ARMY RESEARCH LABORATORY

Human-Robot Interface: Issues in Operator Performance, Interface Design, and Technologies

by Jessie Y.C. Chen, Ellen C. Haas, Krishna Pillalamarri, and Catherine N. Jacobson

ARL-TR-3834 July 2006

NOTICES

Disclaimers

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5425

ARL-TR-3834 **July 2006**

Human-Robot Interface: Issues in Operator Performance, Interface Design, and Technologies

Jessie Y.C. Chen, Ellen C. Haas, Krishna Pillalamarri, and Catherine N. Jacobson
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* July 2006	**2. REPORT TYPE** Final	**3. DATES COVERED (From - To)** June 2004 to June 2005
4. TITLE AND SUBTITLE Human-Robot Interface: Issues in Operator Performance, Interface Design, and Technologies		**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Jessie Y.C. Chen; Ellen C. Haas, Krishna Pillalamarri, and Catherine N. Jacobson (all of ARL)		**5d. PROJECT NUMBER** 62716AH70 **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory Human Research and Engineering Directorate Aberdeen Proving Ground, MD 21005-5425		**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-3834
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report tries to examine salient issues in robotic operator performance and reviews some of the promising user interface solutions, in designs and technologies. The first section concerns general issues in human-robot interaction and operator control unit designs. The second section presents the controlling of teleoperated and semi-autonomous robots and its associated human performance issues as well as user interface solutions. The last section surveys potential innovative technologies for enhancing the performance of robotic operators. Specifically, it concerns multimodal technologies, including voice recognition/synthesis systems, bone conduction and throat microphones, and tactile systems.

15. SUBJECT TERMS
auditory control and display, haptic display, human-robot interface, human-robot interaction, multi-modal, robotic operator performance

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	100	Jessie Y.C. Chen
Unclassified	Unclassified	Unclassified			**19b. TELEPHONE NUMBER** *(Include area code)* 407-384-5435

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

Contents

List of Figures

List of Tables

Acknowledgments

The authors wish to thank Michael J. Barnes of the U.S. Army Research Laboratory's Human Research and Engineering Directorate for his support and guidance throughout the process of preparing this report. Mr. Barnes is the co-manager of the Technology for Human-Robot Interaction (HRI) Soldier-Robot Teaming Army Technology Objective (ATO) (now under the Robotics Collaboration ATO) and has provided much valuable advice about important HRI issues to us. The authors also wish to thank the reviewers for their helpful comments.

1. Introduction

The presence of robotic technologies and the concurrent existence of research and development programs is growing in many field applications such as space exploration, search and rescue, national defense, entertainment, police special weapons and tactics operations, health care, and personal assistance. Although the use of robotic assets in different applications introduces concerns of human-robot interaction (HRI) that are unique to its particular application, several principles and issues of HRI transcend situational circumstances in which robotic assets are employed. This report tries to examine some of the most salient issues in robotic operator performance and reviews some of the promising user interface solutions, both in designs and technologies.

This report consists of three sections. The first section concerns general issues in HRI and operator control unit (OCU) designs. This section discusses the status of robotics as they are employed within various operational environments. A discussion of the application of robotic assets within different social sectors (e.g., civilian and military) is followed by specific HRI issues. In the second section, controlling of teleoperated and semi-autonomous robots and its associated human performance issues and user interface solutions are presented. In the last part of this section, issues related to human-robot teaming are examined. The third section surveys potential innovative technologies for enhancing the performance of the robotic operators. Specifically, it concerns multimodal technologies, including voice recognition and synthesis systems, bone conduction and throat microphones, and tactile systems. The usefulness of these systems to human-robot teams are presented.

1.1 The Use of Robots

1.1.1 Civilian Efforts

The use of robotic assets in the civilian arena is continually growing. Space research is increasingly incorporating autonomous technologies as a means of conducting field operations when human effort is unsafe, infeasible, or simply not cost effective. The National Aeronautics and Space Administration (NASA) is currently debating the use of robotic assets to assume the tasks of maintaining and servicing the Hubble telescope (David, 2004). NASA officials are in the process of determining the extent to which robotic servicing of equipment, given the current state of robotic technologies, is a workable alternative. NASA's two unmanned robots (Opportunity and Spirit) are completing their missions on Mars as part of the effort to employ robots in outer space exploration (Associated Press, 2004). The robots' missions are to cover a quota of miles of ground and to conduct specific photography tasks. The Defense Advanced Research Projects Agency's (DARPA's) Grand Challenge illustrates the ongoing efforts to push technological advancements in unmanned robotics (Markoff, 2005). DARPA invited teams to enter their robots in a 150-mile race across the Mojave Desert with a price of $2,000,000 to be awarded to the winner (increased from $1,000,000 in 2004). The public call for race entries and the substantial purse was in answer to Congress' call to accelerate robotic research and development initiatives. Although no entries completed more than 5% of the 150-mile trek in the 2004 contest, the race instilled motivation in engineers to build a better robot.

There is a growing body of research and development projects that focuses on the use of robots for search and rescue missions. Researchers in academia and industry are working together to improve robotic assets used for urban search and rescue (USAR) operations. In 2001, the joint international project team RoboCup, in conjunction with the American Association of Artificial Intelligence (AAAI) hosted the 2001 AAAI/RoboCup Robot Rescue Event (Casper & Yanco, 2002). Similar to DARPA's Grand Challenge, the event was geared toward pushing researchers to continue in their efforts to design better robots for use in USAR operations. The event also provided a simulated environment for researchers to further their understanding of the multiple facets of HRI.

Much of the research on robotics has focused on social acceptance of autonomous technologies and the effects of interface design on the operator. Such projects have been largely conducted in laboratory environments with controlled settings. A study conducted by Burke, Murphy, Coovert, and Riddle (2004) examined the interactions of humans and robots in operational environments with a focus on the human side of the interaction. In this experiment, robots were employed in a mockup of a collapsed building where data could be collected in simulated field applications across a span of 16 hours of drill time. Researchers assessed team processes, communication between operators, shared mental models, and the associated levels of situational awareness (SA).

The September 11, 2001, attacks on the World Trade Center provided an (albeit unfortunate) opportunity for robots to be employed in a full-scale non-simulated technical search task (Casper & Murphy, 2003). Representatives from several industries worked together under supervision of the Center for Robotic Assisted Search and Rescue (CRASAR) to employ unmanned robots to search for victims, transport medical supplies, and examine areas beneath the rubble to support the work of structural engineers. For this unstaged USAR event, six different robots were employed, each with its own set of "skills" and corresponding OCU. Once the robotic missions were complete and the representative teams were demobilized, a *post hoc* analysis of HRI was performed. CRASAR researchers also assessed the human-robot ratio and characteristics of communication between agent and operator for each type of robotic asset as well as the general work flow of robots during use.

The use of robots in the aftermath of the 1995 Oklahoma City bombing and the 2001 attacks on the World Trade Center have led to an increasing interest in the development of rescue robots. As an emerging liaison between laboratory researchers and disaster response teams, CRASAR warns that in the rush to deliver material solutions, engineers must consider the needs of the search and rescue community to effectively employ robots for USAR-specific tasks (Murphy, 2004). The relationship between laboratory researchers and the USAR community is continuing.

In the field of entertainment, robotics is developing its own niche. Digital entertainment companies have been and continue to apply substantial effort to the development of realistic and satisfying robotic companions as evidenced by the work of researchers and engineers in Sony's robotic entertainment sector (Arkin, Fujita, Takagi, & Hasegawa, 2003). Sony has created a dog-like robot (AIBO[1]) as well as a humanoid robot (Sony dream robot), both of which have evolved from extensive research in such areas as ethnology (study of animal behavior) and human psychology. Such research allows humans to identify with the robotic behavior and interact with their robots in predictable ways, thus promoting the process of bonding with their robotic companions.

1.1.2 Military Efforts

The Army's Future Combat System (FCS) Brigade Combat Team (BCT) incorporates a wide array of unmanned assets, including aerial and ground vehicles as well as unmanned sensor platforms. FCS is actually the first Army program to include unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) in a significant context within the force structure. Research and development efforts are currently under way in academia, and industry, and Department of Defense (DoD) laboratories. For example, the U.S. Army Research Laboratory's (ARL's) Robotics Collaborative Technology Alliances (RCTA) program developed autonomous mobility technology that was capable of operating in rolling, desert, and urban terrain, and the RCTA has demonstrated the capabilities with the Demo III eXperimental unmanned vehicle and a 10-ton Stryker platform (Robotics Collaborative Technology Alliances, 2004). The RCTA described their capabilities as "enhancing Soldier physical security and survivability, improving SA and understanding, and conducting reconnaissance, surveillance, targeting and acquisition missions in an era of rapidly evolving operational and technological challenges" (p. 1).

The Army's FCS program staged the unmanned combat demonstration (UCD) as part of an ongoing effort to integrate robotic assets into the Army's force structure (Kamsickas, 2003). The UCD, one of several FCS technology demonstrations, primarily focused on determining a realistic span of control (operator workload) in the manning of remotely controlled vehicles during operations in a tactical environment. An understanding of how participants employed the conceptual model of the UGV was used to determine realistic functional requirements for that system. The effort combined Government and industry to characterize and evaluate the Soldier workload associated with manning robotic assets (UGV), with the overall objective of assessing Soldier effectiveness. The demonstration employed only the armed reconnaissance vehicle (ARV) since this asset has many capabilities, thus placing a wide range of control on the operator to employ those capabilities. In a three-phase process whereby test and evaluation

[1]AIBO, which is not an acronym, is a registered trademark of Sony Corporation.

progressed from a simulated to a virtual environment, it was found that a realistic span of control (workload) is one Soldier to multiple UGVs in non-volatile environments and one Soldier per UGV during times of attack. The UCD also brought to light the need for tactics, techniques, and procedures (TTPs) for the employment of unmanned assets that are specific to the tasks, features, and characteristics of those systems. It was noted that during the UCD, Soldiers relied on TTPs for manned vehicles when they operated the ARV.

The FCS command and control (C2) program, led by DARPA and the Communications and Electronics Command, has examined future battle command at the small-unit level. In a series of experiments conducted from 2001 to 2005, several 2-week-long commander-in-the-loop experiments were conducted in which several 90-minute-long battle exercises were conducted with participants sitting in mock C2 vehicles (C2Vs) and infantry carrier vehicles (ICVs). As part of the series of experiments, control of UAV and UGV assets resided in the C2V crew tasks. Although not central to these experiments, control of unmanned assets was embedded as component to the FCS concepts tested here.

In August 2003, a demonstration of Warrior's Edge Technologies[2] was conducted in which Soldiers provided their opinions regarding several Army assets, many of which were prototypes of unmanned systems such as the all-terrain reconnaissance vehicle, small UAV, an unmanned vehicle called PackBot, and unmanned ground sensors. Soldiers were asked to provide their impressions of these new technologies with respect to enhanced SA, workload levels, and decision making, all within the context of a military operations on urban terrain (MOUT) site. Surveys were designed to highlight deficiencies and successes for those systems presented in this demonstration. In terms of unmanned assets, the multifunctional utility-logistics and equipment vehicle was identified as a key source of information. Feedback regarding the unmanned technologies suggests that their usefulness is well regarded and that their contribution to battlefield understanding is substantial. Specific challenges in the design of systems presented at the demonstration, although not specifically directed at unmanned assets, also provided useful feedback for continued development.

Blackburn, Laird, and Everett (2001) provided a detailed review of lessons learned from several UGV vehicle programs, which is presented on line at http://www.spawar.navy.mil/sti/publications/pubs/tr/1869/tr1869.pdf.

1.2 HRI

1.2.1 Metrics

Although metrics for evaluating the HRI are commonly derived from the specific circumstances within which the robotic system is employed, it is believed that research and development of

[2]Warrior's Edge is a program that brings network-centric warfare to the dismounted Soldier through a combination of data fusion, wireless network connectivity, and the use of lightweight portable robotic sensor platforms.

robotics has reached a point where some generalities of HRI transcend specific applications (Fong et al., 2004). Fong and colleagues have proposed a set of metrics through which task-oriented HRI can be evaluated. Specifically, the metrics are designed to assess the level of effort required on behalf of the human and the robot in order to jointly accomplish tasks. For this study, task and common metrics are discussed (Fong et al., 2004). In defining the task-specific metrics that are applicable to the operation of mobile robots, we identified five tasks: (a) navigation from points A to B, (b) perception of remote environment, (c) management of robot and human tasks, (d) manipulation of remote environment by robot, and (e) tasks involving social interaction. This study concedes that certain factors inherent in human-robot teams (e.g., communication limitations, robot response time, user limitations) present confounds.

1.2.2 Principles

In an attempt to design robot technologies to minimize workload bottlenecks and error potential within the HRI, Goodrich and Olsen (2003) developed a set of principles that are based on the concept that because of technological limitations involving HRI and the robot-environment interaction, all human intent for robot performance is transformed into an augmented version of what was desired and what could actually be performed. The authors of this study (Goodrich & Olsen, 2003) suggest seven principles for effective interface design. The bases for these principles are neglection time (how long a robot can perform a task effectively without human interaction), interaction time (the time it takes a robot's performance to rise from threshold to maximum after human interaction begins), robot attention demand (how much time is required to operate a robot as a function of the mathematical relationship between neglection time and interaction time), free time (the amount of time remaining for secondary tasks during HRI—also a function of neglection and interaction times), and fan out (the number of HRIs that can be performed simultaneously, given that the robots are the same). These five concepts lay the foundations for the seven principles of efficient interface design developed by Goodrich and Olsen. The following is a brief summary of the seven principles:

- The first principle stipulates that switching between different interaction and autonomy modes should require as little time and effort as possible. No mental model should be required to switch between modes; knowledge of how to act in each mode should be sufficient.
- The second principle requires that cues provided to the robot should be natural whenever possible. The use of natural cues taps our pre-existing database of expressions used to convey intent. Map-based sketching is an example of a natural cue. Skubic, Bailey, and Chronis (2003) investigated the use of this naturalistic cue as an effective means for conveying intent to robots. This is addressed in more detail in a subsequent section of this report.
- The third principle emphasizes the ability of the operator to have as much direct contact with the target environment as possible to reduce interfacing with the robot. An example

of a direct link between the human and the target environment is a touch screen that displays an image of the environment. The operator touches a point of interest on the screen in order to, for example, indicate a new destination point for the robot. Touching the screen at the area of interest is essentially the command input that directs the robot's movements. A direct link between the operator and the target environment reduces operator workload because the operator needs only a mental model of the environment and not of the robot in order to successfully initiate commands for the robot.

- The fourth principle arises from the concession that a direct link between the operator and the target environment is not always achievable. When direct links are not possible, it is best to design the interface so that operator focus remains on the target environment. A status display of the environment (e.g., terrain detail and temperature) exemplifies an effort to keep the operator focused on the target environment in which the robot is employed and not on the robot itself.
- The fifth principle of an effective interface requires that information provided to the operator should be open to manipulation as needed. For example, specific feedback about the status of a robot (e.g., altitude of a UAV) should allow for manipulation of that feedback (e.g., change the altitude of the UAV).
- A sixth principle, designed to reduce cognitive workload and thus increase the operator's ability to multitask, involves externalizing information that would normally reside in the operator's short-term memory. In reference to the control of mobile robots, externalizing memory may include displays of surrounding terrain features that are not immediately within the robotic sensors but are necessary for one to keep in mind when traversing across a target environment.
- The final principle is aimed at ensuring that the interface design allows for proper management of the operator's attention so that it is directed to critical information at the proper times.

Goodrich and Olsen's (2003) seven principles of effective interface design for mobile robots represent a general trend in the robotics literature to begin summarizing and generalizing the information to date about HRI design concepts.

1.2.3 Human Role in HRI

The role of the human in human-robot teams has been defined and described in many ways and for many reasons. Burke, Murphy, Rogers, Lumelsky, and Scholtz (2004) developed a taxonomy of human-robot teams within which different operator roles are defined. Much of the research on HRI is built from studies in which operators and robots alike play specific roles (e.g., human as teleoperator, human as commander or as bystander). The roles of humans and robots can vary within an operational exercise as well, given such concepts as *traded control* (Schreckenghost, 1999) whereby the human and robot roles change in response to changing environmental situations. Scholtz and Bahrami (2003) define three roles (supervisor, operator, and peer) which are further subdivided into more specific roles. Supervisors are responsible for oversight and intervention when necessary. The role of operator is divided between operator and mechanic. The operator manipulates, configures, and programs the robot while the mechanic resolves technical and hardware malfunctions. The peer is further subdivided into the teammate and the bystander. Teammates work in multiple human-robot teams while bystanders are not directly associated with the robot and therefore do not require formal training. Bystanders generally engage in some social interaction with robotic assets either directly or indirectly. Others have defined the roles of humans and robots in terms of operators and problem solvers (Murphy, 2004) where operators have control over manipulation of the robot and problem solvers are those who direct the overall robotic missions and analyze the data received by the robot(s).

1.2.4 Workload

Generally speaking, robotic operator's workload tends to be higher when s/he has to teleoperate a robot or manually intervene when the robot's autonomous operation encounters problems compared to managing autonomous robots (Dixon, Wickens, & Chang, 2003; Schipani, 2003). However, the level of reduction in workload with automation greatly depends on the reliability of the autonomous system (Dixon & Wickens, 2004). According to Dixon and Wickens (2004), reliability levels at about 60% to 70% may fail to provide any benefits to performance. In addition to the reliability issues, a prominent factor in the workload associated with operating robots is the concept of *context acquisition*. When the operator must switch between tasks (e.g., switching from navigation based on one set of sensory input to data analysis based on another set of sensory input), the mental effort required to reach a certain speed on each task amounts to an increased demand on the operator's cognitive resources as well as increased time required to perform the necessary mental processes to make the switch. In terms of evaluating the usability of robotic interfaces, context acquisition is considered one of several metrics (Olsen & Goodrich, 2003). Externalizing the memory required when one is switching from one task to another is one solution to reducing the workload; making historical images or data available on the interface allows operators to release cognitive resources that would originally be required to remember such historical data. The robotic operator's workload can also be affected by various factors in the robotic controlling environments. The following sections discuss those factors and potential user interface solutions in greater detail.

2. HRI and Its Associated Human Performance Issues

2.1 Teleoperation

The levels with which human operators interact with the robots range from manual control (pure teleoperation) to minimal control (full autonomy). This section focuses on human performance issues in the areas of controlling teleoperated robotic entities. Potential user interface designs to enhance operator performance are also presented.

Teleoperated robots have been used in a variety of situations, ranging from extra-planetary exploration (e.g., NASA's Mars rovers), military operations (e.g., surveillance/reconnaissance or detecting/removing hazardous materials), search-and-rescue activities (e.g., searching for survivors at the World Trade Center after September 11, 2001) to robotic surgery (Associated Press, 2004; Cao, Webster, Perreault, Schwaitzberg, & Rogers, 2003; Casper & Murphy, 2003; Johnston, Wilson, & Birch, 2002; Nguyen et al., 2001; Schenker, Huntsburger, Pirjanian, & McKee, 2001).

Robots can be teleoperated through a wide variety of control media, ranging from hand-held devices such as personal digital assistant (PDA) systems (Fong, Thorpe, & Glass, 2003; Quigley, Goodrich, & Beard, 2004) and cellular phones (Sekmen, Koku, & Zein-Sabatto, 2003) to multiple panel displays with control devices such as joysticks, wheels, and pedals (Kamsickas, 2003). Typical control stations include panels displaying (a) sensor view and/or data transmitted from the robots, (b) commands issued to the robots, (c) health status of the robots, and (d) map displays to maintain the operator's SA and to facilitate navigation. PDA user interfaces, on the other hand, frequently employ touch-based interactions (e.g., stylus) and multi-modal systems such as natural language and visual gesturing (Fong, Thorpe, & Glass, 2003; Keskinpala, Adams, & Kawamura, 2003; Perzanowski et al., 2003). The sizes of the unmanned vehicles (UV) range from just a few inches in dimension to multi-ton vehicles such as modified M1 tanks (Carlson & Murphy, 2004; Malcolm & Lim, 2003).

Human performance issues involved in teleoperating UV generally fall into two categories, namely, remote perception and remote manipulation. Teleoperation tends to be challenging because operator performance is "limited by the operator's motor skills and his ability to maintain situational awareness…difficulty building mental models of remote environments…distance estimation and obstacle detection can also be difficult" (Fong, Thorpe, & Baur, 2003, p. 699). In real-world operations, operator performance sometimes is degraded even further because of robotic system failures. Carlson and Murphy (2004) reviewed data from 10 studies of 15 different UGVs in USAR and modern MOUT applications. Generally, reliability of the UGV performance in the field tends to be low (i.e., between 6 and 20 hours between failures). The common causes include "unstable control systems, platforms designed for a narrow range of conditions, limited wireless communication range, and insufficient bandwidth for video-based feedback" (p. 1). Some of these issues affected the human operator's remote perception, and some affected the remote manipulation task (which includes remote navigation). In the studies reviewed, the most common type of failure was effector failure (e.g., immobility because a rock or debris was stuck in the track mechanism, track slippage, etc.). The UGV operators also frequently encountered sensor failures, especially problems with the cameras and lighting. Camera lenses were often occluded by obstacles, moisture, or mud. Changes in lighting intensities, on the other hand, sometimes made it difficult for the camera's iris to adjust enough and therefore made the robot operator's control from the remote OCU more challenging. In addition, lack of depth perception was cited as a problem needing to be resolved. In terms of communications failures, limited bandwidth often caused video dropout and static, which hindered operator performance. Communications were especially problematic in non-line-of-sight situations. The authors indicated that limited bandwidth was more of an issue for the military than for other domains because of the military rules about allowed frequencies. The following paragraphs present a more detailed discussion of human performance issues in the area of remote perception, followed by a discussion of issues in remote manipulation.

