Army Research Laboratory

Development of Principles for Multimodal Displays in Army Human-Robot Operations

by Michael D. Coovert, Matthew S. Prewett, Kristin N. Saboe, and Ryan C. Johnson

ARL-CR-651 **June 2010**

prepared by

MDC and Associates, Inc.
4004 Stanley Rd.
Plant City, FL 33565

and

University of South Florida
4202 E. Fowler Ave.
Tampa, FL 33620

under contract

DAAD19-01-C-0065

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-CR-651 **June 2010**

Development of Principles for Multimodal Displays in Army Human-Robot Operations

Michael D. Coovert
MDC and Associates, Inc.

Matthew S. Prewett, Kristin N. Saboe, and Ryan C. Johnson
University of South Florida

prepared by

MDC and Associates, Inc.
4004 Stanley Rd.
Plant City, FL 33565

and

University of South Florida
4202 E. Fowler Ave.
Tampa, FL 33620

under contract

DAAD19-01-C-0065

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
June 2010	Final	August 2009–September 2009

4. TITLE AND SUBTITLE Development of Principles for Multimodal Displays in Army Human-Robot Operations	**5a. CONTRACT NUMBER** DAAD19-01-C-0065
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
I 6. AUTHOR(S) Michael D. Coovert,[*,†] Matthew S. Prewett,[†] Kristin N. Saboe,[†] and Ryan C. Johnson[†]	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MDC and Associates, Inc. 4004 Stanley Rd. Plant City, FL 33565 University of South Florida 4202 E. Fowler Ave. Tampa, FL 33620	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory ATTN: RDRL-HRM-D Aberdeen Proving Ground, MD 21005-5425	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) ARL-CR-651

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

[*] MDC and Associates, Inc., 4004 Stanley Rd., Plant City, FL 33565

[†] University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620

14. ABSTRACT

Work in the area of robots and human-robot interaction is exploding. This report reviews part of the literature and provides recommendations for future research. Three sections within the report outline topics of special interest: workload, autonomy, and visual displays. Further information on these sections can be found in an online database.

15. SUBJECT TERMS

human-robot interaction, workload, vision, autonomy

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Linda Elliott
Unclassified	Unclassified	Unclassified	UU	60	**19b. TELEPHONE NUMBER** (*Include area code*) 706-545-9145

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures **v**

List of Tables **vi**

Acknowledgments **vii**

1. Introduction **1**

2. Workload in Human-Robot Interaction: A Review of Manipulations and Outcomes **2**

2.1 Introduction 2

2.1.1 Workload Manipulations in HRI 2

2.1.2 Purpose 3

2.2 Method....3

2.2.1 Literature Search 3

2.2.2 Coding Procedure and Inclusion Criteria 3

2.3 Results 4

2.4 Conclusions 7

3. Autonomy and Automation Reliability in Human-Robot Interaction: A Qualitative Review **8**

3.1 Introduction 8

3.1.1 Levels of Autonomy/Control (LOA)....9

3.1.2 Automated Aid Reliability 9

3.1.3 Purpose 10

3.2 Method....10

3.2.1 Literature Search 10

3.2.2 Inclusion Criteria and Procedure....10

3.3 Results 11

3.3.1 Overall Analysis of LOA 14

3.3.2 Overall Analysis of Automated Aid Reliability 14

3.4. Discussion 15

3.4.1 Guiding Principles 15
3.4.2 Unresolved Issues and Future Directions 15
3.4.3 Summary and Conclusion 16

4. Effectiveness of Visual Devices in Human-Robot Interaction: A Qualitative Review 17

4.1 Introduction 17
4.1.1 Time Resources 18
4.1.2 Contextual Resources 18
4.1.3 Visual Processing Resources 19
4.2 Method 19
4.2.1 Definitions and Inclusion Criteria 19
4.2.2 Procedure 20
4.3 Results and Discussion 20
4.3.1 Time Resources 27
4.3.2 Contextual Resources 27
4.3.3 Results for Visual Processing Resources 28
4.4 Conclusions 29

5. Recommendations for Further Research 30

5.1 Meta-Analysis of Coded HRI Articles 30
5.2 Cumulative Sum Methodology for Modeling Training Effectiveness/Skill Decay 30
5.3 Focus Groups/Structured Interviews for Lessons Learned 31
5.4 Structural Equation Modeling of HRI Effectiveness 32
5.5 Multilevel Data Structures and Models 32
5.5.1 The Linear Multilevel Model 33
5.5.2 HRI Research 35
5.5.3 Multilevel Models of Repeated Measures 36
5.6 HRI Operators Field of Vision 36
5.7 Robot Autonomy and Errors 37

6. References 39

Distribution List 47

List of Figures

Figure 1. Visual modality. 18
Figure 2. Model depicting active areas of HRI research. 31

List of Tables

Table 1. Study summaries on multirobot control....5
Table 2. Summary of studies manipulating single task demands....6
Table 3. Summary of studies examining LOA....12
Table 4. Summary of studies examining automated aid reliability....13
Table 5. Summary of studies manipulating FR....22
Table 6. Summary of studies manipulating latency/time delay....23
Table 7. Summary of studies manipulating SS and MS visual cues....24
Table 8. Summary of studies manipulating FOV....25
Table 9. Summary of studies manipulating camera perspective....26

Acknowledgments

This project was funded by two U.S. Army Technology Objectives: (1) the Army's Robotics Collaboration Army Technology Objective and (2) the Army's Enhanced Situation Understanding for Maneuver Teams. The authors wish to thank Linda Elliott, Elizabeth Redden, and Michael Barnes of the U.S. Army Research Laboratory's Human Research and Engineering Directorate for their support throughout this effort. We are particularly indebted to the many graduate and undergraduate students at the University of South Florida for their dedication and long hours worked in order to identify and code several hundred articles for the literature review and analyses.

This work was supported by government contract number DAAD19-01-C-0065, task order 83. The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

INTENTIONALLY LEFT BLANK.

1. Introduction

Warfighters working with robots are at the cutting edge of the Future Combat Systems (FCS) fighting forces. These individuals work with a diverse set of land, air, sea, and undersea vehicles capable of a variety of missions. The missions vary and can include unattended sensors, reconnaissance, search and rescue, medical support, and direct contact with enemy assets, with the systems ranging from single sensors to multirobot systems. Examples include FCS technologies network, TALON, iRobot, PackBot, the SPARTAN Advanced Concept Technology Demonstration, and the Family of Integrated Rapid Response Equipment sensors and vehicles (Powell et al., 2006). Just as the missions and systems vary greatly, so do the operator control units and multioperator control unit interfaces employed to operate the robots. This variety of missions, robot types, and interfaces can be difficult to train for and manage. It is therefore essential to identify the cognitive and task demands being placed on the warfighter to ensure successful mission outcomes.

Several different approaches are necessary to cover the criterion space of these cognitive and task demands. The main strategy utilized here is an evaluation of the existing literature on human-robot interaction (HRI). Existing documents from the academic and the U.S. Army Research Laboratory literatures were examined and coded. The major dimensions of classifications uncovered included the number of platforms controlled, task difficulty comparisons, level of control by platforms, cuing/decision-making reliability, stereoscopic (SS) vs. monoscopic (MS) display, comparisons between modalities, comparisons within modalities, frame rate (FR), field of vision (FOV), latency/time delay, and camera perspective. A summary of these documents is available upon request.

This report contains several sections that support the taxonomy and provide recommendations for future multimodal displays and research. Sections 2–4 were originally three separate papers, each elaborating on specific aspects of the taxonomy. Each section covers a particular topic in HRI. Section 5 presents proposals for follow-on HRI research.

Due to size constraints, a separate, in-depth analysis of HRI cognitive task dimensions is not presented here but is available upon request from the authors. The in-depth analysis exists in two parts. The first portion is in this report and the second exists online. A database was created in RefWorks (2009) of articles eligible for meta-analysis. The coding sheet for the articles and instructions for using this database are also available from the authors. The database itself exists online and is available via the Web at http://www.refworks.com/.

Especially notable are any guiding principles culled from each article. Section 6 concludes with a references list of the articles in the meta-analysis folder of the REFWORKS database. These studies have been screened and coded as being eligible for meta-analysis.

2. Workload in Human-Robot Interaction: A Review of Manipulations and Outcomes

The current study reviews the relationship between manipulations of teleoperator workload and task outcomes, using multiple resource theory as the underlying framework. Results indicated that controlling more than two platforms is detrimental to many performance indices (reaction time [RT], error rate [ER]), but overall productivity improves. For studies that manipulated workload for a single robot task, visual demands were a limiting factor, and interventions that reduced visual demands improved performance. We conclude with guiding principles for managing workload and improving teleoperator performance.

2.1 Introduction

Autonomous agents have become an essential tool for a myriad of tasks. Through the use of unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs), service personnel can carry out tasks with a reduced risk to their safety. In recognition of these aforementioned advantages, there has been an increased interest in understanding and improving HRI (Chen et al., 2007). From a human factors perspective, understanding and mitigating the impact of workload should improve performance in HRI. This section addresses the issue of workload in HRI through a review of the experimental literature. Existing research has examined a multitude of manipulations and outcomes of workload demands, but a synthesis is needed to understand the state of the current research. The current review provides this need by integrating HRI studies according to manipulations, tasks, and outcomes in order to draw guiding principles.

2.1.1 Workload Manipulations in HRI

This section utilizes multiple resource theory (MRT) as the framework for workload in HRI, as described by Wickens (2002). The main tenets of MRT suggest that multiple cognitive resources allow for multitasking or time-sharing performance. Specifically, tasks requiring different cognitive resources can often be effectively performed together, but competition for the same resource(s) can produce interference. Much of the recent work on MRT has defined these resource channels while predicting the degree to which information from strained resource channels can be effectively offloaded to less-used channels. To summarize, tasks may strain cognitive resources through verbal, manual, or sensory demands (for a complete review, see Wickens [2002]).

Controlling a platform or interacting with an artificial agent imposes many demands, such as executing menu functions, navigating to waypoints, manipulating a foreign object, processing information from data uplinks, and communicating with team members. Most manipulations of HRI workload stem from changing the number of robots available or manipulating the demands of a single task or resource. Multirobot control affects workload by increasing the number of subtasks (monitoring, navigating, and executing). Although providing a user with more than one platform to control will certainly increase workload, will this additional strain outweigh the benefit of having multiple robots to execute task actions? Addressing this question may depend upon the tasks being performed and the criteria desired. Thus, we examine the issue of multirobot control by reviewing the HRI literature according to the tasks and criteria studied.

In contrast to manipulations of robot quantity, other manipulations of workload focus on a single task or cognitive resource. These interventions frequently include changing the performance standard (e.g., number of targets to process) or changing the environmental complexity (e.g., terrain detail). Whereas environmental complexity should impact primarily sensory (visual) demands, performance standards are more likely to affect responding demands. A review of these manipulations should reveal the practical limitations of various cognitive resource channels for HRI tasks.

2.1.2 Purpose

Now that MRT and the common workload manipulations in HRI have been outlined, the purpose of this section is to draw guiding principles for teleoperator[*] workload and performance. A qualitative review will allow us to compare the effects of distinct workload manipulations across a variety of tasks and study criteria. To analyze the literature, a systematic coding process was applied to the extant database, described next.

2.2 Method

2.2.1 Literature Search

The literature search included a query using several scientific and military electronic databases, including the Defense Technical Information Center (DTIC), the Association for Computing Machinery (ACM), and the Institute of Electrical and Electronics Engineers (IEEE). References from a recent HRI review (Chen et al., 2007), as well as obtained experimental studies, were also checked for eligibility. Finally, a hand search was conducted on the following journals and proceedings for the past 5 years: *Human Factors*, *Presence*, *Human Computer Interaction (HCI)*, and *IEEE*.

2.2.2 Coding Procedure and Inclusion Criteria

Before coding, raters reviewed the variables of interest, constructed a coding sheet to reflect them, and accordingly screened articles for eligibility. Five studies were then selected and coded by all raters to examine validity and agreement. Based on acceptable agreement, one out of five raters coded the studies for this review based upon the definitions described in the following paragraph.

---

[*]The word "teleoperator" is broadly defined here and refers to an individual operating a device from a remote location.

To be included in the present review, an article was required to report a study that experimentally compared operator performance between different workload conditions. Furthermore, tasks had to utilize artificial agents or involve teleoperation. Thus, studies that used equipment for non-HRI tasks (e.g., cockpit simulators) were excluded from this review. Criteria included measures of (1) production (e.g., number of actions), (2) errors (e.g., incorrect actions), (3) RT, (4) efficiency (e.g., time to task completion), (5) perceived workload (e.g., the National Aeronautics and Space Administration Task Load Index [NASA-TLX] scores), and (6) situational awareness (SA). Finally, study characteristics such as the design (e.g., repeated measures), sample (e.g., student), task, and apparatus (e.g., UAV) were noted during coding.

2.3 Results

Table 1 lists the citations for the 18 studies assessing multirobot control, the number and type of platform used, the measured task outcomes, and key findings. In general, samples ranged from students to aviation and HRI professionals. Tasks predominantly included navigating platforms to targets or areas of interest, executing an action (e.g., inspection, manipulation), and monitoring and responding to system gauges and alerts.

When examining results by the task performance measures, we observe an emerging trade-off between production and other measures. In many studies, teleoperators could execute more total actions as they controlled more platforms (e.g., Crandall and Cummings, 2007; Lif et al., 2007; Squire et al., 2006). However, increasing the number of platforms also increased ERs in targeting and navigation (e.g., Dixon and Wickens, 2003; Galster et al., 2006), and it tended to increase RTs (e.g., Chadwick, 2006; Levinthal and Wickens, 2006). These results suggest that the control of multiple platforms allows the teleoperator to accomplish more tasks overall because of the increased resources. However, this added productivity comes at a cost of accuracy and efficiency. Although the control of one robot was optimal for task errors and RT across studies, the control of two robots did not inhibit performance to nearly the same degree as control of four or more robots (Adams, 2009; Chadwick, 2006; Ruff et al., 2002). Thus, control of two platforms might provide an optimal fit for maximizing both speeded performances and ER.

Finally, automation and multimodal feedback were examined as methods of improving the cognitive workload from additional platforms. In the case of automation, reliability made a much greater impact than the degree or type of automation (Levinthal and Wickens, 2006; Ruff et al., 2004). The addition of audio feedback, on the other hand, provided a consistently more positive effect (Wickens et al., 2003; Dixon and Wickens, 2003).

Table 2 presents the manipulation and the task affected as well as key findings for the 17 studies examining task demands. The types of devices used had more variability in this sample than in multirobot samples, including a robotic arm interface (Park and Woldstad, 2000), a decision-making simulation (Hendy et al., 1997), and virtual environments (VEs) from a variety of perspectives.

