- Kaber, D.B., Onal, E., & Endsley, M.R. (1997). Design and development of a comprehensive user interface for teleoperator control in nuclear applications. *Proceedings of the 7th Topical Meeting on Robotics and Remote Systems* (pp. 947–954). LaGrange Park, IL: American Nuclear Society.

- Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, 36, 447–462.
- Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35, 1243–1270.
- Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40, 153–184.
- Lee, J. D., & See, K. A. (2004). Trust in technology: Designing for appropriate reliance. *Human Factors*, 46, 50–80.
- Lee, J. D. (2006). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1570–1596). Hoboken, NJ: Wiley.
- Lee, J. D. (2008). Review of a pivotal Human Factors article: "Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 50(3): 404-410.
- Lewandowsky, S., Mundy, M., & Tan, G. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology – Applied*, 6, 104–123.
- Masalonis, A. J., R. Parasuraman, et al. (1999). Trust as a construct for evaluation of automated aids: Past and future theory and research. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Vols 1 and 2*: 184-188.
- Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*, 47, 35–49.
- Miller, C. A. and R. Parasuraman (2007). "Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control." *Human Factors* 49(1): 57-75.
- Muir, B. M. (1987). Trust between humans and machines, and the design of decision aides. *International Journal of Man- Machine Studies*, 27, 527–539.
- Muir, B. M. (1994). Trust in automation: 1. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37, 1905–1922.
- Muir, B. M., & Moray, N. (1996). Trust in automation: 2. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39, 429–460.

- Muir, B. M. (1989). Operators' trust in and use of automatic controllers in a supervisory process control task. *Unpublished Ph.D. dissertation*, University of Toronto, Toronto, Canada.
- National Aeronautics and Space Administration (NASA). (2008). *National research announcement #IAFD*. Washington, DC: Author.
- Necka, E., 1997. Attention, working memory and arousal: Concepts apt to account for the 'process of intelligence'. In: G. Matthews, ed. *Cognitive science perspectives on personality and emotion*. Amsterdam: Elsevier.
- Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38, 665–679.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cybernetics – Part A: Systems and Humans*, 30, 286–297.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230–253.
- Parasuraman, R. and C. D. Wickens (2008). Humans: Still vital after all these years of automation. *Human Factors* 50(3): 511-520.
- Parasuraman, R., & Wilson, G. F. (2008). Putting the brain to work: Neuroergonomics past, present, and future. *Human Factors*, 50, 468–474.
- Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95–112.
- Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6), 1296-1308.
- Schmorrow, D., Stanney, K.M., Wilson, G., & Young, P. (2006). Augmented cognition in human-system interaction. In G. Salvendy (Eds.), *Handbook of Human Factors and Ergonomic* (3rd Ed.). New Jersey: Wiley and Sons, Inc.
- Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.
- Sheridan, T. B. (2002). *Humans and automation: Systems design and research issues*. New York: Wiley.

- Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Individual differences in monitoring failures of automation. *Journal of General Psychology*, 120, 357–373.
- Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 991–1006.
- Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, 52, 701–717.
- Starter, N., and Woods, D. 1994. Pilot Interaction with Cockpit Automation II: An Experimental Study of Pilot's Model and Awareness of the FLIGHT MANAGEMENT SYSTEM. *International Journal of Aviation Psychology* 4:1–28.
- Sarter, N., Woods, D., & Billings, C. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). New York: Wiley.
- Sturges, R.H. 1990a. A quantification of machine dexterity applied to an assembly task. *International Journal of Robotics Research* 9(3).
- Taylor, R.M. Situational awareness rating technique (SART): the development of a tool for aircrew systems design. *Proceedings of the NATO Advisory Group for Aerospace Research and Development (AGARD) Situational Awareness in Aerospace Operations Symposium (AGARD-CP- 478)*; October 1989, p. 17.
- Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors*, 2, 99–126.
- Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effect of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, 2(4), 352.
- Wilkison, B. D., Fisk, A. D., & Rogers, W. A. (2007). Effects of mental model quality on collaborative system performance. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, Baltimore, MD, 31 1506.
- Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, 43, 355–365.
- Zuboff S. 1988. *In the Age of the Smart Machine*. New York: Basic Books

ACKNOWLEDGMENTS

The successful completion of this master's project is due to the support I received from several people. First, I would like to express my gratitude toward my thesis advisor, Dr. Richard Stone, for his encouragement and guidance throughout my thesis development. His positive advice was critical to my success. Additionally, I thank my thesis committee members for their support and patience with me. Dr. Gary Mirka and Dr. Stephen Vardeman were helpful in shaping my experiment design and executing the experiment as well as Dr. W. Robert Stephenson in the Statistics department. I am grateful to each of these people for sharing their valuable time for my benefit.

Finally, I thank my parents, my dear grandmother and my friends in Iowa State University, especially my research assistant Wang. Their love, encouragement, patience and support of my ideal were the biggest contributor to my success.