2.1.1 Remote Perception

Remote perception is essential for effective teleoperation. In the teleoperating environments, human perception is compromised because the natural perceptual processing is de-coupled from the physical environment. This de-coupling affects people's perception of affordances in the remote scene and often creates problems in remote perception such as scale ambiguity (Woods, Tittle, Feil, & Roesler, 2004). Simple tasks could be challenging because there was no motion feedback in remote visual processing and because of the unmatching viewpoint that could result when the camera was placed at a height that did not match normal eye height (Tittle, Roesler, & Woods, 2002). Poor perception has a detrimental effect on SA and therefore, on teleoperating tasks. For remote-manipulation tasks such as bomb disposal, operators often need to estimate the absolute sizes of objects so they can decide whether it is safe for the robot to maneuver in the remote environment (e.g., without getting stuck in a depression) (Drascic, 1991). Studies of rescue robots (e.g., robots for search and rescue at the site of the World Trade Center after September 11, 2001) demonstrated that human operators' performance was often compromised because of poor spatial awareness caused by inadequate video image from the cameras and/or sensors on the robots (Casper & Murphy, 2003; Murphy, 2004). In some cases, remote human operators had difficulty estimating the sizes of clearings and whether it would be possible to climb over an obstacle (Casper, 2002). In a study by Darken, Kempster, & Peterson (2001), the participants' performance of spatial orientation and object identification in a remote environment was degraded in comparison to performance in a live walk-through condition. Expert operators of bomb disposal devices complained that the monochrome and monoscopic video they had to use made their tele-manipulation tasks very difficult, especially when "dealing with small objects outdoors or in bright sunshine and shadow conditions" (Drascic, 1991, p. 9).

In Fong et al. (2004), a framework of task metrics for HRI was presented. In the domain of remote perception, the authors suggested the following categorization:

- Passive Perception (interpretation of sensor data)
  - *Identification*: Detection and recognition of mission-related objects
  - *Judgment of extent*: Quantitative judgments about the environment (e.g., absolute and relative judgments of distance, size, or length)
  - *Judgment of motion*: Estimates of the velocity of egomotion (i.e., robotic movement) or movement of other objects

- Active Perception (seeking sensor data to enhance SA, usually involving manipulation of the camera and/or the robotic movement)
  - *Active identification*: Recognition tasks that involve mobility and/or manipulation of the camera
  - *Search*
    - Stationary search: Search tasks that do not involve mobility and usually involve camera control or data fusion from sensors
    - Active search: Search tasks that involve mobility and usually involve camera control or data fusion from sensors

The following paragraphs discuss how remote perception is affected by factors such as limited view, degraded depth perception, camera viewpoint, degraded video image, and time delay. The effects of these factors on the tasks in Fong et al. (2004) framework are presented.

2.1.1.1 Limited View

The use of cameras to capture the environment in which the robot is navigating sometimes creates the so-called "keyhole" effect (Woods et al., 2004; Murphy, 2004). In other words, only a portion of the environment can be captured and presented to the operator and it requires extra effort to survey the environment (by manipulating the cameras) in order to gain SA comparable to direct viewing. Switching from camera to camera or from one view to another also poses potential memory problems for the operator since s/he has to remember what has been seen previously and incorporate it with the current view (Olsen & Goodrich, 2003). Teleoperation is often prone to poor spatial awareness of the remote environment because of the impoverished representations from video feeds which could omit essential cues for building teleoperator's mental models of the environment (Darken & Peterson, 2002; Tittle et al., 2002). In real-world operations such as the World Trade Center rescue effort reported in Casper and Murphy (2003), operators often have to rely on the video from the robot's eye view to diagnose problems encountered by the robot when automatic proprioception information is not available. For example, in the World Trade Center case, a robot was stuck because it lodged itself on a metal rod. The operator could not diagnose the problem based on the video feed from the robot.

A restricted field of view (FOV) affects remote perception in a number of ways. Tasks such as target detection and identification of self-location in a virtual environment were found to be negatively affected when participants were asked to perform the tasks by viewing the remote environment through video (Darken et al., 2001). Thomas and Wickens (2000) demonstrated that operators tended to show "cognitive tunneling" when viewing the remote environment with the use of an immersive display (such as the ones typically used for ground robots) instead of displays with exocentric frame of reference (similar to views from a UAV), which had a greater FOV. Furthermore, important distance cues may be lost and depth perception may be degraded when FOV is restricted (Witmer & Sadowski, 1998). With a reduced FOV, drivers have more difficulty in judging the speed of the vehicle, time to collision, and perception of objects or locations such as obstacles and the start of a sharp curve (Van Erp & Padmos, 2003). Wider FOV is often used to broaden the scope of the visual scene in indirect driving and teleoperation situations to compensate for the limited FOV generated by on-board cameras. Wide FOV is especially useful in tactical driving tasks where turning and navigation in unfamiliar terrain are involved (Scribner & Gombash, 1998). However, with increasing FOV, the speed of travel tends to be perceived as increased because of the scene compression and drivers usually respond by reducing their speed (Smyth, Gombash, & Burcham, 2001). In addition, the decreased resolution and increased scene distortion associated with scene compression increase cognitive workload for tasks such as driving and locating objects as well as motion sickness symptoms. Motion sickness can also be induced by the increased ocular stimulation and motion in the peripheral vision that comes with a wider FOV. On the other hand, Smyth et al. (2001) found that spatial rotation and map planning performance was improved with the wide FOV display, and they suggested that wide FOV had a similar priming effect on spatial cognitive functioning as the direct viewing. They concluded that for indirect vision driving, optimal performance might be achieved if unity vision display were employed with the capability to electronically change FOV.

2.1.1.2 Degraded Depth Perception

The use of monocular cameras and its effects on teleoperator's depth perception has been investigated in various contexts. Basically, projecting three-dimensional (3-D) depth information onto a two-dimensional display surface results in compressed or "foreshortened" depth perception (Thomas & Wickens, 2000). The compression is worse with the ground robots than with the aerial robotic vehicles because of their low viewpoints. Using monocular cameras, the teleoperator has to rely on cues such as interposition, light and shadow, linear perspective, and size constancy of objects to judge depth of the remote scene (Rastogi, 1996). In unfamiliar or difficult terrain such as the rubble pile at the World Trade Center scene where objects are disorganized and deconstructed, depth perception is extremely challenging because of the lack of apparent size cues (Murphy, 2004).

Degraded depth perception affects teleoperator's estimates of distance and size and can have profound effects on mission effectiveness. It is well documented that humans underestimate distances more in virtual environments (VE) than in the real world (Lampton, Singer, McDonald, & Bliss, 1995; Witmer and Kline, 1998; Thompson et al., 2004). According to Witmer and Kline (1998), the texture and pattern of the floor in the VE did not significantly affect observers' judgment of distance, nor did the movement method employed by the observer (e.g., moving via a treadmill versus using a joystick). Thompson et al. found that underestimation of distance in the VE compared to the real world was consistent, regardless of the quality of graphics rendering (photographs, low-quality computer-generated graphics, and wireframe computer graphics were used to represent graphics with different levels of quality). Therefore, research about distance estimation conducted in the VE is applicable to robotic control environment, since the imagery for the latter is essentially of photographic quality. In a usability test of a mixed initiative robotic system, Marble, Bruemmer, and Few (2003) reported that "most participants indicated a desire for the interface to overlay the video with a depth indicator, especially in teleoperated mode" (p. 451).

Scribner and Gombash (1998) examined stereovision in a teleoperated environment and found that there were significant differences between mono- and stereo-vision for error rate (i.e., number of obstacles contacted). Their data also supported the findings of other driving-related research that stereo-vision enhances performance of tasks that require depth positioning, identification of negative obstacles, or navigation in unfamiliar environments. Green, Dougherty, and Savacool (2003), on the other hand, did not find the stereo-vision system beneficial in enhancing operator's depth the perception in shipboard crane handling tasks. In addition, as observed by Scribner and Gombash (1998), artificially induced binocular stereo-vision tends to increase motion sickness and the operator's stress ratings.

2.1.1.3 Camera Viewpoint (context)

A human operator's perception of the remote environment often relies on the video feeds from the camera(s) mounted on the robot. For robots with extended manipulators (e.g., arms), cameras can be placed on the gripper of the manipulator and capture the remote scene egocentrically (Rastogi, 1996). Alternatively, cameras can be placed on the body of the robot and provide an exocentric view of the movement of the manipulator. Depending on the placement of the cameras, which may or may not match the normal eye sight of the operator, remote perception (e.g., position estimation) may be degraded by the unnatural viewing angles for the human (Murphy, 2004; Van Erp & Padmos, 2003).

Multiple camera viewpoints are usually employed to enhance remote perception (especially object identification) (Casper & Murphy, 2003). Hughes and Lewis (2004) found that using a separate camera that was controlled independently from the orientation of the robot increased the operator's overall functional presence (e.g., improved search performance). Hughes and Lewis suggest a two-screen approach, where one screen is under human control and the other screen is sensor driven (i.e., a sensor would direct the operator to a particular viewpoint of interest). However, it was suggested that the differences between eye point and camera viewpoint may induce motion sickness (Van Erp & Padmos, 2003). In addition, when one is handling multiple robots, it can be challenging for the operator to acquire the different contexts rapidly when switching among the robots (Fong, Thorpe, & Baur, 2003; Olsen & Goodrich, 2003). The user has to remember, for example, the surroundings for each robot and what tasks have been and have not been performed (Casper & Murphy, 2003). Moreover, literature about change blindness suggests that information in one scene may not be encoded sufficiently to be compared or integrated when accessed subsequently (Levin & Simons, 1997; Thomas & Wickens, 2000). Therefore, some changes may go undetected when viewpoints are changed. It is even more challenging when the robots are heterogeneous and with different capabilities.

Future warfare employing the FCS may need to integrate information from multiple platforms, potentially from aerial and ground sources. The UAV generally provides an exocentric view of the problem space (i.e., the battlefield) while the UGV presents a viewpoint that is egocentric and immersed in that environment. Displays for integrating information from different frames of references (e.g., exocentric and egocentric) present potential human performance issues that need to be carefully evaluated (Thomas & Wickens, 2000). Research has shown that integrating information across egocentric and exocentric views can be challenging for the operator (Olmos, Wickens, & Chudy, 2000). In addition, operators may be susceptible to saliency effect and anchoring heuristic/bias. Salient information on one display may catch most of the operator's attention, and the operator may form an inaccurate judgment because information from the other sources is not properly attended to and integrated.

It is sometimes difficult to perceive the attitude (i.e., pitch and roll) of the robots with fixed cameras when the robots are on a grade or in an environment where regularly referenced objects for orientation (e.g., horizon, walls, and ceilings, etc.) are not available (Lewis, Wang, Manojilovich, Hughes, & Liu, 2003). Misperception of attitude was believed to be a major contributing cause to teleoperation accidents at Sandia National Laboratories, New Mexico, in which the uniformly slanted terrain was perceived to be horizontal by the operators (McGovern, 1991).

2.1.1.4 Degraded Video Image

The communication channel between the human operator and the robot is essential for effective perception of the remote environment. Factors such as distance, obstacles, or electronic jamming may pose challenges for maintaining sufficient signal strength (French, Ghirardelli, & Swoboda, 2003). As a result, the quality of video feeds that a teleoperator relies on for remote perception may be degraded and the operator's performance in distance and size estimation may be compromised (Van Erp & Padmos, 2003). Common forms of video degradation caused by low bandwidth include reduced frame rate (frames per second), reduced resolution of the display (pixels per frame), and a lower gray scale (number of levels of brightness or bits per frame) (Rastogi, 1996). The product of frame rate, resolution, and gray scale is bandwidth (bits per second), and it is important to determine how to exchange these three variables with a given bandwidth so that operator performance can be optimized (Sheridan, 1992).

Piantanida, Boman, and Gille (as cited in Reddy, 1997) found that participants' depth and egomotion perception degraded when frame rates dropped. Similarly, Darken et al. (2001) demonstrated that people had difficulty maintaining spatial orientation in a remote environment with a reduced bandwidth. The participants also had great difficulty in identifying objects in the remote environment. For applications in VE, many researchers recommend 10 Hz to be the minimum frame rate to avoid performance degradation (Watson, Walker, Ribarsky, & Spaulding, 1998). Van Erp and Padmos (2003) suggest that speed and motion perception may be degraded if image update rate is below 10 Hz. French et al. (2003) suggest that no fewer than eight frames per second be employed for teleoperation of the UGV, based on their experimental results.

A different form of degraded video image, the so-called "jitter," also happens when the amount of time between two signals at the receiving end is different from when they are sent (Fong et al., 2004). The effects of this type of anomaly on human remote perception remain to be investigated.

2.1.1.5 Time Delay

Time delay (i.e., latency, end-to-end latency, or lag) refers to the delay between input action and (visible) output response and is usually caused by the transmission of information across a communication network (MacKenzie & Ware, 1993; Fong et al., 2004). Studies of human performance in the VE show that people are generally able to detect latency as low as 10 to 20 ms (Ellis, Mania, Adelstein, & Hill, 2004). Meehan, Razzaque, Whitton, and Brooks (2003), on the other hand, reported that participants in a lower latency (i.e., 50 ms) condition had a higher self-reported sense of presence in a stress-inducing virtual environment than did the participants in the higher latency group (i.e., 90 ms) although the difference was not statistically significant. However, the lower latency group did experience a significantly higher heart rate change from the baseline level. Other studies also reported lower subjective ratings of presence associated with latencies (Jung, Adelstein, & Ellis, 2000; Kaber, Riley, Zhou, & Draper, 2000). It is not clear if and how these findings on telepresence in VE can be applied to non-immersive environments. In addition, the effects of time delay are usually investigated in the context of remote manipulation rather than in remote perception and are therefore discussed in greater detail in the following section.

2.1.2 Remote Manipulation

Remote manipulation is a fundamental part of the robotics operator's task. It usually includes a navigation task (i.e., moving the robot from point A to point B) and a manipulation task (e.g., arm-based grasping, non-prehensile motions such as pushing, and discrete actions such as payload management) (Fong et al., 2004). This section discusses how factors such as limited view, degraded video image, time delay, and motion affect these tasks.

2.1.2.1 Limited View

Research in driving performance with restricted FOV shows that the effectiveness of remote driving can be compromised because of the limited view. For example, several studies show that peripheral vision is important for lane keeping and lateral control (Van Erp & Padmos, 2003). Land and Lee (1994) found that when driving on curved roads, drivers rely on the "tangent point" on the inside of the curve. A restricted FOV might hinder the turning task since this tangent point has to be determined 1 to 2 seconds before the bend. Drivers with a limited FOV often initiate their control actions earlier than optimal (Van Erp & Padmos, 2003). Oving and Van Erp (2001) compared driving an armored vehicle with head-mounted displays (HMD) versus periscopes and observed better vehicle control and faster task completion time with the HMD system. However, Oving and Van Erp (2001) and Smyth, Paul, Meldrum, and McDowell (in process) showed that the HMD might induce greater motion sickness in comparison to other viewing conditions.

2.1.2.2 Degraded Video Image

As reported earlier, people have difficulty maintaining spatial orientation in remote environments when video image is degraded because of reduced bandwidth (Darken et al., 2001). Richard et al. (1996) reported that tracking performance degraded for low frame rates (i.e., 7 Hz, 3 Hz, 2 Hz, and 1 Hz) but did not degrade significantly when frame rates dropped from 28 Hz to 14 Hz. Massimino and Sheridan (1994) demonstrated that teleoperation was significantly affected with a rate of five to six frames/second and became almost impossible to perform when the frame rate dropped below three frames/second. Chen, Durlach, Sloan, and Bowens (2005) found that with a 5-Hz frame rate, participants' target acquisition performance was somewhat degraded, although not significantly. Several studies examined the effects of reduced frame rates on driving performance. According to Van Erp and Padmos (2003), lowering the image update rate may affect speed estimation and braking. French et al. (2003) showed that reduced frame rates (e.g., two or four frames per second) affected the teleoperator's performance in navigation duration (time to complete the navigation course) and perceived workload. It was worth noting that no significant differences were found among different frame rates (i.e., 2, 4, 8, and 16 fps) for navigation error, target identification, and SA. The authors recommended that no fewer than eight frames per second be employed for teleoperating UGVs. It appears that increasing the frame rate to higher than 8 Hz might not greatly enhance indirect driving performance. For example, in a study of teleoperation of ground vehicles, McGovern (1991) did not find driving performance degradation when image update rates were lowered from 30 to 7.5 Hz.

2.1.2.3 Time Delay

Sheridan and Ferrell (1963) conducted one of the earliest experiments on the effects of time delay on teleoperating performance. They observed that time delay had a profound impact on teleoperator's performance, and the resulting movement time increases were well in excess of the amount of delay. Based on this and other experimental results, Sheridan (2002) recommended that supervisory control and predictor displays be used to mitigate the negative impact of time delays on teleoperation (more on user interface design is presented in a later section). Generally, when system latency is more than about 1 second, operators begin to switch their control strategy to "move and wait" instead of continuously commanding and trying to compensate for the delay (Lane et al., 2002).

Several researchers have been investigating the human performance degradation in interactive systems caused by time delays less than 1 second (compared to several seconds in the Sheridan & Ferrell study). In a simulated driving task, driver's vehicle control was found to be significantly degraded with a latency of 170 ms (Frank, Casali, & Wierville, 1988). According to Held, Efstathiou, and Greene (1966), latency as short as 300 ms would make the teleoperator decouple his or her commands from the robotic system's response. Warrick (1949, as cited in Lane et al., 2002) also showed that participants' compensatory pursuit tracking performance degraded with a latency of 320 ms. Lane et al. (2002), on the other hand, did not find any performance degradation in a 3-D tracking task until the latency was more than 1 second, although the authors also reported that it took the participants significantly longer to complete a position (i.e., extraction and insertion) task when the latency was more than 500 ms. In a study of target acquisition using the classic Fitts' law paradigm, MacKenzie and Ware (1993) demonstrated that movement times increased by 64% and error rates increased by 214% when latency was increased from 8.3 ms to 225 ms. A model of modified Fitts' law (with latency and difficulty having a multiplicative relationship) was proposed, based on the experimental results. In another study of latency effects on the performance of grasp and placement tasks, Watson et al. (1998) found that when the standard deviation of latency was above 82 ms, performance degraded (especially for the placement task, which required more frequent visual feedback). It was suggested that a short variable lag could be more detrimental than a longer fixed one (Lane et al., 2002). Over-actuation (e.g., over-steering and repeated command issuing) is also common when system delay is unpredictable (Kamsickas, 2003; Malcolm & Lim, 2003).

Time delay has been associated with motion/cyber sickness, which can be caused by cue conflict (i.e., discrepancy between visual and vestibular systems) (Stanney, Mourant, & Kennedy, 1998; Kolasinski, 1995). In Oving and Van Erp's (2001) study of indirect driving of an armored vehicle, several participants in the HMD driving condition had to withdraw from the experiment because of motion sickness. The authors suspected the delays in the HMD system might have contributed to motion sickness by creating "discrepancies between the visually displayed head orientation and the vestibularly and proprioceptively sensed orientation" (p. 1376).

2.1.2.4 Motion

As planned for the FCS of the U.S. Army, operators will sometimes need to control their robotic assets from a moving vehicle (e.g., C2V). The effects of motion on teleoperation performance therefore present important issues and need to be carefully examined. The FCS lead system integrator performed a demonstration for the concept and technology development phase, in which operator's teleoperated robotic vehicles from a moving command vehicle (Kamsickas, 2003). The results showed that motion made all tasks more difficult, compared to an exercise in a simulated environment, and some tasks (e.g., editing plans and maps, and target acquisition) became almost impossible to perform. The operators needed to rely on stabilization points to brace their hands when performing some tasks. The operators also tended to over-steer their robotic vehicles when their own vehicle turned one way but the robot needed to turn the other way. A study by Cowings, Toscano, DeRoshia, and Tauson (1999) reported that the C2V crew's health and performance was degraded when the crew had to perform computerized tasks on a moving platform. Intermittent short halts and different vehicle configurations did not appear to reduce the severity of sickness and performance degradations.

2.1.3 Interface Designs for Teleoperation

User interface design is paramount to effective robotic teleoperation. Innovative techniques and technologies have been designed to enhance operator performance and ameliorate potential performance degradation discussed before. This section reviews several of these display designs and the human performance issues they try to resolve. Further information about the multimodal systems and stereoscopic displays (SD) is presented in sections 2 and 3 of this report.

2.1.3.1 Attitude Displays

Attitude (i.e., pitch and roll) of a robotic vehicle may be easy to reference when there are other familiar objects (e.g., horizon, buildings, trees, etc.) in the remote environment. However, if those reference points are absent and the on-board cameras are fixed, operators sometimes find it surprisingly difficult to accurately assess the attitude of their robotic vehicles (Heath-Pastore, 1994). In fact, misperception of attitude was cited as the only problem in an egocentric tele-operation accident at Sandia (McGovern, 1991). Essentially, the operators were not aware that their robotic vehicles were on a grade until they rolled over. Other near-roll-over incidents have been reported and it was determined that insufficient awareness of the attitude of the teleoperated vehicle caused the incidents (Aviles et al., 1990). In the World Trade Center search-and-rescue efforts, the operators had similar problems and were not aware of the orientation of the surface until their robots flipped or rolled (Murphy, as cited in Lewis et al., 2003). Lewis and his colleagues (Wang, Lewis, & Hughes, 2004) developed a gravity-referenced view (GRV) display (see figure 1) and observed that operators were more situationally aware of the robotic vehicle's attitude by using this display, although the terrains were extremely challenging and visually complex (e.g., lacking reference points for orientation). They also selected better routes (i.e., more direct and flatter) and completed their navigation tasks in shorter times. The authors cautioned that the conditions favoring the use of GRVs may be limited to those involving confusing environments and stressful operations.

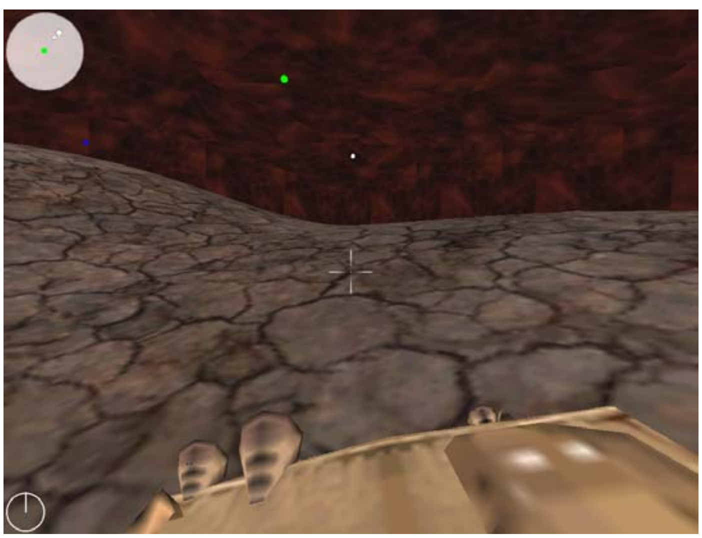

Figure 1. Attitude display (adapted from Wang, Lewis, & Hughes, 2004, with permission).