Table 1. Study summaries on multirobot control.

Study	Manipulation	Criteria (by Task Type)	Key Findings
Adams, 2009	One, two, or four UGVs	No. of actions, efficiency, and workload for search and transfer	• Slight differences between one and two UGVs, but efficiency and perceived workload were worse with four robots.
Chadwick, 2005	One or two UGVs	Targeting errors, navigation errors, and perceived workload	• No significant differences between groups.
Chadwick, 2006	One, two, or four UGVs	RT to hit target, RT to correct navigational error	• Response times degraded slightly from one to two UGVs. • Response times degraded markedly from two to four UGVs.
Chen et al., 2008	One or three UGV and/or UAVs	Errors, efficiency, SA, and workload in targeting (with navigation)	• Targeting errors were equal between three platforms and single UAV or UGV, but perceived workload and efficiency suffered.
Crandall and Cummings, 2007	Two, four, six, or eight UGVs	Errors and efficiency in navigation and target detection/transfer	• Four and two UGV conditions exhibited fewest lost robots. • Six and eight UGV conditions yielded highest no. of target successes.
Dixon and Wickens, 2003	One or two UAVs	Tracking error, target reporting accuracy, RT to system alerts	• One UAV user had slightly better performance indices than two UAVs. • Adding auditory feedback improved performance across conditions.
Galster et al., 2006	Four, six, or eight UAVs	Targeting accuracy, time processing key targets, RT to probes, workload	• Four UAV users had better accuracy and RT, but equal processing times. • Workload differences between conditions emerged for difficulty.
Humphrey et al., 2007	Six or nine UGVs	Efficiency, workload, and SA in bomb disabling simulation	• No. of platforms also coincided with no. of bombs to diffuse (difficulty). • Performance and workload indices were similar between conditions.
Levinthal and Wickens, 2006	Two or four UAVs	Idle time during UAV navigation, RT to system alerts	• Users were less efficient when controlling four UAVs. • False alarms in automation hurt performance more than false misses.
Lif et al., 2007	One, two, or three UGVs	Number of waypoints reached within given time (production)	• Users visited more waypoints controlling two or three UGVs (equally) than controlling one.
Murray, 1995	One, two, or three sensors	Time to monitoring task completion	• Users were significantly slower completing the tracking task with three platforms than with one.
Parasuraman et al., 2005	Four or eight UGVs	Completion time for game, no. of games won, workload	• Completion time and win rate deteriorated from four to eight UGVs. • As workload increased, automation features had a greater impact.
Ruff et al., 2002	One, two, or four UAVs	Targeting accuracy, correct rejection rate of automation errors, workload	• One UAV user had the fewest rejection errors, two UAV users had the best targeting accuracy, and four UAV users reported the most workload.
Ruff et al., 2004	Two or four UAVs	Targeting and navigation completion, RT to system alerts, workload	• All performance indices were better in two UAV conditions than four. • Reliability of automation, rather than level of automation, had greatest impact.
Squire et al., 2006	Four, six, or eight UAVs	Total number of actions executed (production)	• Users performed increasingly more actions with more platforms.
Trouvain and Wolf, 2003	Two, four, or eight UGVs	No. of inspections, no. of idle robots per second, time delay per inspection, workload	• Users performed more overall inspections with four and eight UGVs, but also had more idling time and efficiency loss.
Trouvain et al., 2003	One, two, or four UGVs	Time to navigation task completion, deviation from optimal path (errors)	• Users of one UGV had optimal navigation performance. • Two and four UGV users were equal in performance.
Wickens et al., 2003	One or two UAVs	Tracking error, system failure RT and errors, targeting time and errors	• One UAV user demonstrated faster reaction and targeting times. • Errors in tracking and system failure detections were equivalent.

Table 2. Summary of studies manipulating single task demands.

Study	Manipulation	Criteria (by Task Type)	Key Findings
Chen and Joyner, 2009	Dense or sparse targeting area	Targeting errors	• Errors increased with more distractor objects around the target. • In difficult conditions, manual control outperformed semi-autonomy.
Cosenzo et al., 2006	No. of targets to photo w/ UAV	Errors in targeting, RT to navigational decisions	• As no. of targets increased, targeting errors and RT to navigational stimuli increased.
Darken and Cervik, 1999	Ocean or urban environment	Efficiency in navigation	• Users had stronger performance in visually sparse ocean environments than in complex urban environments, regardless of the type of camera.
Draper et al., 1991	No. of alerts needing responses	Errors and RT in responding to UAV alerts	• Performance degraded as system alerts were more frequent; no interaction between condition and form of responses (manual vs. verbal).
Folds and Gerth, 1994	Dense or sparse targeting area	RT to identify new threat in virtual tracking task	• RT to emerging threat was slower in dense environment. • Auditory warnings improved RT more so in dense environments.
Galster et al., 2006	No. of targets to process	Errors, efficiency, and workload in processing targets; RT to probes	• Workload differences emerged favoring the low target condition. • Four UAVs yielded better performance with more targets than six or eight UAVs.
Hardin and Goodrich, 2009	200 or 400 distractor targets	Efficiency and errors in VE search and rescue	• No. of distracters had a significant effect on efficiency but not on errors; introducing autonomy did not mitigate this impact.
Hendy et al., 1997	Low, medium, or high time pressure	Efficiency, error, and workload in air traffic control	• Performance dropped only at high levels of time pressure. • Workload indices increased sharply beyond low time pressure.
Mosier et al., 2007	Low or high levels of time pressure	Errors and efficiency in diagnosing system problem in flight simulator	• Adding time pressure increased pilot efficiency but also increased diagnosis errors; this was worsened by system information conflicts.
Murray, 1995	Complex or simple images	Efficiency in monitoring and tracking targets in VE	• Increasing image complexity increased target detection time. • Automated mobility improved user performance in complex conditions.
Park and Wolstad, 2000	Size of destination for placement	Efficiency and workload in object transfer with robotic arm	• Less efficiency and higher workload in conditions with smaller targets. • 3-D displays helped performance with small targets.
Schipani, 2003	Navigation distance	Workload ratings in VE navigation	• Workload increased with greater distance to travel. • Line of sight with the operator did not impact workload.
Sellner et al., 2006	Simple or complex images	Efficiency and errors on task decision making (on stimuli)	• Simple displays decreased decision time but also increased errors. • Integrative presentations reduced the time penalty in complex displays.
Watson et al., 2003	Distance in 3-D placement	Errors, efficiency, and usability on virtual object placement (helmet-mounted display [HMD])	• Placement errors increased with greater distances in addition to task completion time; poor frame rate worsened this effect.
Witmer and Kline, 1998 (two studies)	Dense or sparse environment	Errors in distance estimation for VE	• More complex environments did not impact virtual distance estimation.
Yeh and Wickens, 2001	Dense or sparse environment	Errors, workload, and trust on target detection	• Users had better performance with low (vs. high) environmental detail. • With reliably cued targets, the impact of visual detail was reduced.
Yi et al., 2006	No. of targets to photo	Errors and SA in targeting with UAV	• Accuracy and SA decreased with more mission targets. • Workload conditions were not counter-balanced for practice effects.

Increasing task difficulty generally decreased a variety of performance indices across a variety of tasks. This would suggest that task demands are not criterion-dependent, as with control of multiple platforms. Based on this review, HRI task performance is particularly susceptible to strains on visual resources. This is evidenced by several relationships reported in studies. First, users had better performance in visually sparse or simple environments (e.g., Chen and Joyner, 2009; Darken and Cervik, 1999). Second, studies that manipulated visual features to mitigate workload reported a positive impact from their interventions (e.g., Park and Woldstad, 2000; Yeh and Wickens, 2001). Third, as visual demands were increased, audio feedback tended to improve operator performance (e.g., Folds and Gerth, 1994). Because HRI tasks are limited to interface and camera views, the visual channel will inherently receive greater strain than most other resource channels. Based on the evidence presented here, one may remove these demands by reducing visual information (e.g., using integrative displays or lower environmental detail) or by offloading information to other sensory channels (e.g., tactile, auditory).

2.4 Conclusions

The purpose of the current section was to examine the available research and determine guiding principles for managing workload in HRI. Specifically, this section examined manipulations of robot number and task demands separately, highlighting results by task and criteria. Results indicated that control of multiple platforms increases user productivity to the detriment of RT, accuracy, and workload. Results from manipulations of task demands suggested that visual strains are the primary limitation to teleoperator performance.

Results of this section yield several guiding principles for managing workload in teleoperators. First, the benefit from controlling multiple platforms should be explicitly weighed against the deterioration of other performance indices. Researchers and practitioners need to determine which criterion is more critical to task success, which may vary according to the situation. For example, overall productivity may be the critical outcome for search-and-rescue operations, whereas teleoperators disabling explosives are more likely concerned with correct actions. Second, workload from multirobot management may be alleviated through the introduction of practical and reliable automation, and attention management may be facilitated by audio alerts. Results from task demand manipulations suggest that HRI tasks tend to strain visual resources, such that increasing visual demands subsequently increases workload and reduces performance. We recommend that researchers and practitioners consider and limit these demands. Different approaches can reduce these visual demands, including a change in the display type and/or the use of other sensory channels to provide task feedback (e.g., use of audio or tactile cues).

The primary limitation of this section is that it does not provide a quantitative assessment or meta-analysis of the examined relationships. Although a quantitative review is desirable, existing studies are few in number, inconsistent in operational definitions, and lack needed statistics to permit a meta-analysis at this time. The HRI literature would also benefit from further investigations of workload mitigation, such as the use of multimodal feedback and/or automation. Existing studies in this area are promising but too few in number to provide a complete understanding of the advantages and disadvantages of these strategies. In conclusion, the purpose of this section was to guide future research by synthesizing the existing literature on teleoperator workload for a range of HRI tasks and criteria.

3. Autonomy and Automation Reliability in Human-Robot Interaction: A Qualitative Review

The effectiveness and reliability of automation aids are critical topics in the area of HRI. As more tasks are subsumed by robots and autonomous systems, it is important to examine the relationships between these entities and their human operators. Research to date has covered various manipulations of autonomy, but this broad body of research needs focus and consistency. The current study presents a qualitative overview of research regarding levels and reliability of autonomy/control and the effects they have on important HRI-relevant outcome variables. Results indicate that autonomy and automation aids operate uniquely for different tasks, and that there are many complex factors that can affect not only performance but also usability, confidence, and safety. Unresolved issues in the field and challenges and opportunities for future research are also presented.

3.1 Introduction

Robots and automated systems are now intertwined more than ever in our everyday lives. Robot and automated system operators often interact with these tools as they would with human coworkers. As we move toward more seamless and transparent interactions between humans and robotic entities, it becomes increasingly important to understand how these human operators and systems can optimally perform with the help of automation technology.

One purpose of increased automation is to lower the operator workload by taking on additional tasks without prompting the operator for commands. Empirical research in the areas of HRI and automated systems, however, has discovered more complex relationships between the human operator, the automated agent, and performance. The majority of research falls into two broad categories: level of autonomy/control (LOA) and automation aid reliability. Research on levels of autonomy/control focuses on investigating outcomes when the balance of control between human and autonomous agent is manipulated. Cueing and automation reliability research focuses on manipulating the accuracy and frequency of automation aids in the control of robots or complex semi-autonomous systems.

3.1.1 Levels of Autonomy/Control (LOA)

In many applications, human control of complex systems has been slowly replaced by robots and automated systems. Advances in technology increasingly allow human operators to simply observe a process or be minimally involved through safety checks or a simple button press. While technologies and automation have fully replaced humans in many tasks, a multitude of situations still exist in which humans and semi-autonomous systems or robots must work together. In some instances this cooperation stems from a lack of technology to fully subsume a human operator's role (e.g., air-traffic control). In other situations, an autonomous system is technologically capable of fully performing a task, but legal or safety restrictions exist that require a human operator (e.g., hazardous materials handling).

Research in LOA focuses on manipulating either the amount of control a human operator has over an automatic process or the amount of autonomy a robotic entity or system has from a human operator. Existing research in this area falls in one of two general design categories: human teleoperation of one or more robots and human supervision and control of semi-autonomous systems.

Researchers have long noted that the most common implementation of automation in an applied setting involves allocating as much responsibility to an automated system as is technologically possible (Kaber et al., 2000). If multiple tasks can be automated and supervised by a single operator, having a separate employee perform each task is impractical. The resulting consequence is that operators can only observe the process without any system interaction. They are left essentially "out of the loop." Since most automation is inherently imperfect, failures of automation or unsuccessful collaboration can lead to performance decrements worse than if the operator was completing the task without the use of any autonomous aid (Endsley and Kaber, 1999; Muthard and Wickens, 2003).

3.1.2 Automated Aid Reliability

While research on LOA tends to focus on system-level automation, automation does not always occur in every aspect of a given task. Much research exists exploring the use of automated aids and decision-making support systems that augment and assist a human operator–controlled task.

Automation aids typically are used to alert a human to important information that is either necessary for task completion or helpful in completing a task more efficiently or effectively. Some aids simply present the user with raw information in a more salient form, such as an auditory warning (Wickens et al., 2003). Other automated aids are more sophisticated and aggregate different sources of information to make a recommendation or alert to the user by way of complex computer algorithms (Wickens et al., 2005). Existing research in this area falls in one of three general design categories: production systems, targeting tasks, and diagnostics monitoring.

More complex aids aggregate raw data and present recommendations or alerts to operators. For these types of aids, imperfect calculations can lead to misleading information or incorrect decisions. These automation imperfections can take the form of either false alarms or misses (Dixon and Wickens, 2006). While these automation imperfections can be attributed to a myriad of causes (e.g., low-quality video feed, raw data inaccuracy), they are commonly associated with thresholds set in the decision-making computer algorithms that calculate the raw data and produce alerts and cues. In many cases, these thresholds can be adjusted to make an automated aid more or less prone to false alarms or misses (Levinthal and Wickens, 2006; Yeh and Wickens, 2001).

3.1.3 Purpose

The purpose of this section is to explicate the literature on LOA and automated aid reliability as it relates to HRI. Specifically, this study examines the trends present in these related streams of research to date and provides guidelines for future integrative research. From a practical perspective, this investigation seeks to spur critical thinking in settings where these technologies are used with the hope that improvements in the design, performance, and usability of robots and other autonomous systems will result. Unlike a traditional research design or meta-analytical investigation, this study aims to qualitatively integrate the dispersed research on these topics in an effort to encourage more standardized empirical investigation so that future quantitative meta-analyses are a feasible option for aggregating the data.

3.2 Method

3.2.1 Literature Search

The literature search included a thorough exploration of published studies, conference proceedings, and technical reports from a variety of scientific and military electronic databases, including ACM, DTIC, and IEEE. References from a comprehensive HRI review (Chen et al., 2007) as well as obtained studies were also checked for eligibility. Finally, a hand search was conducted on the following journals and conference proceedings for the past 5 years: *Human Factors, Human-Robot Interaction, Human Computer Interaction, Presence,* and *IEEE*.

3.2.2 Inclusion Criteria and Procedure

To be included in the present review, an article was required to report a study that experimentally investigated different levels of control/autonomy present in an autonomous system or robotic control, or explored the reliability and accuracy of cuing or decision-making aids in these scenarios.

3.2.2.1 Criteria. In the HRI literature, researchers have measured user effectiveness and general performance in a myriad of ways. In this section, the most common operations were selected for examination. In order to be eligible for inclusion, a study had to include at least one of the following criteria: ER, efficiency, RT, SA, or perceived workload.

3.2.2.2 Study Coding. Before coding, raters reviewed the variables of interest, constructed a coding sheet to reflect them, and used it to screen for article eligibility. Five studies identified for eligibility were then selected and coded by all raters to examine validity and agreement. Based on acceptable agreement, one out of five raters coded each study on the following six dimensions: (1) article characteristics, (2) sample characteristics, (3) research design, (4) independent variables, (5) task type and apparatus, and (6) outcome measured.

3.2.2.3 Analyses. As research accumulates in this area, meta-analytic methods may be applied to assess the quantitative impact LOA and automated aid reliability have on performance outcomes. Existing studies, however, are few in number and inconsistent in the operations of study variables. As a result, the present analyses consist of qualitative descriptive summaries of the obtained articles.

3.3 Results

Table 3 presents the studies included for analysis with regards to LOA. Included in the table is a brief summary of the specific manipulation (IV-independent variable) and task design, criterion (DV-dependent variable) measurement, and guiding principles for each study. Table 4 presents the same summarized information for studies included for analysis covering automated aid reliability.

Immediately evident in the qualitative analysis of the included studies is the dependency of results on the experimental task employed in each study. For example, in some tasks such as search and rescue, automation led to improvements in performance across the board (Luck et al., 2006). In other tasks such as an air-traffic controlling scenario, the effect of automation is more complex and performance benefits vary (Endsley and Kaber, 1999). Similarly, automated aid reliability research reveals that for some tasks, imperfect automation leads to large performance decrements (Rovira et al., 2007), while for other tasks it leads to a reliance on other strategies to successfully complete a task with only marginal effects on performance (Meyer et al., 2003).

Table 3. Summary of studies examining LOA.

Study	Manipulation and Design (IV)	Criteria Measurement (DV)	Key Findings
1	Manual robot control vs. shared control with robot navigating and operator focused on target ID.	Performance (no. of targets correctly identified)	For novice robot operators, performance is increased with the use of a semi-autonomous (shared control) navigation aid.
4	Ten LOAs in monitoring, generating, selecting, and implementing between human operator and automated system.	Performance (no. of points earned in targeting simulation, missed targets, and collisions)	LOAs that combine human generation of options and automated implementation produce superior results during normal system operations; joint decision making (human/system collaboration) is detrimental to performance.
5	User-controlled vs. sensor-driven control of secondary independent UGV camera.	Performance (no. of targets identified, time spent in visual target inspection)	Sensor-driven control is better; automatic gaze redirection of a UGV camera helps in close-up identification of objects in a search task.
6	Five LOAs and five schedules of automation (automation on then off for a specified time) for system control.	Performance (no. of errors, errors in secondary task), workload, SA	When automation is cycled on and off, performance is best when the human operator develops a strategy that is implemented automatically; workload correlated with secondary task performance.
7	Five LOAs range from simple support to full automation.	Performance (errors, efficiency), workload, SA	Increased automation leads to performance improvements and reduces human operator subjective workload, but also reduces SA for some system functions.
9	No aid, veto-only aid (stop to avoid damage), or semi-autonomous aid (adjusts course away from obstacles) UGV control.	Usability	Users may struggle to adapt strategies around autonomous agent control, and steering/navigation trouble may arise if the operator is unable to adjust.
11	LOA and latency for search-and-rescue UGV.	Performance (errors, time to completion), usability	Increased automation leads to performance improvements in both errors and time. It also acts as a buffer from the negative effects of control latency.
17	LOA for team of three UGVs: full autonomy, mixed control, full control.	Performance (targets identified), behavior, usability	When controlling multiple UGVs, a mixed control paradigm with both manual control of robots as well as some cooperative automation provided best performance; controllers who switched attention between robots more frequently performed better in manual and mixed control scenarios.
18	Single or dual UAV control with no aid, auditory aid, or flight path tracking automation.	Performance (errors, efficiency, RT)	Automation aid helped improve target identification tasks more when operating multiple UAVs vs. single UAV control.

Table 4. Summary of studies examining automated aid reliability.

Study	Manipulation and Design (IV)	Criteria Measurement (DV)	Key Findings
3	UAV targeting task with automation aid of varying accuracy and reliability.	Performance (errors, response time), SA	False alarm–prone automation leads to a decreased use of aids and ignoring raw data; imperfect automation leads to better detection of a target miss; high workload polarizes these effects.
7	Five LOAs range from simple support to full automation; normal operation or unexpected automation failure.	Performance (errors, time to recovery), workload, SA	Increased automation leads to performance improvements and reduces human operator subjective workload; in automation failure, lower-level LOAs with more human control results in the best performance due to increased SA.
10	Search simulation with two or four UAVs controlled with no automation aid, 90% reliable aid, 60% reliable aid prone to false alarms, or 60% reliable aid prone to target misses.	Performance (efficiency, RT)	There is a substantial cost to efficiency as users control more UAVs; automation aids that provide more false alarms are more detrimental to user performance than 90% reliable or 60% reliable automation aids emphasizing misses.
12	Automatic cuing agent for quality control decision-making task: none, low, or high validity; high vs. low overall automation.	Performance (errors)	Higher levels of automation resulted in more reliance on cues; no performance differences between automation types for valid cues, but lower automation outperformed higher automation with less valid cues.
14	Flight simulation with or without reliable attention guidance automation.	Performance (errors, efficiency), subjective confidence/trust	When automation of flight plan selection is used, pilots were more likely to ignore changes in the environment making the flight unsafe after selection; automation is best in selection, but not necessarily implementation/monitoring.
15	Command and control targeting simulation with various levels of information and/or decision-making automation.	Performance (accuracy, RT), workload, subjective confidence/trust	Imperfect information automation and decision-making automation are both detrimental to performance; major component of failures is the lack of operator access to raw information and complacency.
16	One, two, or four UAVs controlled manually or with 95% or 100% accurate automated or by-consent decision-making aid.	Performance (errors, efficiency)	Management-by-consent automation aid resulted in best performance as it left operators “in the loop” but was scalable to increases in workload (more UAVs).
19	UAV simulation with automated diagnostics information: 100% accurate, 60% reliable with false alarms, 60% reliable with misses, or manual control.	Performance (errors, efficiency), behaviors	Increased misses by automation leads to decrease in concurrent task performance driven by reallocation of visual attention while increased false alarms led to slower response to all automation alarms and were followed by more time scanning the environment (raw data) to determine accuracy vs. 100% accurate alarms.
20	UAV targeting simulation with automated 75% or 100% reliable cuing for some targets.	Performance (accuracy), workload, subjective confidence/trust	Partially reliable cuing increases false alarms and eliminates overall performance benefits of cuing; cuing draws attention toward cued target results in other targets being overlooked.

3.3.1 Overall Analysis of LOA

It is clear that some amount of automation does in fact increase overall performance for primary tasks. This is true for novice robot operators (e.g., Hughes and Lewis, 2005) and UGV and UAV operators (e.g., Wang and Lewis, 2007), as well as in targeting simulations (Kaber and Endsley, 2003). In at least some conditions, automation can lead to significant problems, especially if the operator is unable to access raw data (Rovira et al., 2007) or does not know how to regain control of a robot (Krotkov et al., 1996).

While the notion that all technology available be utilized in an automated system seems sensible, our analyses found a trend toward the opposite. One-third of included studies utilized a version of Endsley and Kaber's (1999) 10-level LOA taxonomy. This taxonomy separates tasks into four roles: monitoring, generating, selecting, and implementing. Each level in the taxonomy assigns either a human operator, a computer (autonomous agent), or both to control each role. Results of studies using the taxonomy all indicate that performance is optimal when the human operator generates potential actions and selects the desired action; it is then automatically implemented by the system (e.g., Kaber and Endsley, 2003). In these scenarios, an increase in task or process automation reduces subjective workload and SA of the operator (Kaber et al., 2000).

3.3.2 Overall Analysis of Automated Aid Reliability

Across all included studies, the reliability and accuracy of automated aids has a significant effect on performance. Automation with a high tendency for false alarms results in the greatest detriment to performance. Operators experiencing automated aids with a high level of false alarms tend to use and respond to aids less frequently and tend to ignore raw data in targeting tasks (Dixon and Wickens, 2006). In a scenario when operators were required to respond to imperfect automated diagnostic aids, responses were slower to all automation aids if false alarms were common, and raw data was used more often, reducing overall efficiency (Wickens et al., 2005). This is in contrast to the Dixon finding, which may be attributable to the false alarms, although more research would be welcome to clarify the apparent differences. When raw data is not available to the operator in imperfect automation (e.g., false alarm prone) conditions, complacency led to further decreases in performance (Rovira et al., 2007). In nearly all cases, when workload was increased, the overall detrimental effects of imperfect automation were polarized (e.g., Levinthal and Wickens, 2006).

Imperfect automation aids also influence performance through the reallocation of attention. This can occur in several ways, the simplest being when an incorrectly activated alert or cued target is attended to by an operator while an actual target or event goes unnoticed (e.g., Yeh and Wickens, 2001). Additionally, automation can lead operators to ignore raw data for a portion of a task that has become automated (Muthard and Wickens, 2003), essentially assuring a problematic situation will arise should automation fail. In line with the findings of LOA research, automated aid reliability research fully supports the notion that access to raw data and avoidance of situations where operators are "out of the loop" are critical to performance (e.g., Ruff et al., 2002).

3.4. Discussion

3.4.1 Guiding Principles

When the included studies are looked at as a whole, some general guiding principles arise from the current status of research in both LOA and automated aid reliability. The most important message that current research sends is that technology should not be utilized simply because it is available. Until the relationships between human operators and a given technology are understood as they relate to performance, the application of that technology to work should be limited to making recommendations. This review of research also sheds light on the fact that there are many forms of automation, all of which have unique effects on different dimensions of performance, and operate in a different way for varying tasks. This is an important practical implication and highlights the need for careful application of research findings to real-world contexts. Last, keeping operators "in the loop" with access to all available data is imperative to successful interactions. Until automation is perfect, a human operator will always need to know how to recover successfully from failure by completing a task the old-fashioned way, without any help from defective automation.

3.4.2 Unresolved Issues and Future Directions

While research on LOA and automated aid reliability has covered many important issues surrounding the interaction of humans and autonomous systems and agents, there is room for more investigation. An area that has been largely overlooked in current streams of research is the differences in the experience levels of operators. Whether they are UAV pilots or quality-control supervisors, current research has largely ignored the fact that experience may play a large role in the interactions operators have with automation. Some research has looked at novice operators (Bruemmer et al., 2004), but empirical investigations comparing novices to experienced operators are needed. For example, a novice operator will likely respond very differently to an automation failure than an experienced employee who knows the background processes behind the automation.

Keeping operators "in the loop" with the task they are completing is another important determinant of performance in many scenarios. Research on interface design could greatly facilitate this by investigating display interfaces that aggregate data and present automation aids, but also give operators intuitive access to raw data should they need it. An existing problem with operators who do have access to raw data is the additional workload associated with accessing it. If the information was easily available and intuitively connected to the related automation within an interface, these two problems may be resolved.

Last, as technology allows, adaptive automation schemes should be investigated as a potential buffer to the effects of different operators or tasks. These systems could alter their own LOA based on output performance or operator responses to automation aids. For example, in a semi-autonomous quality-control system, performance data could be fed back into the system, which could then alter the LOA. If a given operator is experienced and performs better with more control of the system, he or she could then be granted more control. On the other hand, a novice operator might benefit from either higher levels of automation when output efficiency is important or from low levels of automation for training purposes. Similarly, autonomous systems or agents might be able to predict failures and correct them before the human operator is even aware of a problem. Researchers must stay one step ahead of the application of new technologies in order to investigate how best to apply the advances in practical settings.

3.4.3 Summary and Conclusion

The primary finding in this study is the general lack of quantitative analysis data in the fields of LOA and automated aid reliability. This is mainly a consequence of both the limited number of available empirical investigations and the extreme diversity in variable operations and measurement in the existing literature. For example, ER is measured in numerous ways, including points acquired, targets identified, and collisions avoided. While these data inform us about the task-specific relationships they examine independently, they cannot be sensibly integrated by traditional meta-analytical means. This discovery brings to light the need for consistency and cooperation among research in these areas. More general investigations are needed that can be flexibly applied to more tasks (Miller and Parasuraman, 2003), and common methods must be agreed upon so that the findings can be better utilized by a wider audience in practice.

The present study's analysis of LOA and automated aid reliability was born of a larger investigation of HRI. For this reason, the literature search and resulting studies focus only on these topics as they relate to HRI. The benefit of this methodology is an in-depth focus on the topics as they apply to HRI. While the consequences may be the exclusion of some important non-HRI work in the areas of LOA and automated aid reliability, this focus exemplifies the need for consistency in these areas of research.

As technology and automation processes continue to alter the way people interact with each other and machines, researchers must clarify how best to use these modern advances. Common sense may dictate that we use whatever technology is available, but careful investigations of the application of automation are important to guarantee optimal use of these complex and often expensive tools.

4. Effectiveness of Visual Devices in Human-Robot Interaction: A Qualitative Review

Visual devices employed in HRI have matured from mere tools to integrated technological extensions of the body. Research must follow suit and strive to create a programmatic framework to address the lack of common operations in previous literature. The current qualitative analysis organizes five commonly manipulated visual modality devices into three categories based on MRT (Wickens, 1980, 2002). This endeavor synthesizes existing research to create a needed foundation for future research. Analytic results suggest that employing robot-enhanced visual systems aids operators' task performance when the visual device is matched with the operators' task. More importantly, results indicate that our current understanding of visual modalities is rudimentary and full of caveats.