2.1.3.2 SDs

SDs, which rely on various techniques to present binocular image to the user, have been suggested as able to provide advantages over monocular displays such as faster and more accurate perception of the remote scene, enhanced detection of slopes and depressions, enhanced object recognition and detection, visual noise filtering, faster learning, and faster task performance with fewer errors (for certain tasks) (Drascic, 1991). According to Dumbreck, Smith, and Murphy (1987, as cited in Drascic, 1991), remote manipulation tasks that involve "ballistic movement, recognition of unfamiliar scenes, analysis of three-dimensionally complex scenes and the accurate placement of manipulators or tools within such scenes" especially benefit from SDs. Empirical studies examining the utility of SDs generally report that SDs might be useful in only certain circumstances. For example, Drascic (1991) found that the benefits of SDs, while longer lasting for tasks that required binocular depth cues, did not last as long for tasks that did not require much binocular depth perception. Participants generally quickly learned how to use the monocular cues available in the monocular displays to accomplish those tasks. Draper, Handel, and Hood (1991) had their participants perform Fitts' Law tapping tasks[3] and reported that SDs were only useful for more difficult tasks and only for inexperienced participants. They suggested that SDs would be useful when the image quality, task structure and predictability, user experience, and manipulator dexterity were suboptimal. Richard et al. (1996) demonstrated the utility of SDs for enhancing tracking performance in low-frame-rate conditions (i.e., slower than 7 Hz). Rosenberg (1993) found that SDs helped depth-matching performance, and the distances between the two cameras affected the usefulness of the SDs. They reported that the best performance was achieved when the inter-camera distance was less than the interocular distance (i.e., 2 to 3 cm versus 6 cm). Green et al. (2003), on the other hand, did not find significant benefits of using SDs (e.g., time and accuracy of task performance, depth perception, etc.). As for user preference, a consistent finding from various studies is that teleoperators generally prefer SDs over monocular displays (Green et al., 2003; Drascic & Grodski, 1993). However, as noted in Scribner and Gombash (1998), artificially induced binocular stereo-vision may increase motion sickness and the operator's stress ratings.

2.1.3.3 Predictive Displays

Predictive (or predictor) displays, using the teleoperator's control input, "simulate the kinematics without delay and immediately display graphically the (simulated) system output, usually superimposed on the display of delayed video feedback from the actual system output" (Sheridan, 2002, p. 108). Some predictive displays employ VE, in which the "phantom robot" reacts to the teleoperator's commands in real time (Kheddar, Chellali, & Coiffet, 2002). Various techniques such as augmented reality, visual tracking, and image-based rendering have been used for VE-based predictive displays (Rastogi, 1996; Deng & Jagersand, 2003; Ricks, Nielsen, & Goodrich, 2004). Although disturbances may exist in the remote environment and make the model of the actual environment imperfect, predictive displays have been shown to be able to reduce task

[3]Fitts' Law is a model to account for the time it takes to point to a target, based on the size and distance of the target object.

performance time by 50% to 150% (Hashimoto, Sheridan, & Noyes, 1986, as cited in Sheridan, 1992; Noyes & Sheridan, 1984). Ricks et al. (2004), on the other hand, reported that their participants finished their navigation tasks 17% faster and had only 1/5 of the collisions using the predictive display (i.e., ecological display), which also presented spatial range informa-tion using 3-D graphic and a tethered perspective, compared with a standard interface (see figure 2). The participants also preferred the ecological display four to one over the standard display.

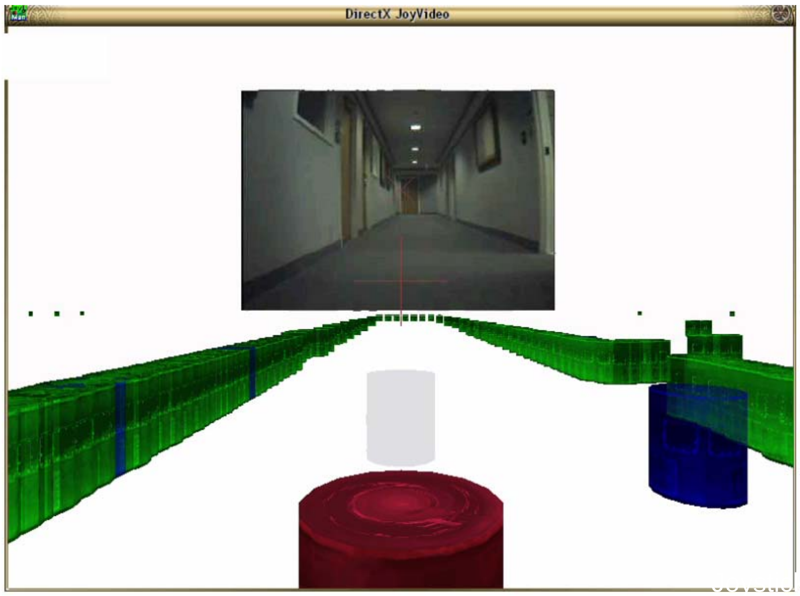

Figure 2. Ecological display (adapted from Ricks, Nielsen, & Goodrich, 2004, with permission).

2.1.3.4 Multimodal Interfaces

Robotic teleoperation has been predominantly a visual task. However, as technology becomes increasingly complex, a single-modality user interface may not allow operators to manage and manipulate their robots effectively (Vitense, Jacko, & Emery, 2003). Multimodal interfaces take advantage of the multiple human sensory channels and can potentially enhance operator performance and alleviate visual workload (Horrey & Wickens, 2004). Draper, Calhoun, Ruff, Williamson, and Barry (2003) reported that speech-based input was more effective than manual input in enabling UAV operators to navigate through menus and select options more quickly and accurately. However, depending on the missions, speech-based interfaces may not be practical (e.g., stealth conditions). In addition, auditory stimuli may draw attention away from the visual tasks because of their onset's intrinsic alerting characteristics; the operator also needs to address auditory information immediately because it fades from working memory quickly (Horrey & Wickens, 2004). These limitations present challenges to effective multimodal user interface designs. More on multimodal displays is presented in section 2.

Haptic/tactile displays are also promising technologies for robotic control. Many haptic systems are developed for robot-assisted telesurgery (Bar-Cohen et al., 2001; Kennedy, Hu, Desai, Wechsler, & Kresh, 2002; Tholey, Pillarisetti, Green, & Desai, 2004). Calhoun, Draper, Ruff, Fontejon, and Guilfoos (2003) evaluated the usefulness of a tactile alert system for a UAV ground control station operation. They found that tactile alerts (delivered via a wrist-worn vibrating tactor) were more effective in informing the operators than were visual alerts (i.e., reaction time was lower for the tactile condition). Haptic interfaces have also been found to be useful in conveying a robot's spatial perception and reducing collisions (Barnes & Counsell, 1999; Diolaiti & Melchiorri, 2002; Zelek & Asmar, 2003). Aleotti, Bottazzi, Caselli, and Reggiani (2002) presented a teleoperation system that employs tactile feedback and gesture-based interaction for remote object exploration. According to Vogels (2004), synchronization is an important issue for multimodal interfaces. Vogels (2004) demonstrated that people were able to detect asynchrony between a visual and haptic stimulus at about 45 ms. However, it remains unclear how an operator's performance might be affected by temporal delay between visual and haptic stimuli.

2.1.3.5 Sensory Ego-sphere

A sensory ego-sphere robotic interface (see figure 3) is based on the concept that a visual representation of a discrete geodesic dome on which sensory data reside encompasses an ego-center (i.e., a robot) (Johnson, Adams, & Kawamura, 2003). A sensory ego-sphere interface is one solution to the coordination of multiple sensors into an intuitive display of the sensor data. A bird's eye view of the robot sitting inside its dome with all sensory data input embedded in the geodesic structure is provided on a screen display. The intent of the sensory ego-sphere interface is to reduce mental workload and increase SA. Johnson et al. demonstrated that the use of a sensory ego-sphere interface reduced teleoperators' mental workload while increasing SA compared with other traditional interfaces (albeit statistically significant results were not found). The user interface designs discussed so far are summarized in table 1.

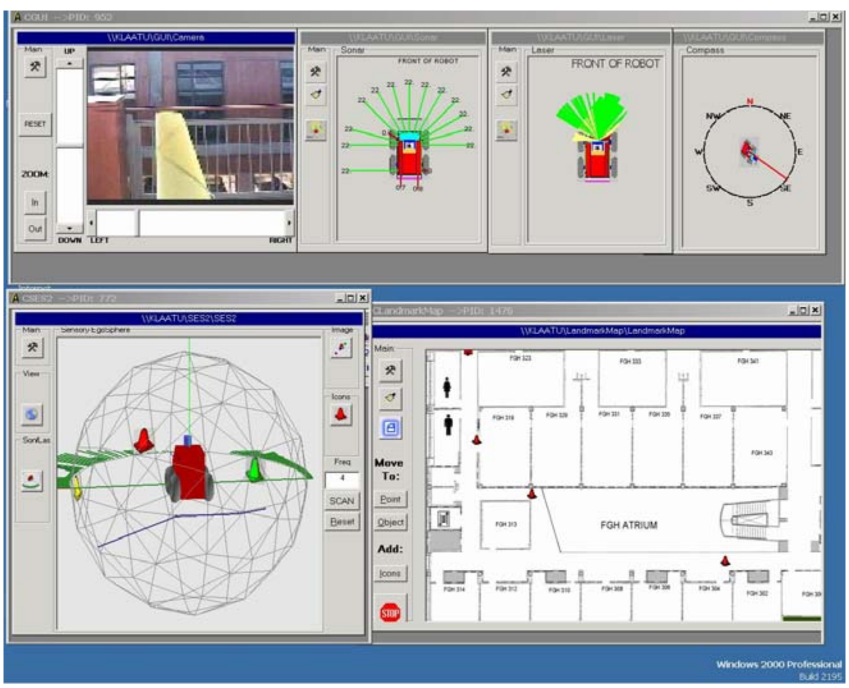

Figure 3. Sensory ego-sphere (adapted from Johnson, Adams, & Kawamura, 2003, with permission).

Table 1. Types of user interface and innovative for teleoperation.

Display	Advantages	Disadvantages/Caveats
Attitude Displays	• Attitude (i.e., pitch and roll) of a robotic vehicle may be easy to reference.	• May be limited to those involving confusing environments and stressful operations.
Stereoscopic Displays	• Faster and more accurate perception of the remote scene, enhanced detection of slopes and depressions, enhanced object recognition and detection, visual noise filtering. • Faster learning and faster task performance with fewer errors (for certain tasks).	• May increase motion sickness and operator's stress ratings.
Predictive Displays	• Reduce task performance time and errors (e.g., collisions).	• Disturbances may exist in the remote environment and make the model of the actual environment imperfect.
Multimodal Interfaces	• Enhance operator performance and alleviate visual workload. • Speech-based input is more effective than manual input in enabling operators to navigate through menus and select options more quickly and accurately. • Tactile alerts are more effective (i.e., reaction time is lower) in informing the operators than are visual alerts.	• Speech-based interfaces rely on voice recognition and may not be feasible in all environments (e.g., stealth missions).
Sensory Ego-sphere	• Provide a robot-centric display which presents sensory data in a more intuitive way. • May reduce mental workload and increase SA.	• Usability data are inconclusive • Some users may find it frustrating and stressful to use.

2.2 Control of Semi-autonomous Robots

Technological advances have expanded the capabilities of robotic assets as well as the nature and complexities of HRI. While much research has been devoted to the effects of increased automation in domains such as aviation and industrial settings such as nuclear and automotive plants, research on the effects of automation on robotic operations is not as robust.

Parasuraman, Sheridan, and Wickens (2000) provide a general model for the different types of automation and the levels of interaction between humans and automated systems. The authors suggest that with such a framework provided, designers can determine the level of automation that is optimal for any given human-machine system (i.e., what part of a system should be automated and to what extent). As Parasuraman et al. (2000) state, automation does not replace the work of humans; rather, it alters it. In the proposed model, there are 10 levels of automation across a four-stage view of information processing. In the lowest level of automation (level 1), there is no automated assistance and the human makes all decisions and takes all actions. As levels progress upward, the authority that automation has in making decisions and executing tasks increases. In the mid-levels of automation, the machine can make suggestions that may or may not be enacted by the computer, depending on the specific level of automation. At level 4, for example, the computer may provide suggestions only. At level 5, however, the computer may take action to follow a suggestion, provided it receives human approval. At the highest level, the computer acts autonomously and essentially does not regard human input. The levels of automation are applied to a four-stage model of information processing wherein the first stage (sensory processing) involves the receipt of information for various sources; the second stage (perception/working memory) involves the manipulation of received information; the third stage (decision making) where decisions are made based on results from stage three; and the fourth stage (response selection) where decisions are executed. The consequences that the levels of automation have on human performance (e.g., workload, SA, complacency, and skill degradation) during specific information processing stages can be delineated so that designers of automated systems can maximize performance and minimize adverse impacts of automation.

The effect of automation on human performance is widely studied. Parasuraman et al. (2000) discuss four human performance issues: mental workload, SA, complacency, and skill degradation. Several references to increased automation resulting in decreased mental workload are cited; however, the authors provide numerous examples in which automation can increase mental workload. With regard to SA, automation is also a double-edged sword. Although automation can provide more information in a timely manner, it can also deprive the human of knowing when changes occur in the status of the system and can prevent the human from developing an overall picture of a situation based on information that has been received and processed by the computer. Continual information processing without human intervention can result in complacency on behalf of the human. The impact of complacency occurs when the automated system malfunctions, and as the human slips in vigilance for monitoring the automated processes, the failure is not detected. Skill degradation, which also occurs when automation assumes a task previously performed by the human, is most notable when automation fails and the human must perform the tasks. As memory erodes and skills weaken over time, the ability for humans to intermittently do normally automated tasks decreases. The design of automated systems must reduce the consequences that they have on human performance. Kaber and Endsley (2004) also looked at the impact of automation on human performance and how certain forms of automation, *adaptive automation* and *intermediate levels of automation*, can relieve some of the negative impacts that automation has on human performance, such as workload and SA.

2.2.1 Interface Designs for Controlling Semi-autonomous Robots

Several interfaces have been developed and tested for controlling autonomous agents, all of which present benefits and challenges unique to each. Several studies have looked at the use of various interfaces, most of which are rather specific in terms of robot functions for which the interfaces control as well as the operational environment in which the robot performs. Steinfeld (2004) reported interviews of experts from the Robotics Institute at Carnegie Mellon University (CMU) and their recommendations, and they observed challenges for controlling fully and semi-autonomous mobile robots. The following is a partial list of the lessons learned:

- For multiple operators, consider giving veto power to the operator with a direct line of sight of the robot.

- Video and map views are useful, but it is not a requirement for both to be visible at the same time.
- A dashboard layout on the bottom of the screen to represent key information is useful.
- Controlling and navigating with 3-D interfaces can be difficult.
- Gauges and state information that changes color or pops up when a threshold is crossed are useful.
- There should be a central error and health summary.
- Integration and color coding information is useful.
- Communication delays must be accounted for.
- We should design for potentially substandard operator environments and conditions.

The following paragraphs review some novel techniques/devices for controlling (semi) autonomous robots. Potential utility and challenges are also discussed.

2.2.1.1 Cellular Phone and PDA

Sekmen, Koku, and Zein-Sabatto (2003) investigated the use of cellular phones to control the actions of robots. While participants indicated satisfaction in using cellular phones, their tiny screens and the ability to control more than one robot with one cellular phone were two challenges presented.

Lightweight control devices such as PDAs are also becoming increasingly popular for use in controlling robotic assets (Fong, Thorpe, & Baur, 2003; Fong, Thorpe, and Glass, 2003; Perzanowski et al., 2003; Quigley, Goodrich, & Beard 2004; Skubic et al., 2003). See figure 4 for an example of a PDA-based user interface. Keskinpala et al. (2003) looked at the use of touch-based (as opposed to stylus-based) PDA robotic interfaces that attach to the arm of the human operator like a wristwatch. Like the cellular phone interface, the amount of display space is at a premium, so screen display must be designed to maximize available space. Furthermore, touch-based PDAs must provide icons and screen items that are large enough to accommodate human fingers. PDAs also have limited software capacity and computing capabilities because of their smaller size. Fong, Thorpe, and Glass (2003) also investigated a PDA-based interface used for teleoperation wherein three modes could be accessed by the teleoperator: direct mode, image mode, and sensor mode. Because of the environment in which unmanned assets often work, the flexibility provided with each mode allows the teleoperator to choose which mode will best assist in directing the robot to fulfill its missions. The direct mode is simply real-time navigation through the ongoing picture of what the sensors detect from the robot's perspective. Image mode allows the operator to freeze the sensor images in order to create an overlay that will map the intended path(s) for the robot to take. Sensor mode allows the teleoperator to choose which sensor image will be displayed and used for navigation.

Figure 4. PDA-based user interface (adapted from Quigley et al., 2004, with permission).

2.2.1.2 Sketch Interfaces

Skubic et al. (2003) looked at the usability of sketch interfaces on PDAs for controlling robotic movements wherein the controller sketches an intended path for the robot to take by specifying the robot's positions relative to the landmarks (figure 5). The task representation is based on relative position instead of absolute position of the robot. Skubic et al. state that sketching is a natural and intuitive way to interface with the system since it simulates human-to-human communication in which hand-drawn route maps are effective in conveying geographic information. However, an obvious limitation is that the system needs to correctly and consistently interpret the stylus markings of the user, which may vary from person to person and even from occasion to occasion for the same user. The authors did not address the effect of sketch interface display on robot performance but focused on the usability of the interface for operators. Although users indicated in this study that they were satisfied with the sketch interface, they expressed concern about the small size of the PDA on which the interface resided. Skubic and her colleagues have also developed a sketch-based interface to control a team of robots, and they performed a usability study (Skubic, personal communication, September 23, 2005). The researchers reported that the sketch interface appeared to be easy to learn and use.

Other sketch interfaces have been developed for robot navigation (Setalaphruk, Ueno, Kume, & Kono, 2003) and for military strategic planning purposes (Ferguson, Rasch, Turmel, & Forbus, 2000). Users can create course-of-action diagrams using the qualitative spatial reasoning techniques. However, the utility of the sketch interface has not been demonstrated in a military environment where robotic assets are involved.

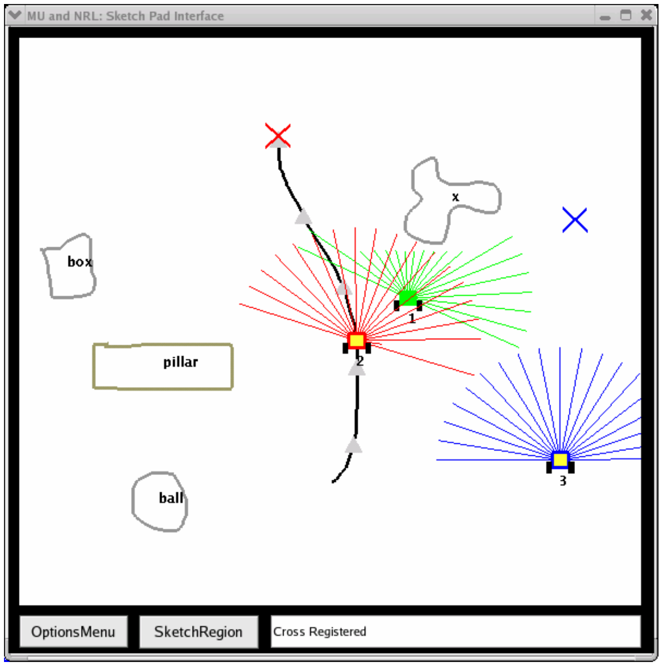

Figure 5. Sketch interface (adapted from Skubic et al., 2005, with permission).

2.2.1.3 Multimodal Interfaces

2.2.1.3.1 Natural Language and Gestures

The use of natural language and gestures to communicate intentions to robotic assets eases the effort required for learning the tactics for successful HRI. Perzanowski et al. (2003) state that there are two communication settings in which human-robot teams exist: basic settings support face-to-face communication gestures such as hand and eye movements and non-basic settings (generally arising from remote locations) require other modes of interaction. A multimodal interface allows the successful communication between human and robot across basic and non-basic settings. In a study of multimodal interfaces in a non-basic setting, Perzanowski et al. (2003) implemented natural gestures such as arm movements and pointing motions as well as verbal expressions directly to the robot or indirectly via the use of a PDA. Furthermore, commands could be qualified with a touch-based PDA (e.g., move chair A wherein chair A is selected by the operator through the touch-based PDA). Findings from this study suggest that verbal expressions, when applicable, are much more widely used than gestures. Furthermore, variability between verbal expressions existed and terse verbal commands (e.g., “move here”) were often supplemented with use of the touch-based PDA interface for clarification (e.g., operator indicates where “here” is).

2.2.1.3.2 Vibro-tactile Displays

Zelek and Asmar (2003) have investigated the incorporation of a secondary tactile modality to the existing visual cues used to guide robotic movement. The authors argue that not only will such a secondary modality enhance the operator’s ability to receive, process, and act on incoming sensor information, the development of a tactile receptor will allow visually impaired individuals to manipulate robotic assets within any given environment. Furthermore, vibro-tactile displays can be used in environments where visual cues are not available to the operator (e.g., aviation in instrument meteorological conditions) or in settings where high noise and low visibility conditions prevail. The authors of this study evaluated the effectiveness of a vibro-tactile device (a glove with individually vibrating motors) that receives visual information. The visual information is sent to the glove and is transformed into a series of patterns of vibrations that correspond to the navigational cues in the environment that are detected from the image sensors. Challenges presented to designers of vibro-tactile interfaces, specifically those that interpret visual information, arise from the limited bandwidth that is available in this modality. The complexities of vision simply cannot be entirely duplicated through tactile interfaces; therefore, engineers must choose which visual cues are selected for transfer into a tactile representation. A navigation lexicon assists in transforming visual information into tactile information by segregating the environment into different facets (e.g., spatial prepositions such as *down*) and sub-facets (e.g., compounds such as *down to* and intransitive prepositions such as *downward*). Outfitted with vibro-tactile gloves, participants of this study navigated through a small indoor obstacle course. Although the sample size was too small for statistical inferences, the data do suggest that use of a vibro-tactile interface has potential for use in personal and robotic navigation tasks.