4.1 Introduction

HRI exemplifies the use of technology as a "force multiplier," in military terms, which increases the physical and mental abilities of operators beyond what was previously feasible. This allows operators to outsource cognitively taxing but predictable tasks to interactive technology systems. Research has yet to reach a methodological and technical consensus, however, on how to maximize HRI given quickly advancing technology, nor has it achieved a systematic, coherent approach to studying visual modalities. The purpose of this qualitative review is to provide a summary, organized by a proposed framework, of the current state of HRI visual modality research. This review will also highlight inconsistencies among variable manipulations and operations in an effort to guide future research. The need for a more systematic research agenda should not be dismissed given technology's ever-growing pervasiveness in military and civilian life.

The present review identifies common themes within HRI literature addressing technology's enhancement of visual perception. According to Wickens' (1980) MRT, some tasks can be performed in parallel while others cannot due to mental workload constraints. Notably, tasks requiring different perceptual resources (e.g., simultaneously performing an auditory and visual task) can typically be performed together, whereas two tasks straining the same modality will mutually interfere with task performance (e.g., performing two visual tasks simultaneously). Perceptual and cognitive overload due to the latter can occur within and between modalities. The present study applies MRT solely to visual modalities and perception. Resources can be constrained by a variety of factors, including time, cognitive processing, and contextual factors. The framework proposed in this study is built upon these three resource constraints (Wickens, 2002).

Current manipulations of visual modalities in the HRI literature can be organized into five categories: FR, latency/time delay, SS and MS visual cues, FOV, and camera perspective. This review organizes these five manipulations into three conceptual dimensions (figure 1): time resources, contextual resources, and visual processing resources.

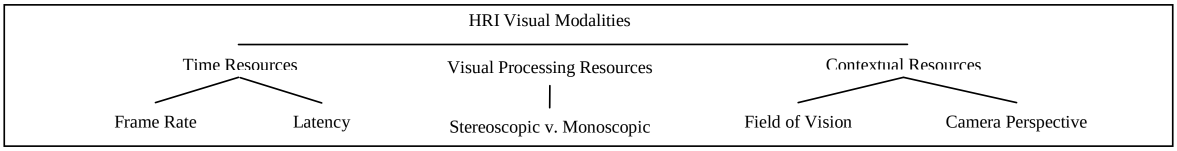

Figure 1. Visual modality.

4.1.1 Time Resources

Time resources, including picture latency/time delay and FR, describe an approach in which time-related HRI system features are altered. Such alterations affect an operator's ability to visually integrate multiple screen views over time. For example, Luck and colleagues (2006) manipulated two forms of time resources: the time delay between a camera display and its operator's teleoperation of a UGV along with whether the latency was variable or consistent over trials.

FR and latency are frequently addressed simultaneously by experimental methodology or operationalized as dependent system responsiveness features (Chen and Thropp, 2007; Darken et al., 2003). Latency, or time delay, refers to the temporal discrepancy between an actual event and when the event is viewed on a screen. FR is defined as the number of screen shots displayed over time or the image refresh rate of a system (typically measured as frames per second).

4.1.2 Contextual Resources

Contextual resources include manipulations of FOV and camera perspective in which the information given by the environmental perspective is changed to holistically alter the extent to which operators are able to visually perceive their surroundings. Thus, the operator's visible range of sight is physically altered via the grounding and/or positioning of a map or camera view. For example, Darken and Cervik (1999) manipulated a virtual map to either orient "up" as north or in the direction of forward movement.

FOV describes the physical dimensions of the operator's visual screen view. A typical manipulation contrasts a wide-panoramic perspective with a narrow perspective. Camera perspective is characterized by the immersion level of the camera in reference to a target object. Manipulations often compare a third-person, or exocentric, camera perspective with a first-person, or egocentric, camera perspective. The latter would be a fully immersed viewpoint. For tasks, such as in a UAV, which allow for three axes of movement (e.g., left-right/yaw, forward-backward/roll, up-down/pitch), perspective also refers to whether the camera view is gravity- or vehicle-based. External visual manipulations are especially sensitive to effects of cognitive processing such that peripheral vision and direct vision result in unique visual interpretations and the SA of an environment.

4.1.3 Visual Processing Resources

HRI studies comparing SS to MS visual cues make use of visual processing resources, serving as a conceptual bridge between time and contextual resource. Cuing, in this case, is dependent both on nuanced time manipulations (e.g., differential processing of latency for SS and MS conditions) and on contextual information provided by the presence or lack of normative binocular depth cues in SS and MS views, respectively.

MS visual displays consist of a 2-D image presented to both eyes that provides visual cues like object size, shadows, and the interposition of objects (Draper et al., 1991). SS visual displays present a 3-D image representation to both eyes allowing for greater perceived realism and, importantly for cognitive processing, retinal disparity. Retinal disparity, as in typical viewing conditions, allows for richer visual cues, complex depth cues, and enhanced visual acuity. Based on Wickens' (2002) description of visual channel resources, MS displays capitalize on peripheral vision perceptual resources, whereas SS primarily employs focal vision perceptual resources.

4.2 Method

We conducted a literature search via several methods to create a comprehensive HRI database. First, published studies, conference proceedings, and technical reports were obtained via a search of several scientific and military electronic databases, including DTIC, ACM, and IEEE. References from an HRI review (Chen et al., 2007) as well as obtained studies were checked for eligibility. Finally, a hand search was performed on the following journals and conference proceedings for the past 5 years: *Human Factors*, *Presence*, *Human Computer Interaction*, and *Institute of Electrical and Electronics Engineers*.

4.2.1 Definitions and Inclusion Criteria

4.2.1.1 Independent Variables. To be included in the present review, an article was required to report a study that experimentally investigated visual modality manipulations, specifically system latency, FR, FOV, camera perspective, SS vision, or MS vision. Studies that failed to satisfy these dimensions were not included in this analysis.

4.2.1.2 Criteria. Within HRI literature, user performance criteria are often defined inconsistently, hindering between-study comparisons of outcomes. In this section, criteria operations most frequently measured were selected for analysis. A study had to include at least one of the following criteria to be eligible for inclusion: ER, RT, efficiency, SA, or the operator's perceived workload. ER was defined as the number or percentage of incorrect responses (or, if reverse coded, the percentage of correct responses) or as a measure of the task's deviation from an optimal path or solution. Efficiency was coded as the time taken to complete a task, or as a measure of production within a standardized unit of time. RT represents the elapsed time between the presentation of a stimulus to the fitting response by an operator. For the sake of distinction, efficiency assesses the overall task completion time, whereas RT measures focus on a speeded response to specific stimuli of interest (e.g., an alarming cue). SA describes the level of task and contextual knowledge an operator has. Finally, perceived workload reflects self-report measures of the operator's experienced cognitive demands, often measured by the NASA-TLX.

4.2.1.3 Study Characteristics. Characteristics were coded that may affect overall study results because they provide insight on the study design's strength and fidelity to the operators' tasks and environment. For study design, counterbalancing or random assignment were noted for repeated measures and between-group studies. Study fidelity characteristics included the sample population (e.g., military, student, gender, mean age), the type of apparatus used (e.g., high- or low-fidelity simulator), and the type of task(s) being performed by users/operators. Apparatus included actual or simulated UAVs and UGVs, flight simulators, helmet-mounted displays (HMDs), VEs, or simple computer interfaces/simulations. Task type was coded according to the types of functions asked of operators by the experiment. Task category examples include robot navigation, teleoperated manipulation of objects, or targeting critical objects/stimuli on the interface (e.g., point and click).

4.2.2 Procedure

Prior to coding studies deemed relevant in the literature search, raters reviewed the variables of interest, constructed a coding sheet to reflect them, and used it to screen for article eligibility. Of the studies deemed eligible, five were selected and coded by all five raters to assess inter-rater reliability. Acceptable rater agreement was found. Subsequent studies were each coded by one rater on the following six dimensions: (1) article characteristics, (2) sample characteristics, (3) research design, (4) independent variables, (5) task type and apparatus used, and (6) the type of outcome measured.

Analyses consisted of descriptive summaries of the obtained articles since too few reported statistics appropriate for a meta-analytic review. These summaries include information on the workload manipulation and study design, study characteristics, the type of task and apparatus, relationships with dependent variables, and a summary of results, or guiding principles. Studies within articles were coded separately if independent samples were used in each (e.g., one article with three reported studies using independent samples was coded as studies a, b, and c).

4.3 Results and Discussion

This section created a framework around which to organize the various manipulations of visual modalities, based on Wickens' MRT (1980, 2002). In the present analysis, 10 studies manipulated FR (table 5), 7 examined latency/time delay (table 6), 7 compared MS to SS visual cues (table 7), 10 assessed FOV (table 8), and 11 studied camera perspective (table 9). A few studies were included in multiple categories (e.g., Lion [1993] was included in both FR and SS vs. MS cues) since they examined more than one visual modality (Lion; Reddy, 1997; Scribner and Gombash, 1998; Van Erp and Padmos, 2003; Watson et al., 2003). Notably, several of the studies reported under the contextual resource theme share considerable overlap in their conceptualization of FOV and perspective. This analysis conceptualized SS and MS comparisons as a hybrid of contextual and time resources, in many cases. As such, some ambiguity is acknowledged regarding how various studies were organized.

Table 5. Summary of studies manipulating FR.

Study	Manipulation	Criteria	Task Type	Results
Calhoun et al., 2006	FR (update rates)	Performance, SA, usability, workload	Operator-controlled UAV	With higher update rates, SA increased, usability decreased, workload decreased, and performance increased. Objective performance ratings showed no difference between FR conditions.
Darken and Cervik, 1999	FR	Errors, SA, usability	Navigation of building with camera view	No significant differences found between FR video conditions; no significant learning effects.
Lion, 1993	FR (33 or 22 Hz), head motion, MS vs. SS	Performance, errors	Tracking task using 3-D computer interface	Higher FR related to better performance; performance learning effects present.
Massimino and Sheridan, 1994	FR (3, 5, 30 fps), presence of force feedback	Efficiency	Operator-controlled mechanical arm via camera view	Increased FR significantly improved efficiency; the addition of force feedback improved efficiency for all FR conditions.
Reddy, 1997	Study A: FR (2.3, 11.5 Hz), Study B: FR (6.7, 14.2 Hz), FOV	Errors, efficiency	Navigation task in VE	Errors and efficiency decreased with lower FR.
Richard et al., 1996	FR, SS, and MS vision	Efficiency	Track and grasp 3-D moving target	Higher FR coupled with MS compensated for a lack of SS visual cues; learning effects were significant.
Van Erp and Padmos, 2003	FR, spatial resolution of image	Errors, efficiency, SA	Driving task	Higher FR was related with improved performance for all criteria; no significant learning effects.
Watson et al., 1998	Studies A, B, C: FR (9, 13, 17 Hz)	Efficiency, errors, RT, usability	VE track and grasp of object using an HMD	With lower FR, RT increased, usability decreased, and efficiency was reduced; errors were not significantly affected.
Watson et al., 2003	Studies A and B: FR, task difficulty	Errors, efficiency, usability	Operator used an HMD to complete tasks	Efficiency decreased, and errors and task difficulty increased as FR decreased.
Chen et al., 2008	FR (“normal,” “reduced”)	Errors, efficiency, usability, workload, motion sickness, performance	Simulated navigation/targeting using UAVs and UGVs	No significant differences between FR conditions; for UGVs, performance (hit rates) decreased with reduced FR; no performance differences for UAV.

Table 6. Summary of studies manipulating latency/time delay.

Study	Manipulation	Criteria	Task Type	Results
Adelstein et al., 2003	Latency, constant or random head motion rates	RT	Observed VE through HMD	Only interactions were significant—changes in motion patterns resulted in a decrease in operators' discrimination abilities and latency detection.
Ellis et al., 2004	Latency detection, environmental complexity	Errors	Navigation of VE with an HMD	Complexity of environment failed to effect operator errors; learning effects reported.
Lane et al., 2002	Operator input and robot action time delay	Efficiency	Tracking and grabbing in UGV simulator	Increased time delays led to a decrease in efficiency.
Luck et al., 2006	Studies A and B: latency rates, variable and fixed latency lengths	Errors, efficiency, usability	Operator-navigated VE using UGV simulator	Increased latency/time delay led to a reduction in efficiency and more errors; efficiency improved when time delay was fixed as opposed to variable.
Shreik-Nainar et al., 2003	Constant or random time delay	Errors, efficiency	Navigation of VE with an HMD	When time delay was constant, as opposed to variable, errors increased and efficiency decreased.
Watson et al., 2003	Image latency, system responsiveness	Errors, efficiency	Completed with HMD in VE	Significant learning effects for impact of system latency.
Chen et al., 2008	Latency (250 ms vs. none)	Errors, efficiency, usability, workload, motion sickness, performance	Simulated navigation/targeting using UAVs and UGVs	No significant differences between presence or lack of latency; usability decreased with presence of latency.

Table 7. Summary of studies manipulating SS and MS visual cues.

Study	Manipulation	Criteria	Task Type	Results
Drascic and Grodski, 1993	SS vs. MS	Errors	Object manipulation using teleoperated robot arm	SS camera view significantly reduced ERs compared to MS views.
Draper et al., 1991	Studies A, B, and C: SS vs. MS	Errors, efficiency	Placement task using robot arm	No difference was present between MS and SS for low-difficulty tasks; significant differences were present for more difficult tasks such that an SS view resulted in greater efficiency.
Lion, 1993	SS vs. MS, FR, head motion	Performance, errors	3-D tracking task	SS display was significantly related to enhanced performance and a reduction in errors.
Park and Woldstad, 2000	Multiple 2-D vs. 3-D MS vs. 3-D SS	Errors, efficiency, workload	Placement task using robotic arm	No significant difference between 3-D MS and 3-D SS; 2-D display outperformed both 3-D displays.
Richard et al., 1996	Studies A and B: SS vs. MS, multimodal feedback type	Efficiency	Used haptic feedback glove in VE	When FR is high or other modality cues are present, SS does not present a significant advantage to MS; in baseline conditions, SS is more efficient than MS; significant learned effects were present.
Scribner and Gombash, 1998	SS vs. MS, FOV	Errors, efficiency, stress, usability	UAV driving task	SS resulted in fewer errors and reduced stress scores, and was preferred by users (usability) over MS.
Nielsen, Goodrich, and Ricks, 2007	Studies A and B: 2-D vs. 3-D across display types (map, video, map-video)	Errors, efficiency	UGV navigation (study A: simulation; study B: realistic)	Study A: no significant differences in completion time or errors; learning effects present. Study B: map-only display completed slower than map-video (2-D) and video-only (3-D).

Table 8. Summary of studies manipulating FOV.