A detailed survey of multimodal interfaces and their potential use for robotic control is presented in the next section. The user interface designs for controlling semi-autonomous robots are summarized in table 2.

Table 2. Types of user interface and innovative designs for controlling semi-autonomous robots.

Display	Advantages	Disadvantages/Caveats
Cellular Phone & PDA	• Enhanced portability.	• Screen sizes and the ability to control more than one robot with one device may be problematic. • Touch-based device must provide icons and screen items that are large enough to accommodate fingers. • Limited software capacity and computing capabilities because of their smaller size.
Sketch Interface	• Natural and intuitive way to interface with the system by specifying paths using land-marks. • Task representation is based on relative position instead of absolute position of the robot.	• System needs to correctly and consistently interpret the stylus markings of the user.
Natural Language and Gestures	• Ease the effort required for learning the tactics for successful HRI.	• Gestured-based interfaces rely on cameras that may be limited by lighting conditions and FOV.
Haptic/Vibro-tactile	• Enhance the operator’s ability to receive, process, and act on incoming sensor information. • Reduce collisions. • Can be used in environments where visual cues are not available to the operator or in settings where high noise and low visibility conditions prevail.	• Limited bandwidth that is available in this modality. • The complexities of vision cannot be entirely duplicated through tactile interfaces; therefore, engineers must choose which visual cues are selected for transfer into a tactile representation.

2.3 Human-Robot Teaming

The concept of human-robot teaming is based on the interdependence between the human operator and the robot for all that is associated with conducting a robot-assisted mission (e.g., defining the mission and tasks, allocating tasks, two-way feedback between operator and robot, controller input, analysis of information, etc.). As technology increases robotic capabilities, an understanding of the concepts and issues that affect human-robot teams is essential so that given any operational environment, robot-assisted missions are performed successfully with full exploitation of all technological and human capabilities and with minimal adverse impact. Identifying the concepts and issues that define the nature of the human-robot team is just beginning to take shape. Researchers and many from the robotics user community are beginning to step back and study the relationship between the human operator and the robot across many operational fields. In a study that combined the insights from several industries, Burke, Murphy, Rogers et al. (2004) call for the need for a design that creates "synergistic teams" of robots and the human controllers.

In Burke, Murphy, Rogers et al. (2004), a taxonomy of the possible relationships between the human operator and the robot was introduced, and the human-robot relationship was described as 3-D. The *human-robot ratio* refers to how many humans are assigned to a robot, as discussed earlier. The *spatial relationship* defines the level of intimacy (closeness) between the human and robot as well as the point of view. The *authority relationship* determines who (if either) is a supervisor, operator, bystander, etc. Burke, Murphy, Rogers et al. (2004) also address communication between the operator and robot since it is a central issue in HRI. The authors assert that there are two forms of communication, each of which has several different communication modalities. *Direct human-robot communication* involves such modalities as speech and gesture while *mediated human-robot communication* arises from graphical user interfaces and VE. In terms of interfaces, which are drivers of communication, the authors suggest that designing interfaces that promote the efficient use of time and have a high tolerance for workload is essential. Other areas relating to communication that were considered in need of basic research are the "effects of delays, poor synthesis of information, and dynamic interactions" (Burke, Murphy, Rogers et al., 2004, p. 7). The concept of social relationships between human operators and robots is in need of investigation since it is not clear what the effects of various social relationships are on the efficacy of HRI activities.

2.3.1 Human-Robot Ratio

The number of robots that can effectively be controlled by one person, referred to as the human-robot ratio, is a design consideration that is driven by several factors. In an article addressing the use of robots for search and rescue, Murphy (2004) discusses the 2:1 human-robot ratio, which is driven by logistics in transporting, maintaining, and operating the robot and the raw capabilities of the robot. The author notes that additionally, specialists are often involved in operating the robot while other members of the human-robot team specialize in interpreting robot data for overall mission execution (not including maintenance and operation of the robot). Murphy also suggests that increased automation of the robot and more sophisticated sensors do not necessarily impact the number of people assigned to a robot but reduce the operator workload (the author calls it "reducing the role"). Although Murphy's discussion is centered on the use of robots for search and rescue, the concepts presented in the article are general and relevant to the employment of robots in other operational settings.

Chen et al. (2005) examined how robotic operators' reconnaissance performance differed, depending on the type and number of assets available. The robotic assets used in this experiment included autonomous UGVs, semi-autonomous UAVs, and teleoperated UGV (Teleop). The results suggested that giving robotic operators additional assets may not be beneficial. When given three robots, participants failed to detect more targets than when given only the UGV or UAV. Moreover, fewer participants were able to complete the mission in the allotted time. Target detection was poorest for the teleop vehicle, most likely because the demands of remote driving. These findings are consistent with those of other robotic control studies (Dixon et al., 2003; Rehfeld, Jentsch, Curtis, & Fincannon, 2005). In Dixon et al., pilots detected fewer targets with two UAVs than with a single UAV. Automation also appeared to benefit pilots' target detection performance. Rehfeld et al. compared one to two UGVs and found that the additional UGV did not enhance the target detection performance of the operator(s). The results of Rehfeld et al. are further discussed in the following section. The findings of Dixon et al., Rehfeld et al., and Chen et al. (2005) suggest that, regardless of the types and homogeneity of the robotic platforms, additional assets do not appear to be beneficial for reconnaissance types of tasks.

2.3.2 Human-Controller Teamwork

Although the human-robot ratio is seen as "a non-reduced fraction, with number of humans over the number of robots" (Yanco & Drury, 2002, p. 114), the concept of human controller team-work arises when the human-robot ratio is variable. Human controller teamwork involves the interactions and coordination that take place when the human-robot team is expanded to more than one operator-robot dyad. Rehfeld et al. (2005) examined cost benefits of various HRI teaming concepts by conducting a laboratory experiment in a scaled MOUT setting. Rehfeld et al. found that giving one more robotic asset to a single operator or a two-person team did not enhance the individual's or team's target detection performance. In fact, in difficult scenarios, the single operators actually performed worse with two robots than with one. On the other hand, the two-person teams performed more than twice as well as the one-person condition in those difficult scenarios, regardless of how many assets were used. These findings echoed what have been observed in the field (e.g., using robots for search and rescue efforts) in Murphy (2004) that remote perception is still one of the most fundamental challenges for robotic operators.

Yanco and Drury (2002), in combining the concepts, theories, and ideas from research in HRI, human-computer interaction and computer-supported cooperative work, present a set of taxonomies by which the field of HRI can be defined. Team composition is one such taxonomical category that Yanco and Drury (2002) briefly describe. In a discussion of the various human-robot operational configurations, the authors present questions that are central to understanding the dynamics of human-robot teaming such as whether operators work together or independently when issuing commands to robots, and what the effect is on robot workload. The authors present eight human-robot team configurations and indicate that for each configuration, a set of questions arises, the answers to which can characterize the nature of the human-robot work relationships and can reveal issues that relate to the performance of human-robot teams. The eight configurations presented are

- One human-one robot wherein an individual commands the actions of one robot;
- One human-robot team wherein one individual sends commands to multiple robots which, in turn, must sort and declassify the operator's commands;
- One human-multiple robots wherein an individual sends commands independently to several robots;
- Human team-one robot wherein multiple humans coordinate among each other to send commands to a robot;
- Multiple humans-one robot wherein the humans independently send commands to one robot which, in turn, must sort and deconflict those commands;
- Human team-robot team wherein multiple humans coordinate to send commands, and multiple robots coordinate to sort and deconflict those commands;
- Human team-multiple robots wherein a team of humans coordinate to send individual commands to individual robots, and finally
- Multiple humans-robot teams wherein humans send commands independently to a team of robots which, in turn, must sort and deconflict those commands.

Murphy (2004) also addresses the issues of human controller teamwork in terms of information flow (who receives what, the timing of information delivery, and whether it is linear) and distributed communications. The existence of distributed communications presents many challenges to human controller teamwork. In a distributed environment, information flow changes from a one-way linear movement to a more fluid back-and-forth adaptive movement. Because of the nonlinear flow of information, responsibility of information and action requests to the operator and to the robot can result in several conflicts. Furthermore, distributed communications complicate the development of an information display since it must be suited for all consumers of information.

In an attempt to envision the future of human-robot coordination, Woods et al. (2004) bring together the issues facing human-robot teaming (as well other HRI issues) across different operational environments. With a focus on USAR and chemical, biological, or radiological incidents, the authors address the impact that technological developments, industrial needs, and the constraints of human cognitive processing have on the design of robots, the organization of human-robot teams, and teamwork. Three perspectives were brought together for a robust treatment of the issues facing human-robot teams: (a) roboticist, (b) cognitive engineer, and (c) practitioner. The unifying concept among these perspectives is how, in light of changes in robotics, do we "exploit new capabilities or work around new complexities" (Woods et al., p. 2). From the combined discussions from each perspective, it is concluded that the adaptability of individuals working in coordinated human robot efforts is crucial. When the human-robot relationship suffers a breakdown (e.g., because of technological limits or malfunctions), it is the adaptability of human-robot ensembles working in a team environment that can effectively sustain a mission until the breakdown is resolved. Woods et al. also addressed the impact that the responsibility of a given operator has on the organizational architecture of a human-robot team. For the operator who bears ultimate responsibility for the outcome of a robotic mission, it is essential that this individual be allowed to monitor the data input and track the intent of other human-robot ensembles.

Another interdisciplinary attempt to further understand the issues facing HRI divides teamwork into two areas, architecture and task allocation (Burke, Murphy, Rogers, et al., 2004). Architecture refers to the organization of the human-robot team (as configured in any of the combinations listed) so that the benefits of teamwork are maximized; the operational setting may require an authoritarian or a democratic structure, for example. In terms of task allocation, human-robot teams must assign tasks that maximize the capabilities of all team members (robot or human) at any given time. Burke, Murphy, Rogers et al. state that task allocation is not likely to be static since capabilities of team members can change as a result of numerous factors, such as individual workload and the nature of the tasks being assigned. Schreckenghost (1999) investigated the effectiveness of a software interface designed to perform *traded control*, a form of supervisory control in which tasks and task objectives are switched between the robot and the human operator.

In addressing the operations of autonomous workstations on other planets and the accompanying controller teams on Earth, Malin (2000) presents issues that are central to an effective multiple human-robot teams. Malin (2000) introduces the concepts of tight and loose coordination as requirements for effective team-oriented operations wherein the robot agents switch in and out of autonomy. The autonomous workstation on other planets, for example, reverts from working independently to depending on human controller input when the workstation encounters problems or when new situations arise in which solutions do not reside within the robotic agents. Loose coordination involves such activities as keeping team members current via various media (e.g., notes, voice messaging, etc.). In designing and developing autonomous agents with a social capability (e.g., an ability to relay desire, intents, conclusions to human operators), it is necessary to consider how the agents' communication capabilities affect the team in which they interact. For example, Malin (2000) discusses the need for a common ground and "mutual group awareness" (p. 255) in order to facilitate effective human-robot teamwork. Common ground is often achieved through a shared interface that "works with the same representation of information" for the autonomous agents and the controllers (Malin, 2000). The design of autonomous agents should support a common ground and group awareness so that they provide information relating to (a) beliefs and assessments, (b) desires, goals, and priorities, (c) current intentions, plans, and procedures, (d) capabilities for low level sensing and acting, and (e) capa-bilities for communicating and using help. In addition to common ground and shared knowl-edge, human-robot teams must be able to engage in cooperative negotiation in order to resolve conflicts (e.g., new team member perspectives) that arise in new or unfamiliar situations. Groupware[4] solutions must incorporate only the dissemination of essential information in order to ensure that negotia-tions and team coordination are efficient. An observational study of robots used in a USAR mission (Burke, Murphy, Coovert, et al., 2004) reinforces the notion that a common operational picture, shared mental models, and efficient communication flow are necessary for effective human-robot teams. In this study, it was found that team members were attempting to develop shared mental models in order to increase their SA. Furthermore, frequent communication between team members was correlated with high scores of SA.

3. Multimodal Auditory Control and Display Technologies for the U.S. Army HRI

3.1 Background

Within the last several years, the introduction and use of complex equipment and systems made robotics systems relatively complex cognitive environments where Soldiers must simultaneously monitor multiple displays, operate multiple controls, and process large amounts of information. The potential for an increasing span of control (fewer people controlling more robots) would make these tasks still more cognitively demanding. When used to supplement and support conventional manual controls and visual displays, multimodal technologies have the potential for providing the Soldier with a means of reducing workload and improving SA in robotic control and display systems.

Multimodal technologies such as spatial auditory displays, speech synthesis, and haptic (tactile) displays can provide system display information to the Soldier, freeing his or her eyes for other tasks. Automatic speech recognition (ASR) can provide a hands- and eyes-free method for providing voice output for system C2. Alternate microphone technologies such as throat and bone microphones can be used in conjunction with ASR systems to ensure proper processing of Soldier speech commands in noisy environments.

The purpose of this section is to describe how these particular multimodal display technologies can be used by themselves or integrated with each other to fit into the Robotics Collaboration Army Technology Objective (ATO). Within this section, the authors describe audio displays and controls, specifically spatial audio displays and ASR. Alternate bone conduction and throat

[4] That is, software that can be used by a group of people who are working on the same information but may be distributed.

microphone technologies that can enhance the intelligibility of ASR commands are discussed. Next, speech synthesis is described not only as an information display that can provide system warnings but also as a means to provide speech feedback when used in conjunction with ASR systems. In the final section, the authors describe haptic display interfaces that are relevant to the HRI environment.

3.2 Spatial Audio Displays

In spatial audio displays, also known as 3-D audio displays, a listener perceives *spatialized* sounds that appear to originate at different azimuths, elevations, and distances from locations outside the head. Three-dimensional audio displays permit sounds to be presented in different horizontal, vertical, and distance locations that are meaningful to the listener.

Earphones are often used to present spatial audio cues (loudspeakers may be used, although their use may be problematic, as can be seen in Shilling & Cunningham, 2002). Before the audio cues reach the earphones, they are filtered through computerized sound filter functions known as head-related transfer functions (HRTFs). These HRTFs provide the sound with specific time, intensity, phase, and reverberation cues. The result is sound that upon output is heard at different locations in space. A head tracker is often used to provide a stable reference point for the audio cues. Because each sound is presented in a different spatial location, listeners may selectively attend to more than one sound at a time. Three-dimensional audio actually enhances listener performance in situations when listeners must listen to several audio messages that occur simultaneously, such as in tasks involving monitoring communications on multiple radio channels. Wenzel, Wightman, and Foster (1988) describe the theory and technique of the synthesis of localized sound and the psychophysical validation of HRTFs and they discuss several applications.

Although 3-D auditory displays have not been integrated into current U.S. Army systems, some applications have been suggested, including monitoring multiple radio communications channels, waypoint navigation, system location and malfunction warnings, threat warnings, and teleoperation of UV. In cockpit applications with helmet- or head-mounted visual displays with a limited FOV, 3-D audio can be used to direct the attention of the pilot to critical events occurring outside the visual FOV. Haas, Gainer, Wightman, Couch, and Shilling (1997) investigated the use of 3-D auditory displays in helicopter cockpit radio communications tasks. The U.S. Air Force experimented with the use of 3-D auditory displays in providing fixed wing aircraft with waypoint information (McKinley, Ericson, & D’Angelo, 1994).

Several researchers performed basic research with applications to military systems. Folds and Gerth (1994) explored the monitoring of multiple simultaneous independent sound sources, demonstrating the value of spatial auditory signals in reducing visual search time. Elias (1995) examined the effects of dynamic auditory preview in a visual target aiming task and explained the relationship between spatial auditory preview and its visual correlate. Endsley and Rosiles (1995) explored the use of vertical auditory localization for spatial orientation, for use in reducing pilot spatial disorientation. Several researchers showed the effectiveness of spatial auditory cues in enhancing visual search performance. These included Perrott, Cisneros, McKinley, and D'Angelo (1995), Strybel, Boucher, Fujawa, and Volp (1995), Elias (1996), and Fujawa and Strybel (1997). Lee (1997) explored multi-channel auditory search to define the optimum number of simultaneous spatial auditory sources for good listener performance. Brungart (2000) investigated the effectiveness of several speech-based distance cues in controlling the perceived distance of virtual audio speech, to recommend effective distance cues for use in spatial audio displays. Finally, Ericson (2000) explored the simulation of linear auditory motion over headphones and then described several attributes of moving sound sources that enable a listener to judge the velocity of a dynamically moving spatial sound, which is useful for providing a veridical simulation of auditory motion over headphones.

3.3 ASR

Speech recognition is also known as ASR. Rabiner (1994) defined ASR as the process of extracting the message information in a voice signal so as to control the actions of a machine in response to spoken commands. With ASR, spoken words are first digitized and then matched against coded dictionaries in order to identify them. Once they are identified, the resulting information in the spoken output can control the actions of a system or machine in response to spoken commands (Haas & Edworthy, 2002).

The first ASR systems were speaker dependent, meaning that a speaker entered samples of all the words that existed in the system dictionary to "train" the system. Currently, most ASR systems are speaker independent, recognizing words in their vocabulary without any speaker training.

Several researchers have explored speech recognition in military applications. Vidulich and Bortolussi (1988) examined speech control in a single-pilot scout/attack helicopter, demonstrating the use of objective and subjective human performance ratings and described the importance of using multiple assessment techniques to assess speech recognition in demanding environments. These researchers found that although the operational reliability of speech controls could be improved, reliable speech controls could enhance the time-sharing efficiency of helicopter pilots. Fisher (2000) related lessons learned while integrating speech control into embedded systems with no keyboard, mouse, or monitor. He listed critical issues involved in incorporating speech control into an embedded system and described the design of one such system in which the hands-free interface is natural and easy to use. Haas, Shankle, Murray, Travers, and Wheeler (2000) explored the use of ASR with spatial audio communications in a simulated tank environment and found that ASR and spatial audio displays have no deleterious effect upon each other when integrated into a simulated tank environment and have great potential as technologies of interest in high noise, stressful tank environments. Noyes, Baber, and Leggatt (2000) described the use of ASR in tanks and armored fighting vehicles and discussed successful applications in which ASR was used. Williamson and Barry (2000) described the design, implementation, and evaluation of a prototype speech recognition interface to the inclusion in a future upgrade.

One important finding in all these studies is that principles of design or usability are important. Karis and Dobroth (1995) suggested that a successful human factors design of a speech recognition system should involve an early focus on the users of the system and the tasks they will perform, collecting performance data via simulations and prototypes, iterating the process of collecting data, identifying problems, and modifying the system. Nielsen and Molich (1990) suggested basic design principles to optimize the human factors design of ASR systems, including the use of simple and natural dialogue, minimizing demands on user memory load, providing feedback, providing shortcuts, and providing clearly marked exits.

One limitation of ASR is that speech recognition systems may experience loss of message intelligibility in noisy environments, where ambient noise might interfere with the transmission and reception of Soldier speech commands into an ASR system (Noyes et al., 2000; Myers & Cowan, 2003). The following section, which concerns bone conduction and throat microphones, describes some alternate interface technologies that might be useful in enhancing the performance of speech recognition systems in noisy environments.

During the next several years, the number and type of applications of speech recognition will increase dramatically, and attempts will be made to automate fairly complex operations. As noted by Karis and Dobroth (1995), a factor of great importance in the success of future systems is the overall design of these systems with respect to their capabilities for interacting with users. Future systems must take human conversational behavior into account, as well as principles of human factors design. The effectiveness of ASR systems and their acceptance by the Soldier and by other users will depend upon the extent to which ASR systems have been designed to accommodate some of the flexibility inherent in human communication, rather than on an attempt to force users to follow a script or vocabulary in which their input is rigidly constrained.

3.4 Throat and Bone Conduction Microphones

3.4.1 Background

Soldiers using radio headsets in robotic control unit operations may experience loss of message intelligibility, especially in noisy vehicles or dismounted operations where ambient noise might interfere with the transmission and reception of spoken communications. Environmental noise might be a disadvantage in the human-robotic interface, creating potential interference when Soldiers communicate to others to coordinate robotic control, when ASR systems are used for robotic C2, or when Soldiers listen to audio target or positional information transmitted by the robotic OCU. Bone conduction and throat microphone technologies might alleviate some of the ambient noise problems because they can isolate the speech signal from environmental noise, thus preventing degradation of the speech signal sent into and received from the communications system.

3.4.2 Bone Conduction Headsets

Bone conduction headsets enable the user to send and receive spoken communications. With airborne sound, when someone speaks, the sound travels through the ear canal to the eardrum, which vibrates the small bones of the middle ear and transforms sounds into nerve impulses that are interpreted by the brain as sound. With bone conduction reception, sound waves are received as vibrations on the skull or cheekbones, which bypass the outer ear and proceed to the middle and inner ear where they are translated into nerve impulses that are interpreted by the brain as sound. With bone and air conduction, sound waves are perceived in exactly the same way: as nerve impulses interpreted by the brain. Figure 6 is a TEMCO[5] bone conduction headset with a standard (non-bone conduction) boom microphone.

Figure 6. TEMCO bone conduction headset.

Sound signals received by bone conduction are not exactly the same as those received through air transmission. Because bone vibrations are transmitted through bone or skin, the high frequency elements may be attenuated (reduced). In order to produce high-quality, under-standable sound, some bone conduction headphone manufacturers use an equalizing circuitry that restores the high-frequency signal and makes the sound more intelligible to the listener.

With bone conduction microphones, the sound waves generated by the talker are generated as vibrations on the skull or cheekbones. These vibrations are detected by transducers that have close contact with the bone of the skull, making them relatively resistant to transmitting environmental noise. Bone microphone transducers use contact pickups, which are microphone elements designed to detect sound waves in a solid medium such as bone, rather than in the air. Contact pickups are most often piezoelectric devices, although some inertial and mechanical microphones exist. Figure 7 is a TEMCO bone conduction headset with a standard (non-bone conduction) boom microphone.

Bone conduction headsets offer several advantages to the Soldier. Because sound is transmitted through the bones rather than through air, ambient noise will interfere less with the transmitted sound. Because the microphone and receiver work by "hearing" with the bone structure of the head, the ears are completely free and open to hear surrounding sounds or free to be covered and protected against background noise.

[5]TEMCO is not an acronym.

Figure 7. TEMCO bone conduction headset with a standard (non-bone conduction) boom microphone.

3.4.3 Throat Microphones

A throat microphone is a skin vibration transducer, which is worn around the throat and is actuated by vibrations of the larynx. Hypothetically, the throat microphone reduces transmitted environmental noise because of the close contact of the transducer with the throat skin. Throat microphone transducers use contact pickups, which are microphone elements designed to detect sound waves in a solid medium rather than in the air. The contact pickups are most often piezoelectric devices, although dynamic microphones have sometimes been used for this purpose. One source described a contact microphone in the form of a flexible strip, which has gained favor in some sound reinforcement circles (Davis & Jones, 1990). As with bone conduction microphones, throat microphones are used when background noise would obscure the sound of speech. Figure 8 shows a Blue Kangaroo Technologies[6] throat microphone.

Figure 8. Communication task performance.

[6] A company that manufactures specialty audio microphones and headsets.

3.4.4 A Comparison of Bone Conduction and Throat Microphones

Appendix A presents a comparison of significant features of bone conduction headsets and throat microphones produced by major manufacturers. This table and the commentary describing the table are based on manufacturer claims regarding their products. Because an objective comparison and evaluation of headsets have not yet been completed or published, the extent to which these claims are true has not been established. However, the reader can use this text and appendix A to obtain an idea of what different manufacturers have to offer.

As seen in appendix A, several bone conduction headsets and throat microphone manufacturers claim that their headsets are ruggedized and waterproof. Most manufacturers of bone conduction headsets and throat microphones claim compatibility with a wide variety of two-way radios. Many manufacturers also claim precise transmission quality in whisper mode, which ensures that persons in the immediate area cannot hear the speech transmission of the user. Manufacturers of bone and throat equipment claim compatibility with helmets and gas masks and claim a wide variety of push to talk (PTT) switches mountable everywhere from the chest to the wrist and hand. One PTT system incorporates an in-line disconnection to ensure swift and complete disconnection from the PTT mechanism in case of potential accidents or enemy attacks.

Although all bone conduction headsets have bone conduction receivers, the characteristics of headset transmission microphones differ between manufacturers. Some headset manufacturers recommend the use of bone conduction microphones, but one manufacturer (New Eagle, now Atlantic Signal, LLC) recommends a conventional acoustic boom microphone. This manufacturer claims that the acoustic microphone has better sound quality than that provided by a bone conduction transmitter.

New Eagle also claims that their bone conduction headsets allow reception of monaural or stereo radio transmissions. Thus, if the user chooses, s/he can receive and monitor two separate radio transmissions (each transmission is received at a different location on the skull). However, the separate skull transmissions are not perceived as coming from separate locations when processed by the perceptual mechanism of the brain because bone conduction does not process through separate auditory channels of the two ears. Thus, listening to bone conduction stereo may sound jumbled and somewhat confusing since both channels will be heard as if occurring at the same location. Still, stereo bone conduction may be an advantage for people who must monitor several radio channels simultaneously and who want the advantages of bone conduction.

Throat microphone manufacturers claim a wide variety of advantages for their products. Some manufacturers claim compatibility with mobile phones. Others claim a compatibility with confined space and hazardous materials operations, which may be an advantage for Soldiers in a cramped robotic OCU. Several manufacturers claim a high isolating compatibility for environmental noise. However, as described later in this report, Mr. Pete Fisher, an ARL researcher, found that several throat microphone units tended to transmit high levels of environmental noise along with speech.

3.4.5 Bone Conduction and Throat Microphones in the HRI Environment

This section discusses the extent to which bone conduction headsets and throat microphone lend themselves to the HRI environment, including the use of the robotic OCU. Characteristics of the HRI environment of greatest interest in this report are robotic control units mounted in vehicles or used in dismounted operations, which are environments with a potential for high levels of ambient noise.

The use of bone conduction and throat microphones for the robotic interface has not been evaluated by equipment manufacturers or by university, Government, or industrial researchers. Researchers at the U.S. Air Force Research Laboratory and the Defense Science and Technology Laboratory (DSTL) in the United Kingdom (U.K.) report no research exploring the use of throat microphone and bone conduction headsets for robotic applications. However, this does not mean that research in this area does not exist. A detailed literature search revealed that two promising sources of information exist at ARL's Computer and Information Sciences Directorate and at the Stanford Research Institute (SRI).

Pete Fisher is currently conducting an evaluation of bone conduction and throat microphones, primarily for use with ASR. Although his evaluation has not yet been completed, Mr. Fisher reported that several of the throat microphones tested showed a lack of external noise rejection, meaning that they transmit local acoustic noise about as well as they detect speech (Fisher, personal communication, November 21, 2003). Mr. Fisher suggested that one solution to the problem of lack of noise isolation would be to design a noise cancellation system that combines the output from several speech sensors (i.e., a conventional microphone, a throat microphone, a bone conduction microphone, an electromagnetic sensor, and lip reading), to produce noise-free speech in a noisy environment. Mr. Fisher reports that an ARL Small Business Innovation Research (SBIR) contractor, Intelligent Automation, is producing such a system. At present, Mr. Fisher reports that the SBIR product looks promising. Again, work on this project has not been completed, so final results are not available for evaluation.

As part of the RCTA, Dr. Greg Myers and colleagues from SRI used two microphones to improve ASR performance in noisy environments (Myers & Cowan, 2003). Dr. Myers used a standard combat vehicle crewman (CVC) headset microphone as well as the ARL physiological microphone (a throat microphone developed by ARL scientist Mike Scanlon) to process operator speech in an ASR system installed in a ground vehicle. Dr. Myers conducted an experiment in July 2003 in Madera, California. In this experiment, speech data were collected from eight subjects riding in a high mobility multipurpose wheeled vehicle (HMMWV) while traveling at 55 mph with windows open. The noise level in the HMMWV ranged from 96 to 100 dB sound pressure level (SPL). The subjects spoke a total of 295 utterances. On the first run, the phrase error rate with the headset microphone alone was 33.6%. When a probabilistic optimal filtering (POF) algorithm was applied to the conventional CVC microphone, the error rate on the microphone was reduced to 9.5%. When the ARL throat microphone was added to the conventional CVC microphone and the POF algorithm (the POF algorithm incorporated both microphones at this point), the phrase error rate was reduced to 7.1%. Dr. Myers' data demonstrated that the addition of a filtering algorithm and a throat microphone can contribute to better ASR performance than that obtained with a conventional microphone alone. Dr. Myers noted that data were collected at only one single noise level and feels that future testing would be beneficial, especially if conducted at different high noise levels. Dr. Myers noted that he felt that improvements in recognition performance offered with the additional throat microphone and processing algorithm would be even greater at higher noise levels.

3.4.6 Bone Conduction and Throat Microphone System Conclusions

The use of bone conduction headsets and throat microphones could increase the performance of message transmission and reception for robotic operations, especially in noisy environments. However, receiver and transmitter performance has not yet been fully evaluated. For example, some microphones, especially throat microphones, have been observed to lack isolation to external noise. Pete Fisher suggested that one solution to this problem would be to design a noise cancellation system that combines the output from several speech sensors to produce noise-free speech in a noisy environment, as is being accomplished through an ARL SBIR. Results look promising, but work on this project has not been completed and results are not available. Dr. Greg Myers demonstrated that when multiple microphones are used, filtering algorithm software must be produced to incorporate all microphones in tandem.

In conclusion, throat and bone conduction performance is a combination of the quality of several factors, including the bone and throat transducer(s) and the signal and the algorithm used to process that signal. Some of the operations such as signal amplification and equalization, which in the past were usually accomplished by hardware, may now be implemented in software. Bone and throat technology performance should be considered at a *system* level, incorporating transducer and algorithm, rather than relating to hardware alone.

3.5 Speech Synthesis

3.5.1 Background

Speech synthesis is defined as the process of creating a synthetic replica of a voice signal in order to transmit a message from a machine to a person for the purpose of conveying the information in the message (Rabiner, 1994). A speech synthesizer is the software or hardware that is capable of rendering the artificial speech produced by the synthesis process. The speech output by the synthesizer can have a human- or machine-like quality, depending on the application.

The range of speech synthesis applications is growing rapidly. Speech synthesis may operate as text to speech (TTS), in which a text message is transmitted into speech, which is then heard by the user. An HRI-related example of this application would include the generation of spoken prompts originating from the robot during system diagnostics. TTS synthesis has advanced to the point that virtually any ASCII (American standard code for information interchange) text message can be converted into fluent speech, providing an intelligible message to the listener. Speech synthesis may also operate as speech to speech, where the user's speech query regarding robot system status would directly trigger a synthesized report from the robot. Speech translation devices also may work this way, where the user's speech may directly trigger a synthesized speech output in a different language. Speech synthesis is even used in gesture-to-speech systems, such as the iCommunicator[7] which translates American Sign Language gestures to synthesized speech and translates speech or text to a proprietary form of video sign language (VSL).

3.5.2 Composition of a Speech Synthesis System

A speech synthesis system is composed of a front end and a back end. As can be seen in figure 9, which illustrates a TTS system, the front end performs high-level synthesis, acting as a user interface by taking input in the form of text (other types of synthesis systems use speech or gesture), and outputting a symbolic linguistic representation. The back end, which performs low-level synthesis, takes the linguistic representation and produces synthesized speech as acoustic waveforms.

The front end has two tasks. The first is to take the raw text, speech, or gesture and convert it into written word equivalents, which is known as text normalization, pre-processing, or tokenization. The second task is to assign phonetic transcription to each work and to divide and mark the text into various prosodic units such as phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is known as text-to-phoneme or grapheme-to-phoneme conversion. Together, the phonetic transcriptions and information about prosodic units combine into the symbolic linguistic representation that is produced by the front end.

The back end, which is also referred to as the synthesizer, takes the symbolic linguistic representation produced by the front end and converts it into actual sound output. Two main technologies used for generating synthetic speech from the back end are known as concatenative synthesis and formant synthesis.

[7]iCommunicator is a trademark of PPR Direct, Inc.

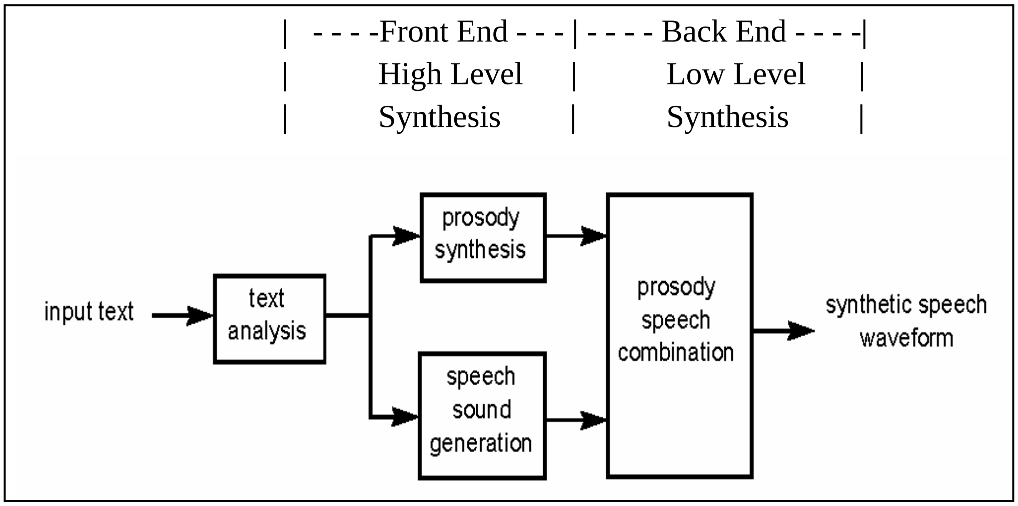

Figure 9. A TTS conversion system.

3.5.3 Types of Speech Synthesis

3.5.3.1 Concatenative Synthesis

Concatenative synthesis is based on stringing together (concatenation) segments of recorded speech. The use of recorded speech produces the most naturally sounding synthesized speech. However, the natural variation in speech and some of the automated techniques for segmenting the waveforms may produce audible glitches in the synthesized output, which detract from the naturalness of the synthesized speech. Three main subtypes of concatenative synthesis are unit selection, diphone synthesis, and domain-specific synthesis.

Unit selection uses large speech databases in which each recorded utterance is segmented into parts, including individual phonemes, syllables, morphemes, words, phrases, and sentences. The division into segments can be done with a number of techniques, including clustering (including the words into similar classes), using a specially modified speech recognizer or manually creating with visual representations of the waveform. An index of the units in the database is created on the basis of segmentation and on acoustic parameters such as fundamental frequency. At run time (when speech is synthesized), we create the desired target utterance by determining the best chain of candidate units from the database (also known as unit selection). This technique is thought of as giving the greatest naturalness to the speech because no signal processing techniques are used on the recorded speech, which is thought to make the speech sound less natural. The advantage of unit selection is that the best unit selection systems may produce speech that is indistinguishable from real human voices, especially in contexts for which the system has been designed. A disadvantage of unit selection is that the speech databases are very large, containing dozens of hours of recorded speech and gigabytes of recorded data.

Diphone synthesis uses a smaller speech database containing all the diphones (sound-to-sound translations) occurring in a given language. Diphones consist of two phonemes (minimal distinctive phonetic units), incorporate transitional sounds, and are thought to produce better sounding speech. There are approximately 1,500 to 2,000 diphones in American English; Spanish has about 800 diphones, while German has about 2,500. In diphone synthesis, only one example of each diphone is contained in the speech database. At run time, the target prosody (the distinctive variation of stress or tone in phrases) is superimposed on these minimal units by means of digital signal processing techniques such as linear predictive coding (LPC), pitch synchronous overlap add method (PSOLA), or multi-band resynthesis overlap add (MBROLA), which takes a list of phonemes as input together with information about the phoneme duration and pitch and produces 16-bit speech samples at the sampling frequency of the diphone database. The quality of the resulting diphone synthesis is generally not as good as that produced by unit selection but is usually more naturally sounding than the output of domain-specific synthesizers. Diphone synthesis has the advantage of requiring a small database but has the disadvantages of sonic glitches of concatenative synthesis. In general, the use of diphone synthesis is declining in commercial applications but continues to be used in research because there are a number of freely available implementations.

Domain-specific synthesis strings together pre-recorded words and phrases to create complete utterances. This technology is very simple to implement and has been in commercial use for a long time. This type of synthesis is used in applications where the variety of output is limited, such as transit schedule announcements or weather report applications. Other applications include talking clocks and calculators. The advantage of domain-specific synthesis is that speech sounds more natural because the variety of sentence types is limited and closely matches the prosody and intonation of the original recordings. The disadvantage is that output is limited by the words and phrases in its database (the database is not general purpose). The domain-specific system is limited to producing only the combinations of words and phrases with which they have been pre-programmed.

3.5.3.2 Formant Synthesis

Formant synthesis does not use human speech samples in producing speech but creates output with an acoustic model. In this model, parameters such as fundamental frequency, voicing, and sound level are varied over time to create a waveform of artificial speech. This method is sometimes known as rule-based synthesis, but some argue that because many concatenative systems use rule-based components for the front end, that the term is not specific enough. Many formant-based systems generate artificial, robotic-sounding speech, and the output would never be mistaken for human speech. However, maximum naturalness is not always the goal of a speech synthesis system. The advantage of formant synthesized speech is that it can be very reliably intelligible, even at very high speeds, without the acoustic glitches that often plague concatenative systems. High speed synthesized speech is often used by the visually impaired for quickly navigating computers using a screen reader. A second advantage of formant synthesis is that it uses smaller programs than concatenative synthesis because a database of speech samples is not involved. Thus, formant synthesis can be used in embedded computing situations where memory space and processor power are often scarce. A third advantage is that formant synthesis provides total control over all aspects of the speech output. Thus, a formant system can output a wide variety of prosody or intonation, conveying not just questions and statements but a variety of emotions and tones of voice.

3.5.4 The Availability of Commercial Speech Synthesis Systems

Appendix B lists many of the currently available speech processing systems. Speech recognition systems are included in this table because synthesis and recognition are often included in the same systems (speech synthesis can provide feedback for speech recognition input). The systems described in the table include commercial off-the-shelf (COTS) hardware and software as well as software suites currently undergoing collaborative development by academia. Some of these products accomplish a specific task, such as ShortTalk[8], which was developed for converting spoken dictation into a written text file. Other systems such as Galaxy[9], Festival[10], SPRUCE[11], and SPHINX[12] are integrated suites of applications, which include speech recognition, speech synthesis, TTS conversion, and dialogue. A dialogue system integrates speech recognition, information retrieval, and TTS conversion into one system. Examples of a dialogue system include telephone information systems that deliver automated airline reservations or stock quotations at user prompts.

Galaxy by MIT has five main functions: speech recognition, language understanding, information retrieval, language generation, and speech synthesis. Pegasus (a Galaxy-based system) provides commercial flight information, while Voyager (another Galaxy-based system) is a guide to navigating the city of Boston.

SPRUCE is predominantly a TTS system with a unique high-level synthesizer configured to drive low-level synthesizers made by others, including Holmes, Klatt, PSOLA, and IBM[13]. It was essentially a research project, and though very promising, has not yet transitioned to a commercially available system.

SPHINX by CMU is a DARPA-funded project founded to create tools for speech applications and to advance the state-of-the-art directly in speech recognition and related areas of dialog systems and speech synthesis. This project resulted in several products, including SPHINX-2,

[8] An unusual method of composing and editing text by speech using shorthand command structures, developed by AT&T several years ago. Not undergoing development anymore.

[9] Galaxy is a conversational platform developed by the Speech Language Systems at the Massachusetts Institute of Technology (MIT). It has several modules inside for specific applications, Voyager for navigating around Cambridge, Massachusetts, Pegasus for airline scheduling, etc.

[10] Festival offers a general framework for building speech synthesis systems as well as including examples of various modules. Developed by the Center for Speech Technology Research, University of Edinburgh, U.K.

[11] A high-level text-to-speech synthesis system developed by a research team at University of Essex, U.K. The expansion of the acronym is unknown.

[12] A collection of real-time speech recognition engines, developed at CMU. Expansion of the acronym is unknown.

[13] IBM is a registered trademark of International Business Machines Corporation.

a real-time, large vocabulary, speaker-independent recognition system. SPHINX-2 includes acoustic models of American English and French in full bandwidth and reduced bandwidth telephone models. The reduced bandwidth version is well suited for hand-held, portable, and embedded devices that can tolerate lower quality of speech but also offers fast response times. SPHINX-3 is a slower, more accurate recognizer used for applications such as broadcast news transcription.

Festival is a multi-lingual speech system. It offers full TTS as well as an environment for the research and development of speech synthesis techniques. It includes a vocabulary of several languages and support for waveform synthesizers.

Products such as Interactive Speech[14], FluentSpeech[15], Aurix[16], and InterSound[17] are hardware and software technologies that are customized and integrated by other developers into consumer-ready end products such as toys, video games, and computer-based training. Aurix is a public and private sector collaborative effort from the U.K. Speech recognition and synthesis in this product are based on techniques developed during 30 years of research by the U.K.'s DSTL. As such, this technology has been developed to meet British military standards.

Mixed excitation linear prediction (MELP) is a technology developed by a partnership of Georgia Tech and Texas Instruments but has since been transferred to several other private sector corporations. Vocal Technologies, Ltd. is one of several companies that embed MELP-based synthesis into their products.

MBROLA by the Circuit Theory and Signal Processing (TCTS) Laboratory of the Faculté Polytechnique de Mons (Belgium) is a set of speech synthesizers for several languages, developed in an effort to boost academic research in speech synthesis and prosody generation. This synthesizer is based on diphone concatenation described previously in this report. Because it does not accept raw text as input, MBROLA is not a TTS synthesizer but is used as a low-level synthesizer driven by a higher level synthesizer such as SPRUCE.

WHISPER (Windows Highly Intelligent SPEech Recognizer) and WHISTLER (Windows Highly Intelligent STochastic taLkER) are a pair of Microsoft Windows[18]-based products integrated into Microsoft's Office Suite of applications including the Encarta encyclopedia. WHISPER and WHISTLER also come in a software development toolkit (SDK) version for other developers to create Windows-based applications.

[14]Interactive Speech is a trademark of Logic-Plus.
[15]FluentSpeech is a trademark of Sensory, Inc.
[16]Aurix is a registered trademark of 20/20 Speech Ltd.
[17]InterSound is a registered trademark of Intel.
[18]Windows is a registered trademark of Microsoft.

Nuance's[19] founders were originally employed at SRI at Menlo Park, California. Nuance's suite of applications is used by many large corporations and includes VoicePrint and Verifier, which are speaker verification and authentication products.

3.5.5 The Utility of Speech Synthesis in the HRI

The utility of speech synthesis in the HRI will depend on system requirements and applications as well as on potential problem areas (also known as challenges) that are specific to the HRI multi-platform environment. These factors are described next, along with a description of an application of speech synthesis in a future robotic system and recommendations for potential good candidates for HRI speech synthesis systems.

3.5.5.1 Speech Synthesis Requirements and Applications

The utility of speech synthesis in the HRI will depend on system requirements and applications. The ability to function with several different operating systems would provide flexibility of use if the HRI uses different robots or robotic controllers that run under different operating systems. Speech synthesis bundled with an ASR system would allow provision of speech feedback to the Soldier when used for voice input for robotic C2. A system robust in high noise environments would allow the user to understand robot feedback in high noise levels found in battlefields or in ground vehicles traveling over rough terrain. Speech synthesis might also be useful in the production of robotic system warnings, messages, diagnostics, and alerts. Because military vocabulary is limited, the synthesizer speech database does not need to be extensive. An embedded speech synthesis system would allow the preservation of robot or control unit system space and power. Finally, system compliance with military standards would ensure that the robot and control systems would have a better chance of success in military applications and environments.