Study	Manipulation	Criteria	Task Type	Results
Parasuraman et al., 2003	FOV and computer opponent strategy	Efficiency, workload	UGV navigation in VE	FOV showed no effects on criteria.
Parasuraman et al., 2005	FOV (three levels)	Efficiency, workload, SA	UGV navigation in VE	Workload increased as FOV decreased; no significant difference was present for efficiency.
Pazuchanics, 2006	Narrow vs. wide FOV	Efficiency, errors, usability	UGV navigation	Widening FOV resulted in improved performance compared to narrower FOV.
Reddy, 1997	Studies A and B: FOV, FR	Efficiency, errors	Navigation task in VE	Errors and efficiency were reduced with wider FOV.
Scribner and Gombash, 1998	FOV, SS vs. MS	Errors, efficiency, stress and motion sickness	UAV driving task	Motion sickness was reported more frequently in wide FOV condition; no significant interaction was present between FOV and MS/SS.
Smyth et al., 2001	FOV (three levels)	Errors, efficiency, workload, stress and motion sickness	UGV driving task	Wider FOV was desired for navigation but the FOV closest to typical vision was preferred for steering.
Smyth, 2002	FOV (three levels)	Errors, efficiency, workload, stress and motion sickness	UGV driving task	Indirect FOV resulted in decreased driving speed and more errors compared to the baseline natural vision condition.
Van Erp and Padmos, 2003	FOV (two levels)	Errors, efficiency, SA	Teleoperation of UGV	Improved performance due to wider FOV was task dependent and was not beneficial when the narrower FOV was sufficient for the task; wide FOV was significantly better when UGV made sharp turns.
Wang and Milgram, 2003	FOV (six levels)	Errors, SA	Teleoperation of UGV using various camera views	SA increased as FOV extended outward from robot; the moderate FOV condition provided the best local SA and ER.
Draper et al., 1991	Narrow vs. wide FOV	Efficiency, errors, usability	UGV simulated search task	Completion times (efficiency) were faster with a wider FOV; efficiency is incrementally improved when wide FOV and warning are present.

Table 9. Summary of studies manipulating camera perspective.

Study	Manipulation	Criteria	Task Type	Results
Darken and Cervik, 1999	Map direction orientation (north-up or forward-up)	Errors, efficiency	UGV driving task using camera/map view	Forward-up alignment was best for targeted search tasks, but north-up alignment elicited the best performance for naïve and primed search tasks.
Heath-Pastore, 1994	Gravity based vs. vehicle based	Errors	UGV simulator driving task	Operators reported greater confidence and SA for gravity-referenced view; gravity-based perspective had fewer errors than vehicle-based perspective.
Hughes, 2005	No. of cameras fixed or independent of UGV, camera alignment	Errors, usability	Search and navigation of robots; target ID	Operator-controlled cameras best for usability.
Lewis et al., 2003	Gravity based vs. vehicle based	Errors, efficiency, usability	Teleoperation of UGV	Efficiency and usability were significantly better for gravity-fixed display.
Murray, 1995	Fixed vs. mobile vehicle-based view	Efficiency	Target detection using camera views	Efficiency was reduced with mobile camera views vs. fixed-position cameras.
Olmos et al., 2000	2-D display vs. 3-D third-person display vs. 3-D split screen display	Error, efficiency, RT	Navigation of VR terrain in-flight simulator	2-D display was detrimental to vertical maneuver performance, 3-D display showed greatest deficits during lateral maneuvers; split screen, when displays were made visually consistent, was significantly different from 2-D and 3-D display.
Schipani, 2003	Line-of-sight view vs. non-line-of-sight	Workload	Navigation of UAV	No significant difference for workload was found between line of sight and non line of sight.
Thomas and Wickens, 2000	Third-person vs. first-person view	Errors, RT, usability	Simulated teleoperation of robot	Third-person view demonstrated faster RT and fewer errors, and operators reported higher levels of confidence (usability) compared to the first-person view.
Draper et al., 1991	Camera view vs. camera view inlaid in larger virtual display	Efficiency, errors, usability	UGV simulated search task	Reported usability reduced when camera perspective is inlaid (picture-in-picture display) VE display.
Nielsen, Goodrich, and Ricks, 2007	Studies A and B: display type and 2-D/3-D comparison: video-only vs. map-only vs. video-map display	Errors, efficiency	UGV navigation (study A: simulation; study B: realistic)	Study A: video-only display resulted in most errors and slowest completion time; no differences between map-only and map + video, learning effects present.Study B: for 2-D displays, more errors present for video-only vs. map-only display.
Drury et al., 2007	Map-centric vs. video-centric display	Errors, efficiency, SA, usability	Simulated UGV search and navigation task	Video-centric display best for usability, movement efficacy, SA, surroundings awareness; map-centric best for location and status awareness.

4.3.1 Time Resources

System latency and FR were categorized as time resource HRI system features. Fourteen studies reported in 10 articles address HRI FR manipulations. Of these studies, 11 measured efficiency and errors, 7 usability, and 3 SA. Two examined workload, and three measured task performance. Overall findings suggest that higher FR (e.g., more frames per second) increases efficiency, reduces errors, and improves usability, among other criteria.

System latency/time delay was manipulated in eight studies within seven articles. Six of these studies measured errors and efficiency. Usability was assessed by two studies, and RT by one. Findings suggest that increased time delays between an operating system and its operator result in decreased efficiency and increased ER. All but one of the studies examining fixed latency vs. variable delays reported that fixed latency delays ameliorate operator efficiency and ER.

The HRI literature on FR and latency frequently made use of the terms interchangeably or inconsistently. Conceptual ambiguities were resolved based upon a study's task and criteria. Generally, higher FR and decreased latencies benefitted user performance. These results are consistent with the notion that a more realistic image will result in less discrepancy between typical visual processing and visual processing of technologically altered stimuli. Frequently, a consistent FR was used throughout studies. Though methodologically consistent, this approach lacks external validity because FR does vary within and across HRI tasks (e.g., Darken et al., 2003). Thus, experimental studies of FR often require less of the operator's attention since conditions are predictable. Operator awareness was also of issue for latency studies. Several studies reported that either learning yielded significant increases in performance criteria and/or pretask awareness training mitigated the deleterious effects of latency on performance measures (Ellis et al., 2004; Watson et al., 2003). Future research should seek to create common operations of FR and latency, assess the specific effects of learning, and determine the threshold for cognitive processing of a realistic/real environment in contrast to a VE.

4.3.2 Contextual Resources

Contextual resources include FOV and camera perspective. FOV was examined in 11 studies within 10 articles; 10 measured efficiency, 9 looked at errors, 4 examined workload, 3 addressed SA, and 2 accounted for self-reported stress, motion-sickness, and usability. The results on FOV are mixed but do suggest a preference for a wide to moderate FOV over a narrow FOV.[*] When a wider FOV is introduced, Scribner and Gombash (1998) reported increased motion sickness rates.

Twelve studies addressed camera perspective, nine reported measures of error, seven assessed efficiency, five usability, two RT, and one workload. Results should be taken as starting points rather than principle, given the variety of manipulations. Overall, performance is maximized

[*]The reader is referred to the Scribner and Gombash (1998) manuscript for complete definitions.

when the camera perspective is an exocentric, third-person view of the environment and/or gravity-referenced (as opposed to being referenced toward the camera's physical direction of movement or tilt). Additionally, when a split-screen display is present (e.g., either a third-person perspective or 3-D image is viewed alongside a first-person or 2-D image, respectively), performance is maximized and is incrementally better to single perspective conditions (Olmos et al., 2000).

Despite a wide range of methodologies and manipulations, contextual resource study results all promote moderation (i.e., FOV within typical visual range) and integration (i.e., perspective and FOV presenting multiple visual displays). For example, when combined with another workload reduction task (such as increasing contextual information), an FOV manipulation allowing an operator to switch between manual and automated operating systems positively affected performance (Pazuchanics, 2006). This suggests that integrating contextual resources with other interface features can be a force multiplier. Relatedly, perspectives that provide either a third-person view or a stable, gravity-based orientation facilitate performance (e.g., Thomas and Wickens, 2000). Results underscored the utility of a user's natural spatial ability, in addition to learning effects, when it comes to increasing performance on criteria (e.g., Darken and Cervik, 1999).

4.3.3 Results for Visual Processing Resources

SS and MS visual cues were examined by 11 studies within 7 articles; 8 reported errors and efficiency while the other criteria—workload, usability, general performance, and self-reported stress—were each assessed within a study. A pattern emerged suggesting SS views are to be preferred over MS views with regard to efficiency and errors. Richard and colleagues (1996) found that when other modalities act as additive cues for the operator or visual conditions are optimal (e.g., high FR), SS is no better than MS.

The benefits of SS displays over MS displays were not overwhelming, as many researchers had hypothesized. In baseline conditions, the added realism and depth cues provided by SS displays did benefit operator performance. However, in the presence of other cues, such as auditory alerts, MS displays fared as well or slightly worse than SS displays. Notably, the small number of studies included in this category and the specificity of task manipulations may bias these preliminary findings. The age of these studies is also of interest. All but two studies were published in the 1990s. Beyond this, these studies lacked consistency among their task purposes and operator instructions. Several studies stress speed, for instance, over accuracy, and vice versa. Though overall results were inconsistent regarding the advantages of SS displays over MS displays, there is a consistent trend favoring SS over MS in high-difficulty situations requiring greater visual acuity. Thus, the advantages of each are highly contingent on the task difficulty and the presence of multimodal cueing.

4.4 Conclusions

This section's review suggests that realistic camera images (e.g., moderate FOV, high FR, low latency delay) are related to higher performance ratings for operators. Increased image realism allows operators to compensate for deficits in visual processing. Such deficits may be the result of reduced retinal disparity and depth cues or overtaxing of focal and peripheral visual processing resources. Our preliminary analysis supports the notion that using robot-enhanced visual systems aids overall task performance but also suggests that our understanding of these relationships is incomplete.

Limitations of this qualitative analysis primarily center on identifying the inclusion criteria, the potential for coding errors within each study, faulty translation of ambiguous terminology, and the small number of studies included within the five categories. A shortcoming of the current visual modality literature, and this analysis by extension, is the absence of a shared mental model within the literature. For example, "third-person camera perspective" was used interchangeably with an "exocentric perspective" or a "god's-eye view." Though study details suggest these terms are equivalent, task demands frequently were given priority over fidelity to the manipulation itself.

Another concern was the impact of learning effects beyond the study variables. Most studies took two approaches to learning effects, either (1) participants completed practice trials prior to a study's data collection to minimize effects or (2) the study included a measure of learning effects as part of the experiment. Beyond this, the majority of coded studies shared significant methodological constraints due to their samples, which were notably small and predominantly male.

As exemplified by the results of this qualitative analysis, researchers need to agree on a program of research, use a common framework to exact a shared mental model, and address numerous literature gaps at both theoretical and practical levels. The need to create and define a programmatic research agenda poses an imminent challenge to researchers. Utilizing the framework proposed in this section will allow for a systematic and theoretically founded means for future studies on visual modalities. In particular, latency/time delay and camera perspective warrant greater attention in order to create a more unified, less fragmented research agenda. Perhaps more confused than the manipulations are the criteria. Aside from reporting ER, no consensus exists for measuring other criteria. For example, within the area of FOV research, studying motion sickness and SA appears to be a fruitful research avenue, but these criteria are neglected in other categories of visual modality manipulations. Thus, both criteria and predictors deserve greater attention and consistency. Without the latter, few guiding principles will arise, and the utility of visual modalities will remain ambiguous in military and civilian operations.

5. Recommendations for Further Research

We reviewed work in the broad area of HRI and it is available as a RefWorks (2009) dataset. This report is more focused, presenting three issues specific to the development of principles for multimodal displays in HRI operations: workload, autonomy and automation, and visual display issues. Each topic has developed strengths and provided guidance to developers. However, further work remains to be done. As such, the following sections present issues targeted at further research in the area.

5.1 Meta-Analysis of Coded HRI Articles

We conducted a literature review to identify articles dealing with HRI topics. Hundreds of articles were screened for those that contained data suitable for meta-analysis. These articles were further coded using the coding scheme developed for the project. The coded articles are contained in an online RefWorks database. The articles were further organized according to the taxonomy of independent variables listed in section 4 of this report.

The next logical step is to further code the articles for a meta-analysis. This involves taking the data from the articles and computing common statistics for the meta-analysis. In certain instances, authors will need to be contacted to provide the requisite data. In addition to the statistical computations, it will be necessary to form appropriate theoretical or methodological groups of equivalent metrics and variables. The initial grouping has been completed and is presented in section 4 of this report. For instances where there are not enough studies for a meta-analysis, further grouping will be undertaken. This involves organizing and collapsing studies into groups that are theoretically similar. For example, when considering the studies of workload and controlling one or more robots, does it make sense to aggregate studies of more than one robot into an ordinal scale? Similar questions exist about the role of potential moderators and their relation to outcomes.

5.2 Cumulative Sum Methodology for Modeling Training Effectiveness/Skill Decay

Section 3 of this report contains an article describing the cumulative sum technique. Used in a variety of areas—most recently for medical error issues—it holds potential as a modeling technique when the purpose is monitoring the process of either acquiring skills and/or skill decay. It is flexible and tailorable, and may prove more effective providing feedback to warfighters training to criterion performance levels (e.g., TALON robot operations). We propose developing the technique as a training feedback system for use both in the United States and with deployed forces.

5.3 Focus Groups/Structured Interviews for Lessons Learned

Figure 2 illustrates some of the main themes that emerge from a review of the academic and scientific HRI literatures. What appears to be lacking, however, is a user perspective on important aspects of HRI for the warfighter. It would be useful to gather experienced users of one or more robot types (e.g., FCS technologies network, TALON, iRobot, PackBot) together for focus groups and/or structured interviews to extract lessons learned on in-theater robot operations. Issues to be addressed include those listed in figure 1 as well as others that emerge from the sessions. This information would be culled together into a report that could be used to guide further HRI research, systems design, and training that would be directed specifically to the needs of the warfighter.