3.5.5.2 Challenges to Synthetic Speech in the HRI

Several factors provide challenges to the use of synthetic speech in the HRI. These include limitations of front end processing of text input (if applicable) into the robot or control unit synthesizer and intelligibility of HRI synthesized speech in high levels of ambient noise found in battlefields or in ground vehicles during travel.

Limitations of front end processing in possible HRI TTS applications include ambiguity because of words that are pronounced differently in context and the use of text numbers. In an example that might be used in HRI system maintenance and diagnostics, ambiguity of maintenance text input can result from the use of words that are pronounced differently in context, such as with the word "project." In the sample sentence, "The purpose of these robotic maintenance projects is to ensure that Part A projects over Part B," the word "projects" has identical spellings but two different pronunciations. Other front end challenges include the conversion of numbers. It is

[19]Nuance, which is based in Burlington, Massachusetts, is a provider of speech and imaging solutions for businesses and consumers. VoicePrint and Verifier are Nuance's software solutions for business applications, in the areas of speaker recognition and authentication.

fairly simple to convert a number into words, with the number “1325” becoming “one thousand three hundred twenty-five.” However, numbers occur in many different contexts in text, and the number “1325” could possibly be read as “thirteen twenty-five” when part of an address, and as “one three two five” if used as the last four digits of an HRI map or robotics coordinate system. In general, TTS systems with intelligent front ends can make educated guesses about how to handle ambiguous words or numbers by examining neighboring words and using statistics about frequency of occurrence.

The intelligibility of synthesized speech is of great concern, especially when it is used in noisy environments such as in moving ground vehicles or battlefield conditions. Morrison and Casali (1994) explored synthesized speech warnings used by drivers with normal hearing and impaired hearing in noisy commercial truck cabs. These researchers found that auditory synthesized speech designed for use in heavy truck cabs must contend with the high noise levels already present in these vehicles and that truck cab noise levels could have a degrading effect on the intelligibility of synthesized voice messages and could pose a substantial risk to drivers. The maximum level of ambient noise encountered in this study was 80 decibels (dBA) SPL, which is representative of the level in truck cabs traveling on smooth freeway pavement. These noise levels are not as loud as those in ground vehicles or tanks, which may reach levels of 113 dBA. Morrison and Casali recommended that the articulation index be used to predict relative synthetic speech intelligibility in noisy environments, although they suggested that higher background noise levels might produce less definitive results. Research is needed to explore the effect of noisy ground vehicle and battlefield environments on the intelligibility of speech synthesis systems in the HRI.

3.5.5.3 Future Uses of Speech Synthesis in Robotic Systems

Researchers at MIT (Fitzpatrick, Metta, Natale, Rao & Sandini, 2003) are building robots with a human-like form, theorizing that this will allow more human-like interactions with people. This robot, known as Cog, incorporates an artificial intelligence (or artificial cognition) device to enable it to learn through social interactions with people. Cog has a human-like face, learns how its own movements alter its sensory input and takes energy efficiency into account during movements. Speech synthesis is used to enable Cog to communicate with people to learn to function in its environment.

3.5.5.4 Recommendations for Speech Synthesis in the HRI

Based on the requirements and challenges described, it is suggested that the InterSound (Intel Corporation) might work well in HRI speech synthesis applications because it provides synthesized speech in embedded systems, works with many operating systems, and has been used in military system applications. The MELP Vocoder (Vocal Technologies, Ltd.) is robust in high noise environments and was selected by DoD digital voice processing consortium for the new 1200- and 2400-baud Federal Standard speech coder. The British-built Aurix system (20/20 Speech), which is primarily an ASR that includes speech synthesis feedback was designed and tested to U.K. military standards. The Aurix system provides hands-free operation of complex system C2. Finally, it is recommended that readers of this report be aware that the information in this report is dated because the speech technology field is constantly evolving because of rapid advances in software and corporate transfer of technology. Corporations and technologies that exist at the time of this report may not exist at a future date.

3.6 Haptic Display Interfaces

3.6.1 Background

Haptic interfaces and displays interact with the skin to present information. One effective example used as an illustration of a haptic display is the "vibrate" function found on most pagers and cell phones (Gemperle, Ota, & Siewiorek, 2001). In this example, one tactile signal is presented through a small, dime-sized vibrator motor to announce the event of an incoming telephone call to the user of a cell phone. Several researchers, including Gemperle et al., found that multiple addressable tactors could be spread across an area of the body to convey complex and coordinated information such as navigation or guidance cues. Gemperle et al. suggested that tactile displays can be used to direct user attention to critical events, especially when user visual or auditory channels are busy or blocked.

An understanding of haptic displays must first begin with a description of the anatomy of the skin, which is the primary organ for this type of display (figure 10). The skin has many different kinds of receptors for receiving sensations, including those of touch, pressure, texture, temperature, pain, and movement of the skin hairs. The human skin provides an extensive haptic space; the skin surface of an average-sized adult human spans 19 square feet (Gemperle et al., 2001). The skin has two layers, the epidermis and the dermis. The epidermis, a thin outer layer ranging from 1/200th to 1/20th of an inch, is composed of dead cells and directly interacts from the environment. Beneath the epidermis is the dermis, a layer of dense connective tissue that averages 1/15th to 1/8th of an inch in thickness. The human body contains several types of skin: glabrous (palms and soles), mutocutaneous (lips), mucus, and hairy. The hairy skin, which covers most of the body, including arms, thorax, and back, is used for tactile display receptors.

Novices in the area of haptic displays often confuse the use of the words "haptic" and "tactile." Webster's Online Dictionary (2004) defines tactile and haptic as "of or relating to or proceeding from the sense of touch". Many researchers define "haptic" to include skin-based as well as proprioceptive (body position, orientation, and movement) information and use the word "tactile" to refer to a type of haptic display that uses pressure or vibration stimulators that interact with the skin (Gemperle et al., 2001). This particular usage is employed in this report.

There are two common techniques used to generate vibration in tactile displays (Van Erp, 2002). The first technique is based on a moving coil driven by a sine wave, while the second is based on a direct current motor with an eccentric weight mounted on it (as found in mobile phones). Other less common actuators are based on piezoelectric benders, air puffs, and electrodes.

Although the actuators or tactors differ in their characteristics as a display element, the basic psychophysics (how people perceive the vibrations) are independent of type of actuator.

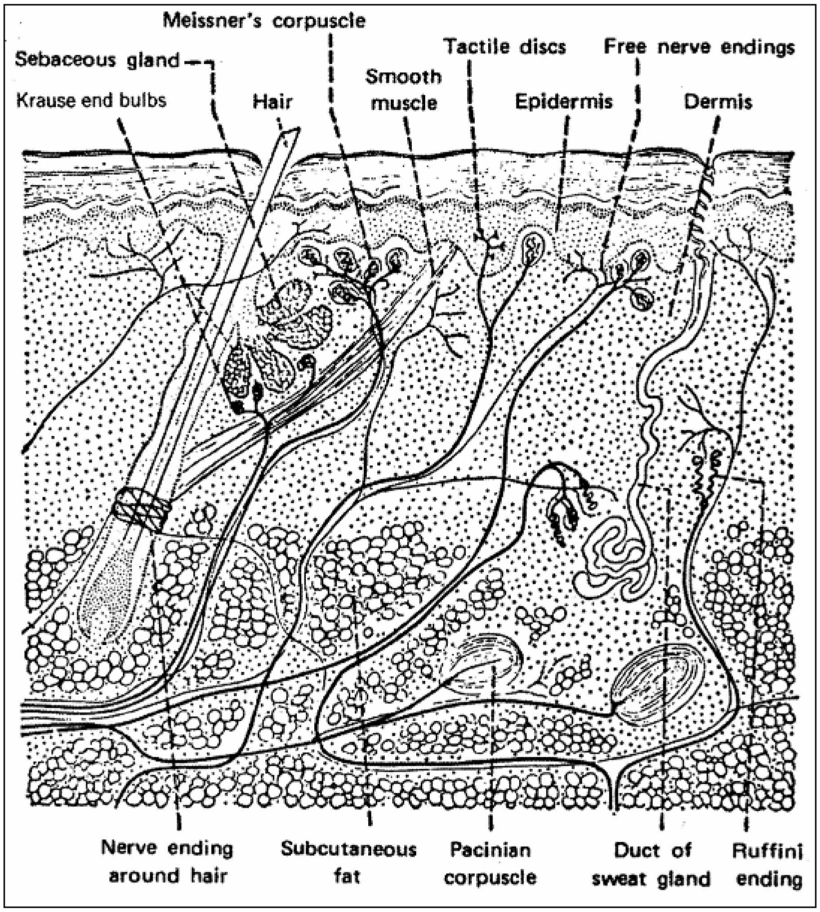

Figure 10. Cross section of the human skin (adopted from Schiffman, 2001).

3.6.2 A Comparison of Tactile Display Technologies

Haptic displays of greatest potential interest to the U.S. Army include tactile displays that provide information through the use of some type of sensor emplaced on the skin. Because skin-emplaced sensors can signal events, tactile displays can be used to provide information when visual or audio cues may not be available. Information provided by tactile displays includes warnings and alarms, navigation and guidance cues, system location and malfunction information, and threat warnings. Gilson, Merlow, Brill, Stafford, and Mathews (2005) demonstrated that tactile cueing lowered participants' response times by more than 1 second in target acquisition tasks compared to visual cueing. As to accuracy, Terrence, Brill, and Gilson (2005) reported that tactile target cueing caused significantly fewer localization errors than did auditory cueing. Terrence et al. also noted that regardless of body orientations of the participants (i.e., supine, kneeling, sitting, standing, and prone), there were significantly fewer localization errors in the tactile condition for five of the eight cardinal directions (the errors for the other three directions were also fewer, but the differences were not statistically significant). The mean reduction in angle differences between presented and perceived cues was 27.11 degrees.

Appendix C lists many of the currently available tactile display systems. The systems described in the table include COTS systems or kits as well as systems currently undergoing collaborative development by academia. The displays of greatest interest to the Robotics Collaboration ATO are described next; they include the TNO[20] Tactile Torso Display, the U.S. Navy Tactile Situational Awareness System (TSAS), the CMU Wearable Tactile Display, the MIT wireless tactile control unit (WTCU), and the University of Central Florida (UCF) Tactile Communication System (TACTICS). These displays, most of which are designed as wearable vests containing multiple tactors, are described next.

3.6.2.1 TNO Tactile Torso Display

Researchers at the TNO Human Factors Research Institute in Soesterberg, the Netherlands, designed and used tactile torso displays in many different applications (Van Erp, Veltman, van Veen, & Oving, 2003). The application of greatest interest to the Robotics Collaboration ATO is the tactile torso display developed by TNO to supplement visual displays used in helicopter hover tasks (figure 11).

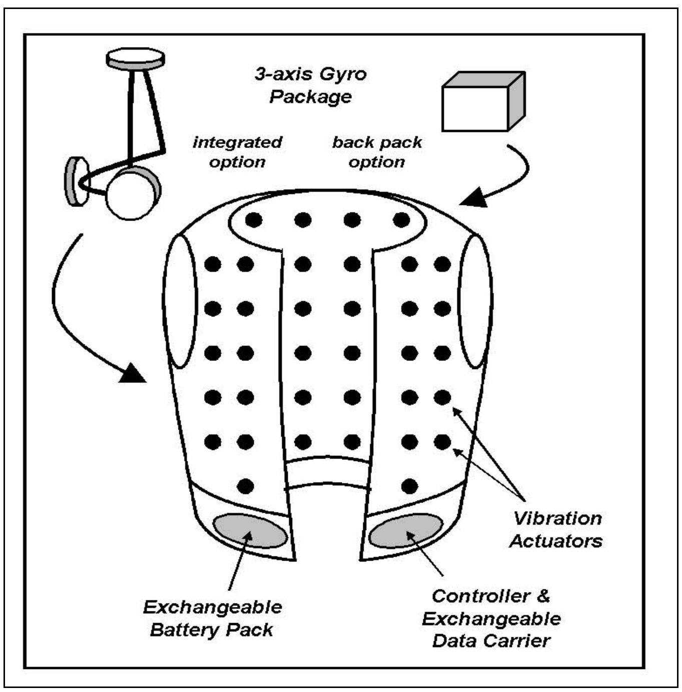

Figure 11. Tactile torso display developed by TNO (Van Erp, Veltman et al., 2003).

Helicopter pilots, who often wear night vision goggles (NVGs) when performing nighttime hover tasks, may not notice aircraft drift. The TNO display was developed to furnish positional cues to

[20]Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek.

enable pilots to correct drift. The display consisted of 64 vibro-tactile elements to present two types of information: a “simple” display to present information of desired direction of aircraft motion only, and a “complex” display to include not only desired direction of motion but additional information regarding current direction. A study was run in which helicopter pilots flew simulated hover tasks in a fixed base helicopter simulator with full vision or with simulated NVGs (Van Erp, Veltman et al., 2003). The results indicated that pilot performance improved when visual cues were supplemented with tactile cues. Results showed a mean reduction in positional error in horizontal and vertical direction when both tactile variations were used, with and without NVGs. Van Erp, Veltman et al. noted that the complex variant of the tactile torso display was less effective than the simple variant, perhaps because of “tactile clutter,” where user confusion arises because too many tactors are used or initiated at one time. This study proved the potential utility of tactile torso displays in reducing drift during hover and that this type of display could even be applied in demanding human-in-the-loop tasks in which complex information is delivered fairly quickly. TNO also developed a vibro-tactile display useful for automobile navigation applications. Van Erp, Meppelink, and van Veen (2002) developed a display in which small vibrators were embedded in a car seat to provide directional and navigation information to drivers. The actuators vibrated on certain sides to alert drivers when a turn was suggested and vibrated faster the closer the car came to a turn. This display was tested in a driving simulator in which participants drove different routes through a simulated city. Vehicle navigation information was presented via a visual display, a tactile display, or both. The results of this study indicated that the addition of a tactile navigation display resulted in better performance and lower driver workload compared to the visual display. Van Erp, Meppelink, et al. noted that tactile automotive display released other heavily loaded sensory channels and may lead to major improvements in driver safety. This system is being adapted to motor vehicles in 2 years, and the haptic seat will make its debut in high-end automobiles (Glaskin, 2004).

Van Erp, Meppelink, et al. (2002) also explored the use of tactile vests in other applications. Their work includes a haptic vest developed for jet pilots who become spatially disoriented during high-speed maneuvers. Tactile vests were also used during experiments on the International Space Station to help scientists understand astronaut motion sickness.

3.6.2.2 TSAS

The TSAS tactile vest was developed by CPT Angus Rupert (Chiasson, McGrath, & Rupert, 2003) for the U.S. Navy (see figure 12). The object of this vibro-tactile display was to inform pilots of their spatial orientation in 3-D space. The TSAS consisted of four vertical columns with five tactors for each column sewn into a vest. The tactors were operated in two modes: a “high” mode using all the sensors to transmit directional information in a sequential pattern of continuous motion across the body, and a “low level” mode in which only three tactors per column were activated to signal warning and alarm conditions.

TSAS testing was conducted by U.S. Navy pilots, who used the vest during hover and flying operations. In addition to aviation, the TSAS was also tested in under-water Navy SEa, Air, Land (SEAL) applications and by Army pilots in an aircraft simulator as part of the Virtual Cockpit Optimization Program. TSAS testing determined that the minimal distance between tactors for differentiating the presence of both tactors was between 2.0 and 2.5 inches. In addition, testing revealed a psychological limit for information density; pilots could not distinguish between signals containing more than two parameters of information (each parameter consisting of altitude, target location, or threat location information). Thus, TSAS researchers limit tactile vests to two layers of information, such as direction and speed, or speed and rotation. The TSAS was considered successful and generated a surge of interest; the TSAS has been integrated into the Touch Lab at MIT and into the Cutaneous Communications Lab at Princeton University.

Figure 12. TSAS vest (Chiasson, McGrath, & Rupert, 2003).

Recently, Chiasson, McGrath, and Rupert (2003) described the use of the TSAS for the Navy Special Forces operations. In this study, a TSAS vest was upgraded to present tactile directional navigation information in high altitude, high opening parachute operations, in ground environments, and in under-water operations. The authors claimed that displays with tactile and visual cues resulted in better human performance than those using visual cues alone and that superior navigational accuracy can be achieved with less mental fatigue on the operator. Chiasson et al. suggested that a tactile display that provides “eyes free” and “hands free” air and ground information may free the user to devote more time to other instruments and tasks when operating in high workload conditions, thus increasing mission effectiveness.

3.6.2.3 CMU Wearable Tactile Display

The Wearable Group at CMU has been designing and testing wearable computers for industrial and military applications for more than 10 years. Part of this effort involved the design and testing of a wearable tactile display, the purpose of which was to interact and interface with wearable and mobile computers. Gemperle et al. (2001) found that their designs are driven by the users’ need to have their hands and eyes free. Their goal is to use multiple addressable tactile stimulators spread across an area of the body to convey complex and coordinated information.

The initial CMU wearable design was a flexible band of tactors that could be worn on the shoulder to provide navigational information. The display was lightweight and the tactors were small. However, when activated, the tactors were loud (75 dBA) and the vest was bulky and consumed a great amount of power. For that reason, the first tactile display was scrapped, although the overall harness styling had some advantages. Gemperle et al. (2001) noted that the harness was comfortable and was easy to don and doff.

A later research project, the CMU wearable tactile display, actually resembled a vest in that it had a familiar waistcoat or vest shape in a design that fit over the torso (see figure 13). The vest was made of heavyweight Lycra[21] and had pockets sewn into the inside to allow the emplacement of tactors at various locations around the torso. Gemperle et al. (2001) noted that this design had several advantages: the vest used tactors that were smaller, silent, and used less power. In addition, a wireless infrared kit allowed the creation of a remote controller to activate the tactors. The range of the remote controller was several feet, which allowed Gemperle et al. to test the device in the lab and around campus with minimal bulk or weight for the subject and no tether or wires to other computers. The wearable tactile display was used to test the presentation of navigational information through the skin and to evaluate body position and signal modulation parameters.

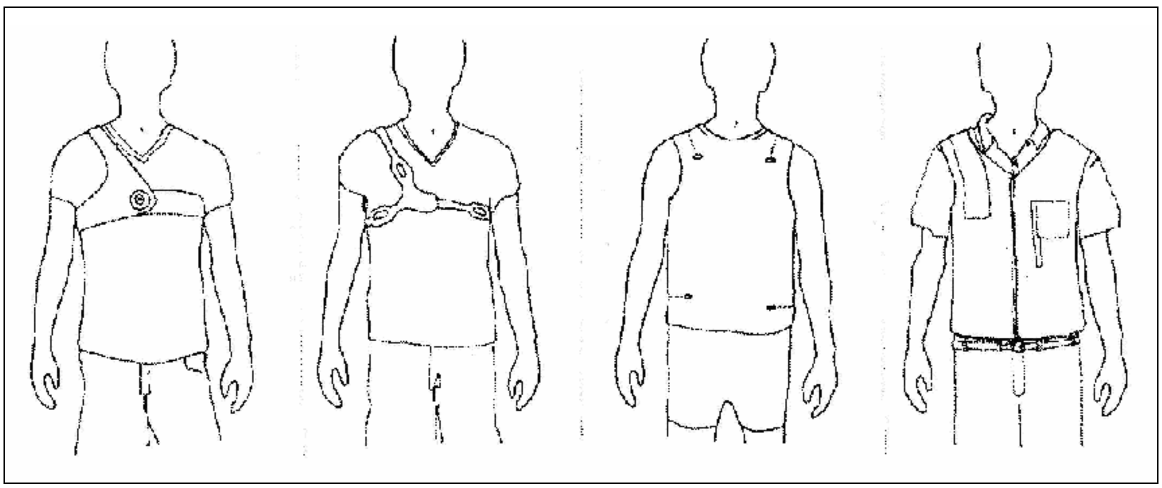

Figure 13. Four designs for a wearable tactile display on the upper torso: from left, an outerwear accessory hiding the function, one showing the function, underwear, and tactile display embedded in a tool vest (Gemperle et al., 2001).

Future CMU research will involve a tactile display with a universal serial bus device that can be plugged into “Spot,” a wearable computer developed by the Wearable Group at CMU. Spot is intended to allow researchers to program complex vibration sequences and then connect these sequences to spatial information. The spatial information can be transmitted from an internal global positioning system (GPS) from an on-campus local area network, allowing the user to test the usability of the tactile display at different locations on campus.

[21]Lycra is a registered trademark of E. I. DuPont de Nemours & Co, Inc.

3.6.2.4 MIT WTCU

Dr. Lynette Jones of MIT, Department of Mechanical Engineering, developed a haptic display as part of the Advanced Decision Architecture, Collaborative Technology Alliance with ARL. Dr. Jones' work (Jones & Lockyer, 2004) was conducted in the context of the design of body-based haptic interfaces that could be used by human operators to interact with computer-generated VE or to control robotic devices. Dr. Jones focused on developing tactile displays that can be worn on the torso or arms and used as navigation aids. In doing so, she explored a range of actuator technologies, including quasi-static mechanical deformation actuators and high-bandwidth vibro-tactile actuators. She is currently building and testing a number of tactile vests that use small electromagnetic motors or shape memory alloy fibers as their actuators and is exploring the development of a thermal haptic display. Several of her recent studies included an analysis of several different tactor types, including vibration motors, roto-tactors, and pancake motors.

Dr. Jones developed the MIT WTCU, which includes an embedded processor that receives commands from a host computer and translates these into patterns of tactor motor actuation (see figure 14). Each tactor actuation pattern can be distinct and is associated with only one command, but the system can be configured to accommodate any number of commands. The system incorporates a micro-controller as well as a motor controller that can be programmed to interface with any wireless transmitter/receiver, as long as it communicates via a universal asynchronous receiver/transmitter protocol. The wireless technology used in this system includes Bluetooth[22] wireless technology, although Wi-Fi[23] and ZigBee[24] transmissions are also possible. The current WTCU uses a MaxStream[25] Bluetooth transmitter that is relatively small (roughly 85 X 40 x 16 mm) and can transmit to distances of 75 m.

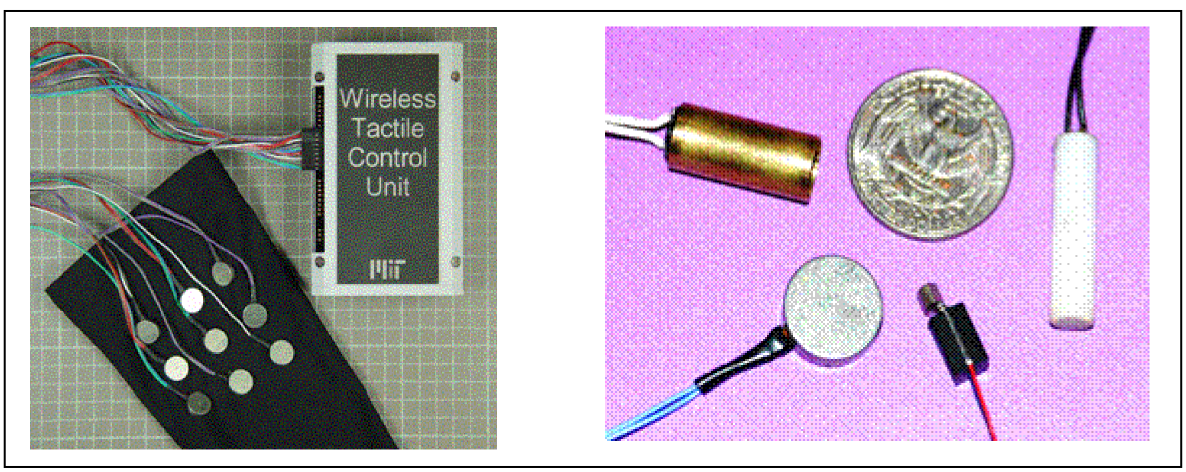

Figure 14. MIT WTCU.

---

[22]Bluetooth is a registered trademark of Bluetooth Special Interest Group.
[23]Wi-Fi is a trademark of Wi-Fi Alliance.
[24]ZigBee is a trademark of ZigBee Alliance.
[25]MaxStream is a registered trademark of MaxStream, Inc.

Dr. Jones' future research will include adding physiological monitoring systems to the WTCU device and wearable display and testing new wireless communication protocols. She will also conduct experiments exploring the discrimination of different vibro-tactile patterns and the use of various tactile cues as navigational aids. The WTCU is currently being used by ARL's Human Research and Engineering Directorate in several experiments, including those conducted for the situational understanding ATO. Future research is planned for the Robotics Collaboration ATO.

3.6.2.5 UCF TACTICS

TACTICS was developed by Dr. Richard Gilson of UCF's Department of Psychology for DARPA. TACTICS contains eight rugged tactors fitted around mid-waist, with optional two-dimensional capabilities (see figure 15).

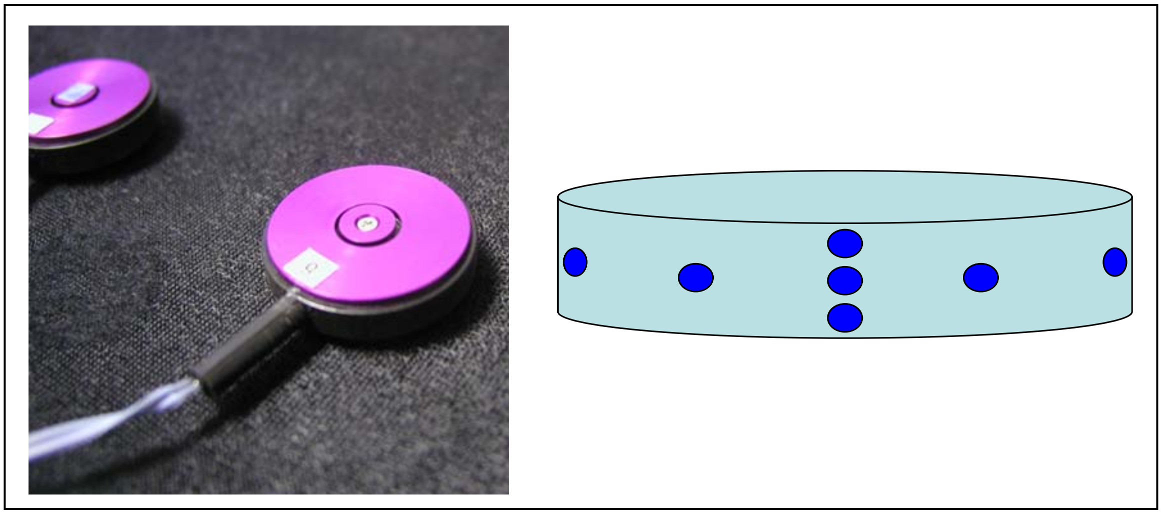

Figure 15. UCF TACTICS.

The TACTICS control unit receives signals from wireless PDA and converts them into recognizable vibration patterns. Currently, six basic tactile patterns (attention, halt, and move out, direction to move, rally, and nuclear-biological-chemical) designed to be analogous to standard Army hand signals are included. According to Gilson et al. (2005), TACTICS has been shown to be effective in providing the following: (a) rapid directional cueing, (b) covert messaging, (c) low interference with visual and auditory tasks, and (d) superiority in degraded conditions. Dr. Gilson suggested that the uses of TACTICS could be potentially extended to HRI in the following areas: (a) mission priority alerts (destination alerts or cloud/ice warning for UAVs), (b) direction of enemy targets acquired by UV (relative to Soldier or UV), and (c) direction of UV. The UCF team is currently conducting field testing at Fort Benning in collaboration with Dr. Elizabeth Redden of ARL.

3.6.3 Guidelines for Tactile Display Design

Van Erp (2002) developed guidelines for designing and using tactile displays. These guidelines, some of which are summarized in this section, discuss the use of tactor amplitude (intensity), frequency, timing and location, and their effect on stimulus detection, information coding, and user comfort.

For a tactile display to be useful, the stimuli must first be detected by the user. Vibration stimuli will be detected when the actuator amplitude exceeds a certain threshold. This detection threshold depends on several parameters, including skin frequency (the skin is roughly sensitive to vibrations between 20 and 500 Hz) and the location on the body. Location is an important design consideration, in that the lowest vibration thresholds (the most sensitive locations) are found on glabrous skin as compared to hairy skin. Vibration frequencies are best in the range of 200 to 250 Hz. Vibration stimuli are more detectable when stimulus duration increases (also known as temporal summation), but this only works for frequencies above 60 Hz. Detection of tactile stimuli is best when there is a fixed ring of rigid material surrounding the vibrating element. Van Erp (2002) suggested that tactor waveform affects detection; a square wave is best because it is most intense; however, a sine wave is smoothest. He suggested that because there is a high variation in the thresholds of sensation and pain, between people and within individuals (age affects perception), the user must be able to adjust the intensity of the tactile stimulus.

Detection of tactile cues is not enough to provide sufficient information to the user. Proper coding of tactile cues allow the user to discern that there is more than one cue being communicated and to understand the nature of the message sent by the tactile display. Tactor information can be coded by magnitude or intensity of vibration. Van Erp (2002) suggested that not more than four different levels be used between the users' detection threshold and their threshold of pain or comfort. Amplitude coding is possible if the intensity of individual actuators is enlarged or if the area of stimulation is enlarged, which can be accomplished by the actuation of two or more tactors at once.

Tactor information can also be communicated by actuator frequency. Van Erp (2002) suggested that not more than nine different levels of frequency should be used for coding information and that differences between frequency levels should be at least 20%.

Temporal pattern is a very useful coding format for tactile signals. Van Erp (2002) suggested that temporal sensitivity of the skin is very high and is close to that of the auditory system and greater than that of the visual system. When one is using a single actuator to communicate temporal information, most often in on/off patterns, Van Erp (2002) suggested that the time between signals must be at least 10 ms. Thus, signals should be 10 ms on time, followed by 10 ms off time, to be detected by the user. However, the display designer should be aware that depending on the type of actuator and load, a vibratory stimulus will take time to reach the set frequency and may smother slowly, so "on" and "off" time may not be easily delineated by the user.

Van Erp (2002) suggested several guidelines for multi-tactor displays. In a multi-element display, actuator location and density are important parameters. When a high tactor density is needed, only certain body parts such as the fingers, hands, and face have sufficiently high spatial resolution to accommodate them. When spatial acuity as low as 4 cm is acceptable, any locus will suffice. When complex tactile messages are used (more than one tactor supplying information), Van Erp (2002) recommended that they be composed of meaningful components. He noted that combining different signals may alter, negatively interfere with, or confuse the user's perception of the signal, perhaps through spatial-temporal interactions. Another potential problem with multiple tactor units is tactile clutter, where the simultaneous or sequential presentation of multiple tactile messages on the same display can result in the user experiencing a reduced comprehension of the display because of sensory overload.

In general, when one is coding tactile displays, Van Erp (2002) noted that it is important to make tactile messages very intuitive (self-explanatory) to the listener. Intuitive displays are important when users will not experience tactile signals continuously and must be able to remember tactile signals during the time period between actuation.

User comfort is an important issue. Van Erp (2002) noted that tactile information presentation requires actual contact between the tactor and the skin of the user, so it is important to ensure user comfort over long periods of time. He recommended that tactile displays worn on the body must be comfortable for the longest intended period of usage. As with signals in other modalities, tactile stimuli may be difficult to ignore if the user does not want to use them, so it is recommended that tactor signals not annoy the user.

Finally, Van Erp (2002) described pitfalls for applying tactile stimulation. He noted that the skin often integrates multiple stimuli, which may result in a tactile percept (perception) that differs completely from the sum of the original stimuli. One example given here was spatial masking, where the location of a tactile stimulus is masked by another stimulus. This may occur when stimuli overlap in time but not in location. As a result, both stimulus detection and identification may be degraded. To avoid this, he recommended using stimuli with different frequencies (one below 80 Hz and one above 100 Hz). Van Erp (2002) also warned about the pitfall of apparent location, in which the perception of a single stimulus is induced by the simultaneous activation of two stimuli at different locations. When this occurs, rather than perceive two stimuli, a third nonexistent or phantom location is perceived by the user to be between the two stimulus loci. The exact position depends on the relative magnitude of the stimuli. To avoid this, both stimuli should be in phase, to evoke a stable perception of multiple stimuli. However, Van Erp (2002) noted that apparent location would be useful in increasing the number of subjective stimulus sites, without our having to increase the actual number of actuators used.

3.6.4 Haptic Display Conclusions and Recommendations

Based on the data given in this report and in the database (see appendix C), the TSAS and the TNO vests stand out as good candidates for the HRI. These systems might work well in HRI OCU applications because they are both relatively mature systems that have been undergoing constant research and development for more than 10 years and because they are robust, having been tested in helicopter cockpits and other military applications. Although the MIT WTCU system is currently used in several studies at ARL, this system is not sufficiently robust for use in a field or crew station environment; this system is recommended for laboratory use only. As can be seen in the comparison database, there are no commercially available systems recommended for use; most tactile display development is currently being conducted by universities or military facilities. Finally, it is recommended that the reader be aware that the information in this report is dated because the tactile technology field is constantly evolving because of rapid advances in software and corporate transfer of technology. Corporations and technologies that exist at the time of this report may not exist at a future date.

Also of interest is the Cybernet Systems Corporation wearable OCU for human-portable robotic applications. Typically, the OCU will be used to guide small, human-portable robots for tactical missions, such as reconnaissance in enclosed spaces such as sewer tunnels. In military parlance, it is a “first man in” situation, where the robot relays video information back to the operator. The OCU also controls movement of the robot. Since this is a tactical situation, the OCU must allow the Soldier to be free to perform other duties without undue hindrance. It is especially important that the Soldier remain free to perform battle tasks while operating the control system. The OCU displays video information and other status information (direction of travel, velocity, tilt angle, etc). It also accepts control commands from the Soldier and transmits the control data to the robot, commanding the robot to move left, right, or forward.

A type of haptic display of interest to Robotics Collaboration ATO in future applications would be a force feedback display, which involves devices that interact with muscles and tendons, which give the human a sensation of a force being applied. Current force feedback devices mainly consist of robotic manipulators that push back against a user, usually the user’s hand, with forces that correspond to the environment in which the effector is located. Displays of this type might be useful in teleoperation tasks because they can be used to signal critical events such as the appearance of terrain hazards or features, which can be useful in guiding or steering. Force feedback displays are also used to guide robot arms that perform other functions, such as cutting skin or other tissues in surgical applications.

4. Conclusions

This report examined human performance issues in robotics control environments and reviewed user interface solutions that could potentially address those issues. As robotics become increasingly prevalent in military and civilian operations, it is important to understand HRI and its associated limitations as well as potentials. In the foreseeable future, it will be more common for humans to work with robotics as a team to perform tasks that humans cannot realistically accomplish alone. Research programs such as the U.S. Army's Robotics Collaboration ATO also started to explore how to enhance operator performance by employing advanced technologies and user interface design concepts (Barnes, Cosenzo, Mitchell, & Chen, 2005). For example, multi-modal user interfaces such as 3-D audio and adaptive automation techniques can be very beneficial, especially in stressful and multi-tasking environments. These solutions and other innovative user interface designs reviewed in this report can hopefully make operators' robotic control tasks less challenging than they currently are.

5. References

Arkin, R. C.; Fujita, M.; Takagi, T.; Hasegawa, R. An ethological and emotional basis for human-robot interaction. *Robotics and Autonomous Systems* **2003**, *42* (3-4), 191-201. Retrieved March 26, 2004, from http://www.csee.usf.edu/~murphy/Courses/RAP/Review/Set3/arkin-emotion.pdf.

Aleotti, J.; Bottazzi, S.; Caselli, S.; Reggiani, M. A multimodal user interface for remote object exploration in teleoperation systems. Retrieved August 5, 2004, from http://rimlab.ce.unipr.it/Publications/HUROIN2002.pdf, 2002.

Associated Press. Mars rover finishes primary missions. Retrieved April 6, 2004, from http://www.cnn.com/2004/TECH/space/04/05/mars.rovers.ap/index.html, 2004.

Aviles, W. A.; Hughes, T. W.; Everett, H. R.; Umeda, A. Y.; Martin, S. W.; Koyamatsu, A. H.; Solorzano, M. R.; Laird, R. T.; McArthur, S. P. Issues in mobile robotics: The unmanned ground vehicle program TeleOperated vehicle (TOV). *SPIE Vol. 1388, Mobile Robots V* (pp. 587-597), 1990.

Bar-Cohen, Y.; Mavroidis, C.; Bouzit, M.; Dolgin, B.; Harm, D. L.; Kopchok, G. E.; White, R. Virtual reality robotic telesurgery simulations using MEMICA haptic system. *Proceedings of SPIE's 8th Annual International Symposium on Smart Structures and Materials*, 4329-47. Retrieved May 28, 2004, from http://ndeaa.jpl.nasa.gov/ndeaa-pub/SPIE-2001/Paper-SPIE-4329-47-MEMICA.pdf, 2001.

Barnes, D. P.; Counsell, M. S. Haptic communication for remote mobile manipulator robot operations. *Proceedings of 8th American Nuclear Society Topical Meeting on Robotics & Remote Systems.* Retrieved May 28, 2004, from http://users.aber.ac.uk/dpb/ans_99.pdf, 1999.

Barnes, M. J.; Cosenzo, K.; Mitchell, D.; Chen, J.Y.C. Soldier robot team as soldier augmentation in future battle spaces: An overview. *Proceedings of the 11th Annual Human-Computer Interaction International Conference*, Las Vegas, NV, 2005.

Blackburn, M. R.; Laird, R. T.; Everett, H. R. Unmanned ground vehicle (UGV) lessons learned (Tech. Rep. 1869). San Diego, CA: SPAWAR. Retrieved August 17, 2005, from http://www.spawar.navy.mil/sti/publications/pubs/tr/1869/tr1869.pdf, 2001.

Brungart, D. S. Speech-based distance cueing in virtual auditory displays. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 3714-3717), Santa Monica, CA, 2000.

Burke, J. L.; Murphy, R. R.; Coovert, M. D.; Riddle, D. L. Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. Retrieved April 27, 2004, from http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2004-5.pdf, 2004.

Burke, J. L.; Murphy, R. R.; Rogers, E.; Lumelsky, V. J.; Scholtz, J. Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Systems, Man and Cybernetics Part C: Applications and Reviews, special issue on Human-Robot Interaction* **2004**, *34* (2). Retrieved April 16, 2004, from http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2004-3.pdf.

Calhoun, G.; Draper, M.; Ruff, H.; Fontejon, J.; Guilfoos, B. Evaluation of tactile alerts for control station operation. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 2118-2122), Santa Monica, CA, 2003.

Cao, C.G.L.; Webster, J. L.; Perreault, J. O.; Schwaitzberg, S.; Rogers, G. Visually perceived force feedback in simulated robotic surgery. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 1466-1470), Santa Monica, CA, 2003.

Carlson, J.; Murphy, R. R. How UGVs physically fail in the field. Retrieved April 16, 2004, from http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2004-16.pdf, 2004.

Casper, J. Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. Unpublished master's thesis, University of South Florida, Tampa. Retrieved Apr. 16, 2004, from http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2002-8.pdf, 2002.

Casper, J.; Murphy, R. R. Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man and Cybernetics Part B* **2003**, *33* (3), 367-385. Retrieved April 16, 2004, from http://crasar csee.usf.edu/research/Publications/CRASAR-TR2003-2.pdf.

Casper, J.; Yanco, H. RoboCup/AAAI robot rescue competition. *AI Magazine* **2002**, *23* (1), 43-49.

Chen, J.Y.C.; Durlach, P. J.; Sloan, J. A.; Bowens, L. D. *Robotic operator performance in simulated reconnaissance missions;* ARL-TR-3628; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2005.

Chiasson, J.; McGrath, B. J.; Rupert, A. H. *Enhanced situational awareness in sea, air, and land environments;* (DTIC XB-NAMRL, pp. 1-10); Naval Aerospace Medical Research Laboratory: Pensacola, FL, 2003.

Cowings, P. S.; Toscano, W. B.; DeRoshia, C.; Tauson, R. *Effects of command and control vehicle (C2V) operational environment on soldier health and performance*; Tech. Rep.: NASA/TM-1999-208786; NASA Ames Research Center: Moffet Field, CA, 1999.

Darken, R. P.; Kempster, K.; Peterson, B. Effect of streaming video quality of service on spatial comprehension in a reconnaissance task. *Proceedings of the 23rd Army Science Conference* (pp. 391-400), Orlando, FL, 2001.

Darken, R. P.; Peterson, B. Spatial orientation, wayfinding, and representation. In K. Stanney (Ed.), *Handbook of virtual environment technology* (pp. 493-518). Mahwah, NJ: Erlbaum, 2002.

David, L. Robotic missions to save Hubble proposed. Retrieved April 6, 2004, from http://www.cnn.com/2004/TECH/space/04/05/robot.hubble/index.html, 2004.

Davis, G.; Jones, R. *Yamaha sound reinforcement handbook* (2nd ed.), p. 118. Milwaukee, WI: Hal Leonard Corporation, 1990.

Deng, Z.; Jagersand, M. Predictive display system for tele-manipulation using image-based modeling and rendering. Retrieved August 5, 2004, from http://www.cs.ualberta.ca/~jag/research/papers/iros03predSys.pdf, 2003.

Diolaiti, N.; Melchiorri, C. Tele-operation of a mobile robot through haptic feedback. *Proceedings of IEEE International Workshop on Haptic Virtual Environments and Their Applications* (pp. 67-72), 2002.

Dixon, S. R.; Wickens, C. D. Automation reliability in unmanned aerial vehicle flight control. *Proceedings of Human Performance, Situational awareness and Automation Conference* (pp. 205-209). Mahwah, NJ: Lawrence Erlbaum, 2004.

Dixon, S. R.; Wickens, C. D.; Chang, D. Comparing quantitative model predictions to experimental data in multiple-UAV flight control. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 104-108), Santa Monica, CA, 2003.

Draper, M.; Calhoun, G.; Ruff, H.; Williamson, D.; Barry, T. Manual vs. speech input for unmanned aerial vehicle control station operations. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 109-113), Santa Monica, CA, 2003.

Draper, J.V.; Handel, S.; Hood, C. C. Three experiments with stereoscopic television: When it works and why. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics* (pp. 1047-1052), 1991.

Drascic, D. An investigation of monoscopic and stereoscopic video for teleoperation. Master's thesis, University of Toronto. Retrieved August 9, 2004, from http://vered.rose.utoronto.ca/people/david_dir/MASC/MASC.full.html, 1991.

Drascic, D.; Grodski, J. Defense teleoperation and stereoscopic video. *Proceedings of SPIE, 1915 (Stereoscopic Displays and Applications IV)*, 58-69. Retrieved August 9, 2004, from http://vered.rose.utoronto.ca/people/david_dir/DND93/DND93.full.html, 1993.

Elias, B. Dynamic auditory preview for visually guided target aiming. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1415-1419), Santa Monica, CA, 1995.

Elias, B. The effects of spatial auditory preview on dynamic visual search performance. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 112-123), Santa Monica, CA, 1996.

Ellis, S. R.; Mania, K.; Adelstein, B. D.; Hill, M. Generalizeability of latency detection in a variety of virtual environments. Retrieved May 28, 2004, from http://www.sussex.ac.uk/Users/km3/hfes.pdf, 2004.

Endsley, M. R.; Rosiles, S. A. Vertical auditory localization for spatial orientation. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 55-59), Santa Monica, CA, 1995.

Ericson, M. A. Velocity judgments of moving sounds in virtual acoustic displays. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 3710-3713), Santa Monica, CA, 2000.

Ferguson, R. W.; Rasch, R. A.; Turmel, W.; Forbus, K. D. Qualitative spatial interpretation of course-of-action diagrams. *Proceedings of the 14th International Workshop on Qualitative Reasoning*. Retrieved May 28, 2004 from http://www.cc.gatech.edu/faculty/Ron.Ferguson/papers/ferguson-rasch-turmel-forbus--coa-diagrams--qrw-2000--dist.pdf, 2000.

Fisher, F. Lessons learned integrating speech control into embedded systems. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 3758-3761), Santa Monica, CA, 2000.