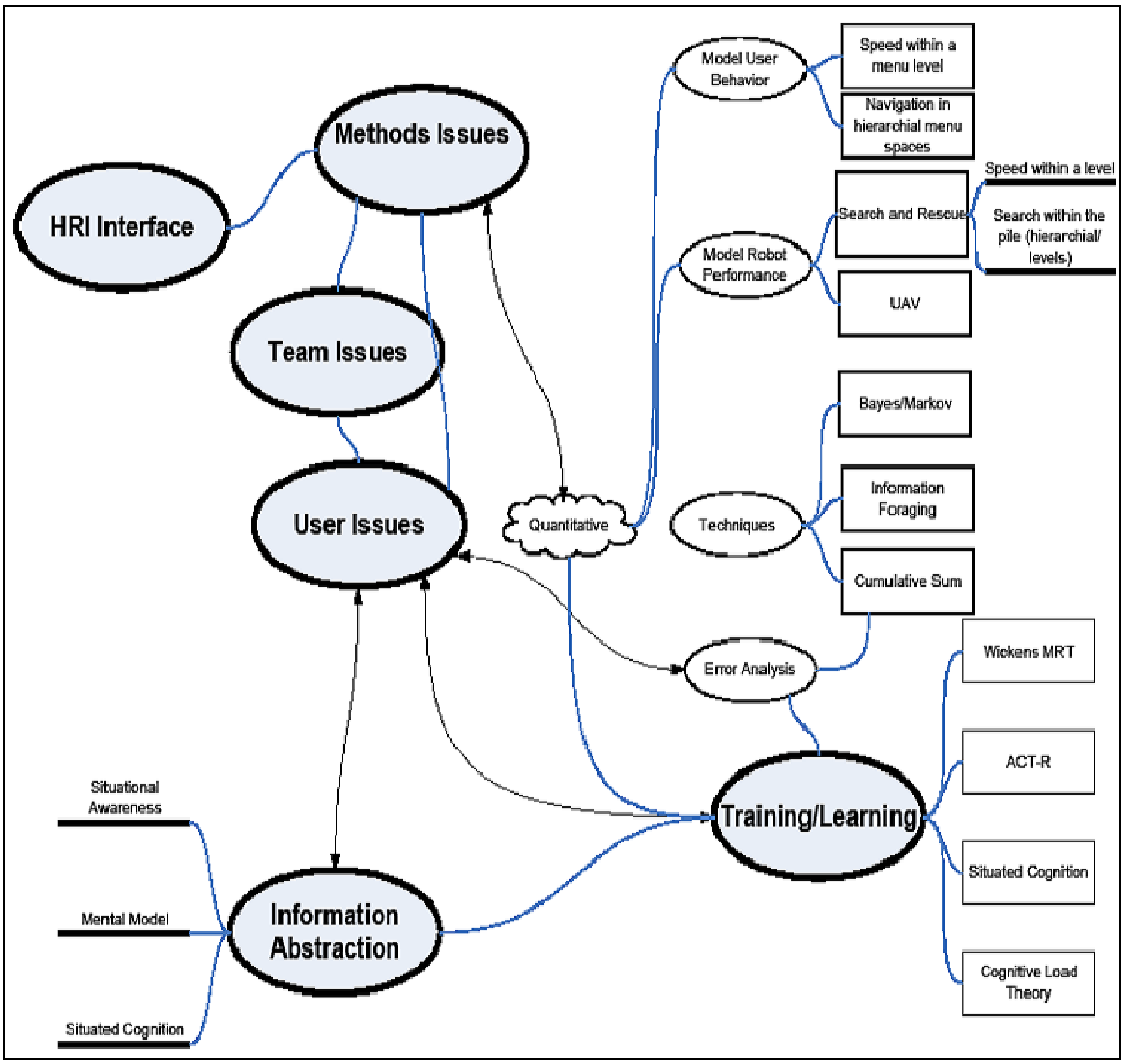

Figure 2. Model depicting active areas of HRI research.

5.4 Structural Equation Modeling of HRI Effectiveness

Structural equation modeling can yield systematic and reliable insight from accumulated data. For example, given a collection of focus group research, information can be collected on the experience and effectiveness of the operators who serve as participants. Experience could then be operationalized via such measures as the amount of time operating specific robot types, number of field exercises, number of missions, and so forth. Effectiveness is operationalized via objective data, when available, or by self-reports based on past missions. This data along with existing attitudinal, aptitude, and training data can be used to develop conceptual/theoretical models of warfighter HRI performance. These models are tested via structural equation modeling to determine the importance of different constructs on warfighter HRI performance.

5.5 Multilevel Data Structures and Models

Research in HRI is needed in many areas, such as workload modeling, team vs. individual operators for different numbers of robots, and the effectiveness of multisensory interfaces given differing task requirements. Many research issues studied thus far and reported in the literature have not employed a multilevel analysis even though the problems and data are hierarchically structured.

Many kinds of data have a hierarchical or clustered structure such that units at one level are nested within units at the next level. Examples include offspring grouped within families, students in classes, and individual workers in teams. For repeated measures data, the observations or measurements are nested within individuals. The lowest level of observation is called level 1 (e.g., offspring, students), followed by level 2 (e.g., families, classes), and so forth. Theoretically, there may be any number of levels to such a structure, but in practice, most empirical studies focus on 2- or 3-level data. Multilevel models are also known as hierarchical linear models, mixed models, and random coefficients models. For an introduction and/or complete treatment of the methodology, see Bryk and Raudenbush (2001), Goldstein (1995), Hox (1995), Kreft and de Leeuw (2000), and Snijders and Bosker (1999).

In multilevel model frameworks, variables are often measured at each level. The subscript i is used to represent unit i in level 1, and subscript j is used to represent unit j at level 2. An example is individual i in group j. The outcome (response) variable measured on each level-1 unit is designated y_{ij} and represents the measure for individual i in group j. Predictor variables at level 1 are designated x_{ij}, and at level 2 are designated z_j.

Analysis of multilevel data in the past has been performed in different ways but each has associated problems. These problems are as follows:

1. Disaggregating (total or pooled regression analysis): Regression is conducted on the full data analysis, and the unit of analysis is the individual. The general model is $y_{ij} = a + bx_{ij} + e_{ij}$, where "$a$" refers to the intercept, b is the beta weight, and e_{ij} is the error. The model can expand to accommodate level-2 variable predictors: $y_{ij} = a + bx_{ij} + cz_j + e_{ij}$, where c is

the beta weight for the level-2 variable z_j. Problems with this approach include severely biased estimates of effects (regression coefficients tending to be too large and standard errors too small) resulting in inflated type-1 ERs. Goldstein (1995) provides a full discussion of this and additional problems.

2. Aggregation: This approach is used to analyze the data at the level-2 unit, or group level in our example. The model for the data is $y_{.j} = a + bx_{.j} + e_j$, and with a level-2 predictor, $y_{.j} = a + bx_{.j} + cz_j + e_j$. The primary limitation of the aggregation approach is that it ignores all level-2 variation (within group in our example). The result is that no interpretations can be made about effects or relationships at the individual level. Attempted interpretations lead to the "ecological fallacy" (Robinson, 1950).

3. Analysis of covariance (ANCOVA): The ANCOVA model has the individual as the unit of analysis, and the independent variable is categorical with levels defined by the level-2 units (groups in our example). The level-1 predictor variable, x, is a covariate. The intent of the analysis is to test for an effect of level-2 units (groups) on *y*, after removing the effect of *x*. The model is $y_{ij} = a_j + bx_{ij} + e_{ij}$, and the coefficient *b* is assumed to be invariant across level-2 units, while the intercept a_j is allowed to vary across those level-2 units. Problems with this strategy include the unrealistic assumption (in most cases) of an invariant slope and the inability to incorporate level-2 variables *z* into ANCOVA.

4. Separate regressions: This approach is to conduct separate regression analyses within each level-2 unit, with the model stated as $y_{ij} = a_j + b_j x_{ij} + e_{ij}$. The procedure yields separate estimates for intercepts and slopes for each level-2 unit, which can be used to compute variability in the resulting estimates of a_j and b_j. These estimates can be used as dependent variables in second-state regressions, predicting these values from the level-2 variable, *z*: $a_j = c_0 + c_0 z_j + e_{aj}$ and $b_j = d_1 + d_1 z_j + e_{bj}$. Results indicate the degree to which variability in intercepts and slopes among level-2 units are predictable from *z*. While it is closer to taking into account the multilevel nature of the data, it is problematic in several ways, including high unreliability in level-2 slopes and intercepts, and no partitioning of the variance of *y* into within- vs. between-group portions. Also, it is impractical with a large number of level-2 units and involves the estimation of a large number of parameters (Goldstein, 1995; Kreft and de Leeuw, 2000).

5.5.1 The Linear Multilevel Model

A basic assumption of the multilevel model is that there exists a population of units at each level, and we obtain a random sample from each population. For example, if our level-2 variable is schools, we assume a population of schools and we obtain a random sample of schools. A level-1 variable is a population of students within each school, and we obtain a sample of students from within each school. The outcome variable y_{ij} is the score for outcome variable *y* for level 1, unit *i*, from level 2, unit *j*. An example is a reading test score for student *i* from school *j*. For level-1 predictors, x_{ij} is the score on predictor *x* for level 1, unit *i*, from level 2, unit *j*.

For example, x_{ij} might represent the number of hours spent reading each week for student i from school j. For level-2 predictors, z_j is the score on predictor z for level 2, unit j, e.g., a binary indicator of public (1) vs. private (0) for school j.

The level-1 model specifies the outcome variable as a linear function of level-1 predictors. In the previous example, the reading test score is expressed as a linear function of the number of hours spent reading each week: $y_{ij} = a_j + b_j x_{ij} + e_{ij}$, where a_j is the intercept for level 2, unit j, and it is the predicted value of y_{ij} when $x_{ij} = 0$. b_j is the slope for level 2, unit j, and represents the predicted change in y_{ij} for a 1-unit increase in x_{ij}. In our example, it is the predicted change in reading test scores for student i in school j when the number of hours spent reading increases by one unit. The slopes and intercepts vary across level-2 units. The residual, e_{ij}, is that portion of y_{ij} not accounted for by x_{ij}. Level 1 has residual variance, $Va\ (e_{ij}) = \sigma^2$, and represents the variance in the outcome variable not accounted for by level-1 predictors. In our example, this is the variance in reading scores not accounted for by the number of hours spent reading.

The level-2 model specifies level-1 random coefficients as a linear function of level-2 predictor(s). For example, level-1 slopes and intercepts in equations for predicting reading test scores from the number of hours spent reading are represented as linear functions of public- vs. private- school indicators. The level-2 model for level-1 intercepts is $a_j = \gamma_{00} + \gamma_{01} z_j + u_{0j}$, where γ_{00} and γ_{01} are intercepts and slopes in the model. γ_{00} is the predicted value of a_j when $zj = 0$. γ_{01} is the predicted change in a_j for a one-unit increase in z_j. These are fixed coefficients. In our example are the intercept and slope for model-predicting level-1 intercepts in the reading test—the number of hours reading relationship from public-vs.-private schools. The residual term in the model is expressed by u_{0j} and represents that portion of a_j not accounted for by z_j. The variance of the residual is $Var(u_{oj}) = \tau_{00}$. The level-2 model for level-1 slopes is expressed as $b_j = \gamma_{10} + \gamma_{11} z_j + u_{1j}$, where γ_{10} and γ_{11} are the intercept and slope in the model. The term γ_{10} is the predicted value for b_j when $z_j = 0$. The predicted change in b_j for a one-unit increase in z_j is expressed as γ_{11}. These are fixed coefficients, and in our example are the intercept and slope for predicting slopes in the level-1 reading test—the relationship between the number of hours reading for public vs. private schools. The residual term is u_{1j} and represents the portion of b_j not accounted for by z_j. The variance of the residuals is expressed as $Var(u_{1j}) = \tau_{11}$. The residuals in the two equations have a covariance, $COV\ (u_{0j}, u_{1j}) = \tau_{01}$.

Through substitution from the level-2 equation for a_j and b_j into the level-1 model for y_{ij}, the overall model for y_{ij} is given as

$$y_{ij} = a_j + b_j x_{ij} + e_{ij}\,, \tag{1}$$

$$y_{ij} = (\gamma_{00} + \gamma_{01} z_j + u_{0j}) + (\gamma_{10} + \gamma_{11} z_j + u_{1j})\, x_{ij} + e_{ij}\,, \tag{2}$$

and

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j + \gamma_{11} z_j x_{ij} + u_{1j} x_{ij} + u_{0j} + e_{ij}\,. \tag{3}$$

It is important to note two aspects that set this equation 3 apart from a conventional regression model. The first is the presence of an interaction between a level-1and a level-2 predictor. This important characteristic allows for the modeling of cross-level interactions. Second are the distinct error terms for different levels and aspects of the model. This takes into proper account the multilevel data structure in defining these errors.

This basic model is easily extended to include multiple level-1 and level-2 predictors through the addition of another subscript on the *x*'s and *z*'s as well as their coefficients.

1. Parameter estimation: The previous model can be fit to a hierarchically structured sample of observations. The model implies a moment structure (means, variances, covariances) of measured variables that are expressed as a function of a set of parameters: level-1 residual variance (σ^2), level-2 fixed coefficients (γ_{00}, γ_{01}, γ_{10}, γ_{11}), and level-2 residual variances and covariances (τ_{00}, τ_{11}, τ_{01}).

2. Model specification: Models will be specified in order to address our research questions. Within each question, a series of models ranging from simple to complex will be specified, first introducing level-1 predictors and then introducing level-2 predictors. This general strategy is recommended by Snijders and Bosker (1999) and Bryk and Raudenbush (1992). Comparison of residual variances between models indicates the variance accounted for by each set of predictors.

5.5.2 HRI Research

Consider the following research question: To what extent is the warfighter's performance with a robot influenced by the (1) warfighter's personal attributes, (2) robot's attributes, and (3) team performance factors.

This can be expressed in a general equation form:

Warfighter and robot performance	=	Warfighter personal attributes	+	Robot attributes	+	Team level factors	. (4)
Outcome variable		Level-1 variables		Level-1 variables		Level-2 variables	

The first set of analyses focuses on predicting the warfighter's and robot's performance from two sets of level-1 predictors (warfighter's personal attributes [e.g., experience], robot's attributes [e.g., interface type], and one level-2 predictor [team level factor]). A series of models ranging from simple to complex will be examined, and a comparison of the residual variances between models will indicate the amount of variance accounted for by each set of predictors (Snijders and Bosker, 1999; Bryk and Raudenbush, 1992). Beginning with the warfighter's personal attributes variables, variables will be entered as a set, and those significantly related to the warfighter and robot performance would be retained. Following this, the robot attributes variables will be entered as a set, and those significantly related to warfighter and robot performance would be retained. Each model would have a deviance statistic reflecting the residual variance. The deviance statistic allows for a comparison of the remaining variance from each model and facilitates model comparisons. The team level variables result in a continuum of performance. This allows for an examination of cross-level interactions between the team level factors (level 2) and the warfighter attributes (level 1) as well as the robot attributes (level 1). These models will be statistically compared via the deviance statistic.

5.5.3 Multilevel Models of Repeated Measures

Another series of analyses focuses on the dynamic relationship that exists in HRI operations. In this repeated measured data, the sample of individuals is measured on repeated occasions. These data can be viewed as having a multilevel structure, where level-1 units are trials and level-2 units are missions. Trails are nested within procedures, and procedures are nested within individuals (level 3). Within the multilevel framework there is no need for the measurements to be the same for each individual. Also, the spacing and number of occasions can vary across individuals.

The basic two-level model is easily extended to accommodate level-3 units (Bryk and Raudenbush, 2001; Goldstein, 1995; Kreft and de Leeuw, 2000; Snijders and Bosker, 1999).

Level-3 predictor measures are all variables that, up to this point, have been considered level-1 predictors (e.g., warfighter's personal attributes, robot attributes, team factors). Following the strategy just described, the level-1 predictors will be examined one at a time, followed by the level-2 predictor procedure, then the level-3 predictors. Models from least to most complex will be evaluated and compared with the deviance statistic

5.6 HRI Operators Field of Vision

Manipulating the type of information that is provided by visual modalities and available to operators of TALON robots in military operations will provide increased knowledge on mechanisms to maximize performance criteria. Key performance criteria to assess include user ERs, efficiency, workload, motion sickness, and general operator usability. The following paragraphs discuss three future research projects that center around manipulations of camera perspective's FOV, environmental immersion, and multimodal feedback.

Previous studies on manipulating an operator's FOV have reported enhanced performance when a wide or "natural" FOV was present compared to a narrow FOV (Parasuraman et al., 2005; Scribner and Gombash, 1998; Smyth, 2002; Smyth et al., 2001; Van Erp and Padmos, 2003). In light of these findings, it is hypothesized that performance will improve (e.g., reduced errors and workload, enhanced efficiency and SA) as the FOV provided by the robot-mounted camera perspective increases. In this experiment, FOV would be manipulated at three levels: narrow, "natural visual range," and wide. In addition, we would measure operators' subjective reports of motion sickness. Scribner and Gombash found that though a wider FOV increases performance, it also is associated with greater incidents of motion sickness. Given the deleterious effects of motion sickness on operators, understanding the advantages and disadvantages of varying levels of FOV is needed.

For operators to effectively navigate a terrain with a teleoperated robot, a robot-mounted camera's perspective is an important consideration. Previous research suggests that an exocentric, or third person, perspective compared to an endocentric, or first person, perspective enhances performance and subjective reports of usability (Olmos et al., 2000; Thomas and Wickens, 2000). Additionally, Olmos et al. found that a split screen display, which enables the operator immediate access to both perspectives, is better than a single perspective. Given the aforementioned findings, it is hypothesized that an exocentric perspective, compared to an endocentric TALON-mounted camera perspective, will decrease errors and increase efficiency and usability. Additionally, a single camera screen shot will be less effective than a split-screen shot of both perspectives. Thus, three conditions will be present: an exocentric perspective only, an endocentric perspective only, and a split-screen endo- and exocentric camera perspective. All conditions will employ 3-D camera displays given that previous studies have reported the benefits of 3-D over 2-D perspectives due to the increased depth cues available, realism, and SA (Drascic and Grodski, 1993; Olmos et al., 2000; Scribner and Gombash, 1998).

Last, another research avenue to pursue involves comparing a single visual modality condition to a multimodal feedback condition. Park and Woldstad (2000) found that the addition of another modality cue alleviated differences between visual perspective conditions that were present when only visual modalities were compared. Thus, a study with conditions comparing visual feedback only (camera view), visual and force feedback (vibrotactile feedback provided as operator control TALON using joystick), and visual and audio cuing alerts would assess whether the simple addition of another sensory cue could improve operator performance without a significant change to the robot's camera perspective.

5.7 Robot Autonomy and Errors

TALON robot operators are an ideal population to research further in the areas of LOA and automated aid and cuing reliability. Manipulating the amount of control an operator has over a TALON robot or the amount of autonomy the robot or teams of robots have is a simple investigation that could yield information directly applicable to the battlefield.

Existing research on LOA supports the notion that increased automation generally leads to increases in performance (Wickens et al., 2003); however, boundary conditions are being identified. For example, Chen and Joyner (2009) state that for nonprimary tasks in a multitasking environment, degradations of human performance are often observed. Furthermore, Chen and Terrence (2009) provide evidence that certain individual differences such as attentional control play a role in operator's interaction with automated systems. Unfortunately, the robots and systems used in this research vary greatly, and most results have been task- or equipment-specific. Additionally, having access to raw data (e.g., a video feed rather than just collision sensors) and changes in workload (e.g., controlling one vs. two or more robots) affected performance differently in the various equipment and scenarios used to date (Rovira et al., 2007). Given the wide array of performance effects indicated by previous research, it is imperative that these factors be examined as they directly apply to TALON operators and the unique demands of the environments they operate within. A future research project in this area would be limited by the technologies capable of being integrated into the TALON systems. If Endsley and Kaber's (1999) 10-level LOA taxonomy were used, monitoring, generating, selecting, and implementing action could be manipulated in a search and rescue scenario. Based on previous research, it is hypothesized that performance will be optimal when the human operator generates potential actions and selects the desired action; that desired action is then automatically implemented (Kaber and Endsley, 2003). Other potential variables of interest in this investigation are the operator's experience level and the number of robots controlled. Important criteria in addition to performance are SA, subjective workload, and general usability.

Previous research on automation aid reliability has found considerable deleterious effects on performance when aids and cues used in the control of robots are inaccurate or not dependable (Wickens et al., 2003). Like LOA research, existing studies on automation aid reliability cover a wide gamut of scenarios and equipment, for which results have been mixed. Overall, automation aids prone to false alarms appear to have the most significant negative impact on performance across various tasks (Dixon and Wickens, 2006). Imperfect automation can also affect performance by incorrectly allocating attention to a noncritical target resulting in a true target or an enemy going unnoticed. The effect of automated aid reliability on TALON operator performance should be investigated to discover whether the same problems arise when aids are imperfect. A simple navigating and targeting scenario would be used with an automated target detection aid to draw the operators' attention to potential targets. The reliability of the aid would be manipulated as well as the proportion of false alarms vs. misses. Based on previous research, it is hypothesized that more targets will be identified under the more reliable aid and that performance will be the worst in the high false alarm condition. Recent research is identifying situations in which individual operator differences interact with false-alarm-prone and miss-prone automated differences. See Chen and Terrance (2009) for these and other findings.

6. References

Adams, J. J. Multiple Robot/Single Human Interaction: Effects on Perceived Workload. *Behavior Information and Technology* **2009**, *28* (2), 183–198.

Adelstein, B. D.; Thomas, G. L.; Ellis, S. R. Head Tracking Latency in Virtual Environments: Psychophysics and a Model. *Proceedings of Human Factors and Ergonomics Society 47th Annual Meeting*, Santa Monica, CA, 2003; pp 2083–2087.

Bosker, R. J.; Snijders, T. A. B.; Guldemond, H. *Power in Two-Level Designs: Estimating Standard Errors of Regression Coefficients in Hierarchical Linear Models for Power Calculations. User's Manual Version 1.6*; University of Groningen: Groningen, The Netherlands, 1999.

Bruemmer, D. J.; Boring, R. L.; Few, D. A.; Marble, J. L.; Walton, M. C. I Call Shotgun! An Evaluation of Mixed-Initiative Control for Novice Users of a Search and Rescue Robot. In *Proceedings of the 2004 IEEE International Conference on Systems, Man, and Cybernetics*, The Hague, The Netherlands, 2004.

Bryk, A. S.; Raudenbush, S. W. *Hierarchical Linear Models*, 2nd ed.; Sage: Newbury Park, CA, 1992.

Calhoun, G.; Draper, M.; Nelson, J.; Lefebvre, A.; Ruff, H. *Simulation Assessment of Synthetic Vision Concepts for UAV Operations*; AFRL-HE-WP-TP-2006-0038; U.S. Air Force Research Laboratory: Wright-Patterson Air Force Base, OH, 2006.

Chadwick, R. A. The Impacts of Multiple Robots and Display Views: An Urban Search and Rescue Simulation. *Proceedings of the Human Factors and Ergonomics Society's 49th Annual Meeting, HFES*, Orlando, FL, 2005; pp 387–391.

Chadwick, R. A. Operating Multiple Semi-Autonomous Robots: Monitoring, Responding, Detecting. *Proceedings of the Human Factors and Ergonomics Society's 50th Annual Meeting, HFES*, San Francisco, CA, 2006; pp 329–333.

Chen, J. Y. C.; Thropp, J. E. Review of Low Frame Rate Effects on Human Performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* **2007**, *37* (6), 1063–1076.

Chen, J. Y. C.; Joyner, C. Concurrent Performance of Gunner's and Robotics Operator's Tasks in a Multi-Tasking Environment. *Military Psychology* **2009**, *21*, 98–113.

Chen, J.; Terrance, P. Effects of Imperfect Automation and Individual Differences on Concurrent Performance of Military and Robotics Tasks in a Simulated Multitasking Environment. *Ergonomics* **2009**, *52*, 907–920.

Chen, J. Y. C.; Haas, E.; Barnes, M. Human Performance Issues and User Interface Design for Teleoperated Robots. *IEEE Transactions on Systems, Man, and Cybernetics* **2007**, *37*, 1231–1247.

Chen, J. Y. C.; Durlach, P. J.; Sloan, J. A.; Bowers, L. D. Human-Robot Interaction in the Context of Simulated Route Reconnaissance Missions. *Military Psychology* **2008**, *20*, 135–149.

Cosenzo, K. A.; Parasuraman, R.; Novak, A.; Barnes, M. Implementation of Automation for Control of Robotic Systems; ARL-TR-3808; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2006.

Crandall, J. W.; Cummings, M. L. Developing Performance Metrics for the Supervisory Control of Multiple Robots. *Proceedings of the Second ACM/IEEE International Conference on Human-Robot Interaction, HRI*, Arlington, VA, 2007; pp 33–40.

Darken, R. P.; Cervik, H. Map Usage in Virtual Environments: Orientation Issues. *Proceedings of IEEE Conference on Virtual Reality*, Houston, TX, 1999; pp 133–140.

Darken, R. P.; Kempster, K.; Peterson, B. Effects of Streaming Video Quality of Service on Spatial Comprehension in a Reconnaissance Task. *Proceedings of I/ITSEC 2001*, Orlando, FL, 2003.

Dixon, S. R.; Wickens, C. D. Control of Multiple-UAVs: A Workload Analysis. *Proceedings of the 12th International Symposium on Aviation Psychology*, Dayton, OH, 2003; pp 1–5.

Dixon, S. R.; Wickens, C. D. Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors* **2006**, *48*, 474–486.

Draper, J. V.; Handel, S.; Hood, C. C.; Kring, C. T. Three Experiments With Stereoscopic Television: When it Works and Why. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, Charlottsville, VA, 1991; pp 1047–1052.

Drascic, D.; Grodski, J. Defense Teleoperation and Stereoscopic Video. *Proceedings of SPIE: Stereoscopic Displays and Applications IV*, Bellingham, WA, 1993.

Drury, J. L.; Keyes, B.; Yanco, H. A. LASSOing HRI: Analyzing Situation Awareness in Map-Centric and Video-Centric Interfaces. *Proceedings of the ACM/IEEE International Conference on Human-Robotic Interaction*, Arlington, VA, 2007; pp 279–286.

Ellis, S. R.; Mania, K.; Adelstein, B. D.; Hill, M. Generalizeability of Latency Detection in a Variety of Virtual Environments. *Proceedings of Human Factors and Ergonomics Society's 48th Annual Meetings*, Santa Monica, CA, 2004.

Endsley, M. R.; Kaber, D. B. Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. *Ergonomics* **1999**, *42* (3), 462–492.

Folds, D. J.; Gerth, J. Auditory Monitoring of up to Eight Simultaneous Sources. *Proceedings of the Human Factors and Ergonomics Society's 38th Annual Meeting, HFES*, Santa Monica, CA, 1994; pp 505–509.

Galster, S. M.; Knott, B. A.; Brown, R. D. Managing Multiple UAVs: Are We Asking the Right Questions? *Proceedings of the Human Factors and Ergonomics Society's 50th Annual Meeting, HFES*, San Francisco, CA, 2006; pp 545–549.

Goldstein, H. *Multilevel Statistical Models*; Halsted: New York, 1995.

Hardin, B.; Goodrich, M. On Using Mixed-Initiative Control: A Perspective for Managing Large-Scale Robot Teams. *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, HRI*, La Jolia, CA, 2009; pp 165–172.

Heath-Pastore, T. *Improved Operator Awareness of Teleoperated Land Vehicle Attitude*; Report # NCCOSC 1659; U.S. Naval Command, Control and Ocean Surveillance Center: San Diego, CA, 1994.

Hendy, K.; Lao, J.; Milgram, P. Combining Time and Intensity Effects in Assessing Operator Information-Processing Load. *Human Factors* **1997**, *30*, 30–47.

Hox, J. J. *Applied Multilevel Analysis*; TT-Publikaties: Amsterdam, The Netherlands, 1995.

Hughes. Task-Driven Camera Operations for Robotic Exploration. *IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems and Humans* **2005**, *35* (4), 513–522.

Hughes, S.; Lewis, M. Task-Driven Camera Operations for Robotic Exploration. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* **2005**, *35*, 513–522.

Humphrey, C. M.; Henck, C.; Sewell, G.; Williamson, B. W.; Adams, J. A. Assessing the Scalability of a Multi-Robot Interface. *Proceedings of the Second ACM/IEEE International Conference on Human-Robot Interaction, HRI*, Arlington, VA, 2007; pp 239–246.

Kaber, D. B.; Endsley, M. R. The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task. *Theoretical Issues in Ergonomics Science* **2003**, *4*, 1–40.

Kaber, D. B.; Onal, E.; Endsley, M. Design of Automation for Telerobots and the Effect on Performance, Operator Situation Awareness and Subjective Workload. *Human Factors and Ergonomics in Manufacturing* **2000**, *10*, 409–430.

Kreft, I.; de Leeuw, J. *Introducing Multilevel Modeling*; Sage: Thousand Oaks, CA, 2000.

Krotkov, E.; Simmons, R.; Cozman, F.; Koenig, S. Safeguarded Teleoperation for Lunar Rovers: From Human Factors to Field Trials. *Proceedings of the International Conference on Environmental Systems*, Monterey, CA, 1996.

Lane, J. C.; Carignan, C. R.; Sullivan, B. R.; Akin, D. L.; Hunt, T.; Cohen, R. Effects of Time Delay on Telerobotic Control of Neutral Buoyancy Vehicles. *Proceedings of IEEE International Conference on Robotics and Automation*, Washington, DC, May 2002.

Levinthal, B. R.; Wickens, C. D. Management of Multiple UAVs With Imperfect Automation. *Proceedings of the Human Factors and Ergonomics Society's 50th Annual Meeting*, ACM Press: New York, 2006; pp 1941–1944.

Lewis, M.; Wang, J.; Hughes, S.; Liu, X. Experiments With Attitude: Attitude Displays for Teleoperation. *Proceedings of IEEE Conference on Systems, Man and Cybernetics*, Washington, DC, 2003.

Lif, P.; Jander, H.; Borgwall, J. Tactical Evaluation of Unmanned Ground Vehicle During a MOUT Exercise. *Proceedings of the 7th International Conference of Engineering Psychology and Cognitive Ergonomics HCI*, Beijing, China, 2007; pp 731–740.

Lion, D. M. Three Dimensional Manual Tracking Using a Head-Tracked Stereoscopic Display. Master's Thesis, Human Interface Technology Laboratory, University of Washington, Seattle, WA, 1993.

Luck, J. P.; McDermott, P. L.; Allender, L.; Russell, D. C. An Investigation of Real World Control of Robotic Assets Under Communication Latency. *Proceedings of the 2006 Human-Robot Interaction Conference*; ACM Press: New York, 2006; pp 202–209.

Massimino, M. J.; Sheridan, T. B. Teleoperator Performance With Varying Force and Visual Feedback. *Human Factors* **1994**, *36* (1), 145–157.

Meyer, J.; Feinshreiber, L.; Parmet, Y. Levels of Automation in a Simulated Failure Detection Task. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, Washington, DC, **2003**; pp 2101–2106.

Miller, C. A.; Parasuraman, R. Beyond Levels of Automation: An Architecture for More Flexible Human-Automation Collaboration. *Proceedings of the Human Factors and Ergonomics Society's 47th Annual Meeting*, Santa Monica, CA, 2003.

Mosier, K. L.; Sethi, N.; McCauley, S.; Khoo, L.; Orasanu, J. M. What You Don't Know Can Hurt You: Factors Impacting Diagnosis in the Automated Cockpit. *Human Factors* **2007**, *49* (2), 300–310.

Murray, S. A. Human-Machine Interaction With Multiple Autonomous Sensors. *Proceedings of the 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, Cambridge, MA, June 1995.

Muthard, E. K.; Wickens, C. D. Factors That Mediate Flight Plan Monitoring and Errors in Plan Revision: Planning Under Automated and High Workload Conditions. *Proceedings of the 12th International Symposium on Aviation Psychology*, Dayton, OH, 2003.

Nielsen, C. W.; Goodrich, M. A.; Ricks, R. W. Ecological Interfaces for Improving Mobile Robot Teleoperation. *IEEE Transactions on Robotics,* **2007**, *23* (5), 927–941.

Olmos, O.; Wickens, C. D.; Chudy, A. Tactical Displays for Combat Awareness: An Examination of Dimensionality and Frame of Reference Concepts and the Application of Cognitive Engineering. *The International Journal of Aviation Psychology* **2000**, *10* (3), 247–271.

Parasuraman, R.; Galster, S.; Miller, C. Human Control of Multiple Robots in the RoboFlag Simulation Environment. *Proceedings of IEEE Conference on Systems, Man, and Cybernetics*, Washington, DC, 2003; pp 3233–3237.

Parasuraman, R.; Galster, S. M.; Squire, P.; Furukawa, H.; Miller, C. Flexible Delegation-Type Interface Enhances System Performance in Human Supervision of Multiple Robots: Empirical Studies With RoboFlag. *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans* **2005**, *35* (4), 481–493.

Park, S. H.; Woldstad, J. C. Multiple Two-Dimensional Displays as an Alternative to Three-Dimensional Displays in Telerobotic Tasks. *Human Factors* **2000**, *42* (4), 592–603.

Pazuchanics, S. L. The Effects of Camera Perspective and Field of View on Performance in Teleoperated Navigation. *Proceedings of Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, CA, 2006; pp 1528–1532.

Powell, D.; Gilbreath, G.; Bruch, M. Multi-Robot Operator Control Unit. *SPIE Proc. 6230: Unmanned Systems Technology VII*, Defense Security Symposium, Orlando, FL, 18–20 April 2006.

Reddy, M. *The Effects of Low Frame Rate on a Measure of User Performance in Virtual Environments*; Report # ECS-CSG-36-97; University of Edinburgh: Edinburgh, Scotland, December 1997.

RefWorks. http://www.refworks.com/ (accessed December 2, 2009).

Richard, P.; Birebent, G.; Coiffet, P.; Burdea, G.; Gomez, D.; Lagrana, N. Effect of Frame Rate and Force Feedback on Virtual Object Manipulation. *Presence* **1996**, *5* (*1*), 95–108.

Robinson, W. S. Ecological Correlations and the Behavior of Individuals. *Sociological Review* **1950**, *15*, 351–357.

Rovira, E.; McGarry, K.; Parasuraman, R. Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task. *Human Factors* **2007**, *49*, 76–87.

Ruff, H. A.; Narayan, S.; Draper, M. H. Human Interaction With Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence: Teleoperators and Virtual Environments 11* **2002**, *4*, 335–351.

Ruff, H. A.; Calhoun, G.; Draper, M.; Fontejon, J.; Abbott, J. Exploring Automation Issues in Supervisory Control of Multiple UAVs. *Proceedings of the 5th Conference of Human Performance, Situation Awareness, and Automation Technology*, Pittsburg, PA, **2004**; pp 218–222.

Schipani, S. P. *An Evaluation of Operator Workload During Partially Autonomous Vehicle Operations*; Report 2010-03-24; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2003.

Scribner, D. R.; Gombash, J. W. *The Effect of Stereoscopic and Wide Field of View Conditions on Teleoperator Performance*; Report A812143; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 1998.

Sellner, B. P.; Hiatt, L. M.; Simmons, R.; Singh, S. Attaining Situational Awareness for Sliding Autonomy. *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI*, Salt Lake City, UT, 2006.

Shreik-Nainar, M. A.; Kaber, D. B.; Chow, M-Y. Control Gain Adaptation in Virtual Reality Mediated Human-Telerobot Interaction. *Human Factors and Ergonomics in Manufacturing* **2003**, *15* (3), 259–274.

Smyth, C. C. *Modeling Indirect Vision Driving With Fixed Flat Panel Displays: Task Performance and Mental Workload*; ARL-TR-2701; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2002.

Smyth, C. C.; Gombash, J. W.; Burcham, P. M. *Indirect Vision Driving With Fixed Flat Panel Displays for Near-Unity, Wide, and Extended Fields of Camera View*; Report A544293; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2001.

Snijders, T.; Bosker, R. *Multilevel Analysis*. Sage: Thousand Oaks, CA, 1999.

Squire, P.; Trafton, G.; Parasuraman, R. Human Control of Multiple Unmanned Vehicles: Effects of Interface Type on Execution and Task Switching Times. *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction. HRI*, Salt Lake City, UT, 2006; pp 26–32.

Thomas, L. C.; Wickens, C. D. *Effects of Display Frames of Reference on Spatial Judgments and Change Direction*; Report ARL-00-14/FED-LAB-00-4; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2000.

Trouvain, B.; Wolf, C. Design and Evaluation of a Multi-Robot Interface. *The Role of Humans in Intelligent and Automated Systems, Symposium*, *RTO HFM*, Warsaw, Poland, 2003; pp 22-1–22-12.

Trouvain, B.; Schlick, C.; Mervert, M. Comparison of a Map- vs. Camera-Based User Interface in a Multi-Robot Navigation Task. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, SMC*, Washington, DC, 2003; pp 3224–3231.

Van Erp, J. B. F.; Padmos, P. Image Parameters for Driving With Indirect Viewing Systems. *Ergonomics* **2003**, *45* (15), 1471–1499.

Wang, W.; Milgram, P. Effects of Viewpoint Displacement on Navigational Performance in Virtual Environments. *Proceedings of Human Factors and Ergonomics Society 47th Annual Meeting*, Santa Monica, CA, 2003; pp 139–143.

Wang, J.; Lewis, M. Human Control for Cooperating Robot Teams. *Proceedings of the 2007 Human-Robot Interaction Conference*, Arlington, VA, 2007.

Watson, B.; Walker, N.; Ribarsky, W.; Spaulding, V. Effects of Variation in System Responsiveness on User Performance in Virtual Environments. *Human Factors* **1998**, *40* (3), 403–414.

Watson, B.; Walker, N.; Woytiuk, P.; Ribarsky, W. Maintaining Usability During 3D Placement Despite Delay. *Proceedings of IEEE Conference on Virtual Reality*, Los Angeles, CA, 2003.

Wickens, C. D. The Structure of Attentional Resources. In *Attention and Performance VIII*; Nickerson, R. S., Ed.; Erlbaum: Hillsdale, NJ, 1980; pp 239–257.

Wickens, C. D. Multiple Resources and Performance Prediction. *Theoretical Issues in Ergonomics Science* **2002**, *3* (2), 159–177.

Wickens, C. D.; Dixon, S. R.; Chang, D. *Using Interference Models to Predict Performance in a Multiple-Task UAV Environment- 2 UAVs*; Report AHFD-03-9/MAAD-03-1; University of Illinois at Urbana-Champaign: Champaign, IL, 2003.

Wickens, C. D.; Dixon, S. R.; Goh, J.; Hammer, B. *Pilot Dependence on Imperfect Diagnostic Automation in Simulated UAV Flights: An Attentional Visual Scanning Analysis*; Report AHFD-05-02/MAAD-05-02; University of Illinois: Champaign, IL, 2005.

Witmer, B. G.; Kline, P. B. Judging Perceived and Traversed Distance in Virtual Environments. *Presence* **1998**, *7*(2), 144–167.

Yeh, M.; Wickens, C. D. Display Signaling in Augmented Reality: Effects of cue Reliability and Image Realism on Attention Allocation and Trust Calibration. *Human Factors* **2001**, *43*, 355–365.

Yi, H.; Song, B.; Ji, D.; Yu, T. Experimental Research on Situation Awareness of the Operators for Unmanned Aerial Vehicle. *IEEE Transactions on Systems, Man, and Cybernetics* **2006**, *36* (6), 1225–1228.

NO. OF COPIES	ORGANIZATION
1 (PDF only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC HRR 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM L 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIM P 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL D 2800 POWDER MILL RD ADELPHI MD 20783-1197
	ABERDEEN PROVING GROUND
1	DIR USARL RDRL CIM G (BLDG 4600)

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HRM A J MARTIN MYER CENTER BLDG 2700 RM 2D311 FORT MONMOUTH NJ 07703-5601
1	ARMY RSCH LABORATORY – HRED RDRL HRM C A DAVISON 320 MANSCEN LOOP STE 115 FORT LEONARD WOOD MO 65473
1	ARMY RSCH LABORATORY – HRED RDRL HRM DI T DAVIS BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1	ARMY RSCH LABORATORY – HRED RDRL HRS EA DR V J RICE BLDG 4011 RM 217 1750 GREELEY RD FORT SAM HOUSTON TX 78234-5002
1	ARMY RSCH LABORATORY – HRED RDRL HRM DG R SPINE BLDG 333 PICATINNY ARSENAL NJ 07806-5000
1	ARMY RSCH LABORATORY – HRED ARMC FIELD ELEMENT RDRL HRM CH C BURNS THIRD AVE BLDG 1467B RM 336 FORT KNOX KY 40121
1	ARMY RSCH LABORATORY – HRED AWC FIELD ELEMENT RDRL HRM DJ D DURBIN BLDG 4506 (DCD) RM 107 FORT RUCKER AL 36362-5000
1	ARMY RSCH LABORATORY – HRED RDRL HRM CK J REINHART 10125 KINGMAN RD FORT BELVOIR VA 22060-5828
1	ARMY RSCH LABORATORY – HRED RDRL HRM AY M BARNES 2520 HEALY AVE STE 1172 BLDG 51005 FORT HUACHUCA AZ 85613-7069
1	ARMY RSCH LABORATORY – HRED RDRL HRM AP D UNGVARSKY POPE HALL BLDG 4709 BCBL 806 HARRISON DR FORT LEAVENWORTH KS 66027-2302

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HRM AJ J HANSBERGER JFCOM FE 115 LAKEVIEW PKWY STE B SUFFOLK VA 23435
1	ARMY RSCH LABORATORY – HRED RDRL HRM DQ M R FLETCHER NATICK SOLDIER CTR AMSRD NSC WS E BLDG 3 RM 343 NATICK MA 01760-5020
1	ARMY RSCH LABORATORY – HRED RDRL HRM AT J CHEN 12350 RESEARCH PKWY ORLANDO FL 32826-3276
1	ARMY RSCH LABORATORY – HRED RDRL HRM AT C KORTENHAUS 12350 RESEARCH PKWY ORLANDO FL 32826
1	ARMY RSCH LABORATORY – HRED RDRL HRM AS C MANASCO SIGNAL TOWERS BLDG 29808A RM 303 FORT GORDON GA 30905-5233
1	ARMY RSCH LABORATORY – HRED RDRL HRM CU 6501 E 11 MILE RD MS 284 BLDG 200A 2ND FL RM 2104 WARREN MI 48397-5000
1	ARMY RSCH LABORATORY – HRED FIRES CTR OF EXCELLENCE FIELD ELEMENT RDRL HRM AF C HERNANDEZ 3040 AUSTIN RD RM 221 FORT SILL OK 73503-9043
1	ARMY RSCH LABORATORY – HRED RDRL HRM AV S MIDDLEBROOKS 91012 STATION AVE RM 348 FORT HOOD TX 76544-5073
1	ARMY RSCH LABORATORY – HRED RDRL HRM CN R SPENCER DCSFDI HF HQ USASOC BLDG E2929 FORT BRAGG NC 28310-5000

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY – HRED RDRL HRM DW E REDDEN BLDG 4 CL 60 FORT BENNING GA 31905-5400
1 (CD only)	ARMY G1 DAPE MR B KNAPP 300 ARMY PENTAGON RM 2C489 WASHINGTON DC 20310-0300
1	ARMY RSCH LABORATORY – HRED RDRL HRM D T DAVIS BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
	ABERDEEN PROVING GROUND
5	DIR USARL RDRL CIM G S FOPPIANO RDRL HR T LETOWSKI RDRL HRM B J LOCKETT RDRL HRS L ALLENDER RDRL HRS D B AMREIN

INTENTIONALLY LEFT BLANK.