Fisher, F.N. Personal communication, U.S. Army Research Laboratory: Adelphi, MD, November 21, 2003.

Fitzpatrick, P.; Metta, G.; Natale, L.; Rao, S.; Sandini, G. What am I doing? Initial steps towards artificial cognition. *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan, May 12-17, 2003.

Folds, D. J.; Gerth, J. M. Auditory monitoring of up to eight simultaneous sources. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 505-509), Santa Monica, CA, 1994.

Fong, T.; Kaber, D.; Lewis, M.; Scholtz, J.; Shultz, A.; Steinfeld, A. Common metrics for human-robot interaction. *Proceedings of IEEE International Conference on Intelligent Robots and Systems* (Sendai, Japan). Retrieved February 23, 2004, from http://vrai-group.epfl.ch/papers/IROS04-TF.pdf, 2004.

Fong, T.; Thorpe, C.; Baur, C. Multi-robot remote driving with collaborative control. *IEEE Transactions on Industrial Electronic* **2003**, *54* (4), 699-704. Retrieved February 23, 2004, from http://vrai-group.epfl.ch/papers/ROMAN01-TF.pdf.

Fong, T.; Thorpe, C.; Glass, B. PdaDriver: A handheld system for remote driving. *IEEE International Conference on Advanced Robotic* (Coimbra, Portugal). Retrieved February 23, 2004, from http://vrai-group.epfl.ch/papers/ICAR03-TF.pdf, 2003.

Frank, L. H.; Casali, J. G.; Wierville, W. W. Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors* **1988**, *30*, 201-217.

French, J.; Ghirardelli, T. G.; Swoboda, J. The effect of bandwidth on operator control of an unmanned ground vehicle. *Proceedings of the Interservice/Industry training, Simulation, and Education Conferences (I/ITSEC)*, Orlando, FL, 2003.

Fujawa, G. E.; Strybel, T. Z. The effects of cue informativeness and signal amplitude on auditory spatial facilitation of visual performance. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 556-560), Santa Monica, CA, 1997.

Gemperle, F.; Ota, N.; Siewiorek, D. Design of a wearable tactile display. *Proceedings of the 5th IEEE International Symposium on Wearable Computers* (pp. 5-12), 2001.

Gilson, R. D.; Merlow, J.; Brill, C.; Stafford, S.; Mathews, M. Briefing and demonstration of TACTICS to the U.S. Army Research Laboratory, Human Research & Engineering Directorate Field Element at Fort Benning, GA, September 2005.

Glaskin, M. Driving by the seat of your pants. The Engineer. *United Kingdom: Centaur Communications*, 2004, May.

Goodrich, M. A.; Olsen, D. R. Seven principles of efficient human robot interaction. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3943-3948). Retrieved April 30, 2004, from http://faculty.cs.byu.edu/~mike/mikeg/papers/GoodrichOlsenCSMC2003.pdf, 2003.

Green, J.; Dougherty, E.; Savacool, E. Stereovision system for shipboard cranes. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conferences (I/ITSEC)*, Orlando, FL, 2003.

Haas, E. C.; Edworthy, J. *The ergonomics of sound: Selections from Human Factors and Ergonomics Society Annual Meetings*, 1985-2000, Santa Monica, CA, 2002.

Haas, E. C.; Gainer, C.; Wightman, D.; Couch, M.; Shilling, R. Enhancing system safety with 3D audio displays. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 868-872), Santa Monica, CA, 1997.

Haas, E. C.; Shankle, R.; Murray, H.; Travers, D.; Wheeler, T. Issues relating to automatic speech recognition and spatial auditory displays in high noise, stressful tank environments. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 3754-3757), Santa Monica, CA, 2000.

Heath-Pastore, T. Improved operator awareness of teleoperated land vehicle attitude (Tech. Rep. 1659). San Diego, CA: SPAWAR. Retrieved April 30, 2004, from http://www.spawar. navy.mil/sti/publications/pubs/tr/1659/tr1659.pdf, 1994.

Held, R.; Efstathiou, A.; Greene, M. Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology* **1966**, *72* (6), 887-891.

Horrey, W. J.; Wickens, C. D. Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors* **2004**, *46*, 611-624.

Hughes, S.; Lewis, M. Robotic camera control for remote exploration. *Proceedings of ACM CHI* (Vienna, Austria). Retrieved May 28, 2004, from http://usl.sis.pitt.edu/ ulab/chi04.pdf, 2004.

Johnson, C. A.; Adams, J. A.; Kawamura, K. Evaluation of an enhanced human-robot interface. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 900-905), 2003.

Johnston, J.; Wilson, R.; Birch, M. *Training for the operation of semi-autonomous systems--Interim Report;* Tech. Rep.: HSIAC-TR-2002-002; Human Systems Information Analysis Center: Wright Patterson Air Force Base, OH, 2002.

Jones, L.; Lockyer, B. *Tactile displays*. Presented at the Army Research Laboratory Meeting on the WTCU, Aberdeen Proving Ground, MD, August 2004.

Jung, J. Y.; Adelstein, B. D.; Ellis, S. R. Discriminability of prediction artifacts in a time delayed virtual environment. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 499-502), Santa Monica, CA, 2000.

Kaber, D. B.; Endsley, M. R. The effects of level of automation and adaptive automation on human performance, situational awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science* **2004**, *5* (2), 113-153. Retrieved March 26, 2004 from *http://www.csee.usf.edu/~murphy/Courses/RAP/Review/Set3/TIES-AA&LOA.pdf.*

Kaber, D. B.; Riley, J. M.; Zhou, R.; Draper, J. Effects of visual interface design, and control mode and latency on performance, telepresence and workload in a teleoperation task. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 503-506), Santa Monica, CA, 2000.

Kamsickas, G. *Future combat systems (FCS) concept and technology development (CTD) phase--Unmanned combat demonstration*; Tech. Rep.: D786-1006102; The Boeing Company: Seattle, WA, 2003.

Karis, D.; Dobroth, K. M. Psychological and human factors issues in the design of speech recognition systems. In A. Syrdal, R. Bennett, & S. Greenspan (Eds.), *Applied speech technology* (pp. 359-388). Ann Arbor, MI: CRC Press, 1995.

Kennedy, C. W.; Hu, T.; Desai, J. P.; Wechsler, A. S.; Kresh, J. Y. A novel approach to robotic cardiac surgery using haptics and vision. *Cardiovascular Engineering: An International Journal* **2002**, *2* (1), 15- 22. Retrieved May 28, 2004, from http://www.kluweronline.com/ *article.asp?PIPS=376550&PDF=1*.

Keskinpala, H. K.; Adams, J. A.; Kawamura, K. PDA-based human-robotic interface. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3931-3936), 2003.

Kheddar, A.; Chellali, R.; Coiffet, P. Virtual environment-assisted teleoperation. In K. Stanney (Ed.), *Handbook of virtual environment technology* (pp. 959-997). Mahwah, NJ: Erlbaum. Retrieved April 30, 2004, from http://vehand.engr.ucf.edu/handbook/ Chapters/Chapter56/ kay.html, 2002.

Kolasinski, E. M. *Simulator sickness in virtual environments*; Tech. Rep. 1027; U.S. Army Research Institute for the Behavioral and Social Sciences: Alexandria, VA, 1995. Retrieved February 13, 2004, from http://www.hitl.washington.edu/scivw/scivw-ftp/publications/ SimSick.rtf.

Lampton, D. R.; Singer, M. J.; McDonald, D. P.; Bliss, J. P. Distance estimation in virtual environments. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1268-1272), Santa Monica, CA, 1995.

Land, M. F.; Lee, D. Where we look when we steer. *Nature* **1994**, *369*, 742-744.

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.* Retrieved February 13, 2004, from www.ssl.umd.edu/publications/SSL02-003.pdf, 2002.

Lee, M. D. Multichannel auditory search: Toward understanding control processes in polychotic auditory listening. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 1343-1347), Santa Monica, CA, 1997.

Levin, D. T.; Simons, D. J. Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review* **1997**, *4* (4), 501-506.

Lewis, M.; Wang, J.; Manojilovich, J.; Hughes, S.; Liu, X. Experiments with attitude: Attitude displays for teleoperations. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1345-1349), 2003.

MacKenzie, S.; Ware, C. Lag as a determinant of human performance in interactive systems. *Proceedings of ACM Conference on Human Factors in Computing Systems-INTERCHI '93* (pp. 488-493). New York, NY: ACM SIGCHI. Retrieved May 27, 2004, from http://www.yorku.ca/mack/p488-mackenzie.pdf, 1993.

Malcolm, A. A.; Lim, J.S.G. Teleoperation control for a heavy-duty tracked vehicle. Retrieved June 21, 2004, from http://www.simtech.a-star.edu.sg/Research/TechnicalReports/TR0369.pdf, 2003.

Malin, J. T. Preparing for the unexpected: Making remote autonomous agents capable of interdependent teamwork. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society/IEA 2000* (pp. 254-257), Santa Monica, CA, 2000.

Marble, J. L.; Bruemmer, D. J.; Few, D. A. Lessons learned from usability tests with a collaborative cognitive workspace for human-robot teams. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics* (pp. 448-453), 2003.

Markoff, J. Robotic vehicles race, but innovation wins. The New York Times (NYTimes.com). Retrieved September 14, 2005, from http://www.nytimes.com/2005/09/14/business/14robot.html, September 14, 2005.

Massimino M.; Sheridan T. B. Teleoperator performance with varying force and visual feedback. *Human Factors* **1994**, *36*, 145-157.

McGovern, D. E. Experience and results in teleoperation of land vehicles. In S. R. Ellis, M. K. Kaiser and A. C. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 182-195). London: Taylor and Francis, 1991.

McKinley, R.; Ericson, M.; D'Angelo, W. 3-Dimensional auditory displays: Development, applications, and performance. *Aviation, Space, and Environmental Medicine* **1994**, *5*, A31-A38.

Meehan, M.; Razzaque, S.; Whitton, M. C.; Brooks, R. P. Effect of latency on presence in stressful virtual environments. Retrieved June 1, 2004, from http://www.cs.unc.edu/~whitton/ExtendedCV/Papers/2003-VR2003-Meehan.pdf, 2003.

Morrison, H. B.; Casali, J. G. Intelligibility of synthesized voice messages in commercial truck cab noise for normal-hearing and hearing-impaired listeners. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 801-805), Santa Monica, CA, 1994.

Murphy, R. R. Human-robot interaction in rescue robotics. *IEEE Systems, Man and Cybernetics Part C: Applications and Reviews, Special Issue on Human-Robot Interaction* **2004**, *34* (2), 138-153. Retrieved April 16, 2004, from http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2004-2.pdf.

Myers, G. K.; Cowan, C. *Robotic collaborative technical alliance quarterly technical status report*; Project 8752-079-43-25, 7/01/03-9/30/03; Stanford Research Institute International: Menlo Park, CA, 2003.

Nguyen, L. A.; Bualat, M.; Edwards, L. J.; Flueckiger, L.; Neveu, C.; Schwehr, K.; Wagner, M. D.; Zbinden, E. Virtual reality interfaces for visualization and control of remote vehicles. *Autonomous Robots* **2001**, *11*, 59-68.

Nielsen, J.; Molich, R. Heuristic evaluation of user interfaces. *Proceedings of the ACM CHI 1990 Conference on Human Factors and Computing Systems* (pp. 1-5), 1990.

Noyes, J. M.; Baber, C.; Leggatt, A. P. Automatic speech recognition, noise and workload. *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 3762-3765), Santa Monica, CA, 2000.

Noyes, M.; Sheridan T.B. A novel predictor for telemanipulation through a time delay. *Proceedings of the 20th Annual Conference on Manual Control*, NASA Ames Research Center: Moffet Field, CA, 1984.

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. *International Journal of Aviation Psychology* **2000**, *10* (3), 247-271.

Olsen, D. R.; Goodrich, M. A. Metrics for evaluating human-robot interactions. Proceedings of PERMIS 2003. Retrieved May 25, 2004 from http://faculty.cs.byu.edu/~mike/mikeg/papers/OlsenGoodrichPERMIS2003.pdf, 2003.

Oving, A. B.; Van Erp, J.B.F. Driving with a head-slaved camera system. *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1372-1376), Santa Monica, CA, 2001.

Parasuraman, R.; Sheridan, T. B.; Wickens, C. D. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics* **2000**, *30* (3), 286-297.

Perrott, D. R.; Cisneros, J.; McKinley, R. L.; D'Angelo, W. R. Aurally aided detection and identification of visual targets. *Proceedings of the 39th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 104-108), Santa Monica, CA, 1995.

Perzanowski, D.; Brock, D.; Adams, W.; Bugajska, M.; Schultz, A. C.; Trafton, J. G.; Blisard, S.; Skubic, M. Finding the FOO: A pilot study for a multimodal interface. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3218-3223), 2003.

Quigley, M.; Goodrich, M. A.; Beard, R. W. Semi-autonomous human-UAV interfaces for fixed-wing mini-UAVs. *Proceedings of IROS 2004.* Retrieved August 26, 2004, from http://faculty.cs.byu.edu/~mike/mikeg/papers/QuigleyGoodrichCIROS2004.pdf, 2004.

Rabiner, L. W. Voice communication between humans and machines-An introduction. In D. B. Roe & J. G. Wilpon (Eds.), *Voice communication between humans and machines* (pp. 5-12). Washington, DC: National Academy Press, 1994.

Rastogi, A. *Design of an interface for teleoperation in unstructured environments using augmented reality displays*. Unpublished master's thesis, University of Toronto. Retrieved August 5, 2004, from http://vered.rose.utoronto.ca/people/anu_dir/thesis/thesis.html, 1996.

Reddy, M. *The effects of low frame rate on a measure for user performance in virtual environments*; Tech. Rep. ECS-CSG-36-97; Department of Computer Science, University of Edinburgh. Retrieved June 2, 2005, from http://www.martinreddy.net/pubs/pdf/ECS-CSG-36-97.pdf, 1997.

Rehfeld, S. A.; Jentsch, F. G.; Curtis, M.; Fincannon, T. Collaborative teamwork with unmanned ground vehicles in military missions. *Proceedings of the 11th Annual Human-Computer Interaction International Conference*, Las Vegas, NV, August 2005.

Richard, P.; Birebent, G.; Coiffet, P.; Burdea, G.; Gomez, D.; Langrana, N. Effect of frame rate and force feedback on virtual object manipulation. *Presence* **1996**, *5*, 95-108. Retrieved June 23, 2005, from http://www.caip.rutgers.edu/vrlab/publications/papers/1996_presence.pdf.

Ricks, B.; Nielsen, C. W.; Goodrich, M. A. Ecological displays for robot interaction: A new perspective. Retrieved August 5, 2004, from http://faculty.cs.byu.edu/~mike/mikeg/papers/RicksGoodrichCIROS2004.pdf, 2004.

Robotics Collaborative Technology Alliances. Robotics collaborative technology alliances. *Proceedings of Collaborative Technology Alliances Conference 2004*, Washington, DC, 2004.

Rosenberg, L. B. The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks. *Proceedings of IEEE Virtual Reality Annual International Symposium* (pp. 27-32), 1993.

Schenker, P. S.; Huntsburger, T. L.; Pirjanian, P.; McKee, G. T. Robotic autonomy for space: Cooperative and reconfigurable mobile surface systems. Retrieved April 30, 2004, from http://robotics.jpl.nasa.gov/people/terry/papers/AM003_paper.pdf, 2001.

Schiffman, H. R. *Sensation and perception: An integrated approach* (5th ed.). New York, NY: Wiley, 2001.

Schipani, S. P. An evaluation of operator workload during partially-autonomous vehicle operations. *Proceedings of PerMIS 2003.* Retrieved February 23, 2004, from http://www.isd.mel.nist.gov/research_areas/research_engineering/Performance_Metrics/PerMIS_2003/Proceedings/Schipani.pdf, 2003.

Scholtz, J.; Bahrami, S. Human-robot interaction: Development of an evaluation methodology for the bystander role of interaction. Retrieved February 23, 2004, from http://www.isd.mel.nist.gov/documents/scholtz/Scholtz_Bahrami.pdf, 2003.

Schreckenghost, D. K. Checklists for human-robot collaboration during space operations. *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 46-50), Santa Monica, CA, 1999.

Scribner, D. R.; Gombash, J. W. *The effect of stereoscopic and wide field of view conditions on teleoperator performance*; ARL-TR-1598; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 1998.

Sekmen, A.; Koku, A. B.; Zein-Sabatto, S. Human robot interaction via cellular phones. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics* (pp. 3937-3942), 2003.

Setalaphruk, V.; Ueno, A.; Kume, I.; Kono, Y. Robot navigation in corridor environments using a sketch floor map. *Proceeding of the 2003 IEEE International Symposium on Computation Intelligence in Robotics and Automation,* Kobe, Japan (pp. 552-557). Retrieved September 23, 2005, from http://ai-www.naist.jp/papers/setala-v/EX/CIRA2003/Paper/cira2003-setala-v.pdf, 2003.

Sheridan, T. B. *Telerobotics, automation, and human supervisory control*. Cambridge, MA: The MIT Press, 1992.

Sheridan, T. B. *Humans and automation: System design and research issues*. Hoboken, NJ: John Wiley & Sons, 2002.

Sheridan, T. B.; Ferrell, W. B. Remote manipulative control with transmission delay. *IEEE Transactions on Human Factors in Electronics* **1963**, *4*, 25-29.

Shilling, R. D.; Cunningham, B. Virtual auditory displays. In K. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 65-92). Mahwah, NJ: Lawrence Erlbaum Associates, 2002.

Skubic, M.; Anderson, D.; Blisard, S.; Perzanowski, D.; Adams, W.; Trafton, J. G.; Schultz, A. C.; Thomas, S.; Cross, E. V. II. Using a sketch pad interface for interacting with a robot team. Retrieved September 14, 2005, from http://www.cecs.missouri.edu/ ~skubic/flyer.pdf, 2005.

Skubic, M.; Bailey, C.; Chronis, G. A sketch interface for mobile robots. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 919-924), 2003.

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*; ARL-TR-2511; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2001.

Smyth, C. C.; Paul, V.; Meldrum, A. M.; McDowell, K. *Examining Alternative Display Configurations for an Indirect Vision Driving Interface;* (in process); U.S. Army Research Laboratory: Aberdeen Proving Ground, MD.

Stanney, K. M.; Mourant, R. R.; Kennedy, R. S. Human factors issues in virtual environments: A review of the literature. *Presence, 7,* 327-351. Retrieved April 30, 2004, from http://web.mit.edu/16.459/www/Stanney.pdf, 1998.

Steinfeld, A. Interface lessons for fully and semi-autonomous mobile robots. *Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA).* Retrieved August 26, 2004, from http://www-2.cs.cmu.edu/~astein/pub/HRI-Lessons-ICRA04pre.pdf, 2004.

Strybel, T. Z.; Boucher, J. M.; Fujawa, G. E.; Volp, C. S. Auditory spatial cueing in visual search tasks: Effects of amplitude, contrast, and duration. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 109-113), Santa Monica, CA, 1995.

Terrence, P. I.; Brill, J. C.; Gilson, R.D. Body orientation and the perception of spatial auditory and tactile cues. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 1663-1667), Santa Monica, CA, 2005.

Tholey, G.; Pillarisetti, A.; Green, W.; Desai, J. P. Design, development, and testing of an automated laparoscopic grasper with 3-D force measurement capability. Retrieved August 31, 2004, from http://www.medicalsim.org/symposium2004/final_pdf/47.pdf, 2004.

Thomas, L. C.; Wickens, C. D. *Effects of display frames of reference on spatial judgments and change detection*; ARL-00-14/FED-LAB-00-4; University of Illinois, Institute of Aviation, Aviation Research Laboratory: Urbana-Champaign, IL, 2000. Retrieved April 4, 2005, from *http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/TechReport/00-14.pdf*.

Thompson, W. B.; Willemsen, P.; Gooch, A. A.; Creem-Regehr, S. H.; Loomis, J. M.; Beall, A. C. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence* **2004**, *13,* 560-571.

Tittle, J. S.; Roesler, A.; Woods, D.D. The remote perception problem. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 260-264), Santa Monica, CA, 2002.

Van Erp, J.B.F. Guidelines for the use of vibro-tactile displays in human computer interaction. In S. A. Wall, B. Riedel, A. Crossman, & M. R. Mcgee (Eds.), *Proceedings of Eurohaptics 2002* (pp. 18-22), University of Edinburgh, 2002.

Van Erp, J.B.F.; Meppelink, R.; van Veen, H. A. *A touchy car* (DTIC TNO-TM-02-B001, pp. 1-19). Soesterberg, Netherlands: Human Factors Research Institute, TNO, 2002.

Van Erp, J.B.F.; Padmos, P. Image parameters for driving with indirect viewing systems. *Ergonomics* **2003**, *46*, 1471-1499.

Van Erp, J.B.F.; Veltman, J. A.; van Veen, H.A.H.C.; Oving, A. B. *Tactile torso display as countermeasure to reduce night vision goggles induced drift*; DTIC, X5-X5, pp. 1-8; Soesterberg, Netherlands: Human Factors Research Institute, TNO, 2003.

Vidulich, M. A.; Bortolussi, M. R. A dissociation of objective and subjective workload measures in assessing the impact of speech controls in advanced helicopters. *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (pp. 1471-1475), Santa Monica, CA, 1988.

Vitense, H. S.; Jacko, J. A.; Emery, V. K. Multimodal feedback: An assessment of performance and mental workload. *Ergonomics* **2003**, *46,* 68-87.

Vogels, I.M.L.C. Detection of temporal delays in visual-haptic interfaces. *Human Factors* **2004**, *46*, 118-134.

Wang, J.; Lewis, M.; Hughes, S. Gravity-referenced attitude display for teleoperation of mobile robots. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 2662-2666). Retrieved September 14, 2005, from http://usl.sis.pitt.edu/ulab/HFES2004-attitude.pdf, 2004.

Watson, B.; Walker, N.; Ribarsky, W.; Spaulding, V. Effects of variation in system responsiveness on user performance in virtual environments. *Human Factors* **1998**, *40,* 403-414.

Webster's Online Dictionary. Philip M. Parker and Webster's Online Dictionary (http://www.webster-dictionary.org/), 2004.

Wenzel, E. M.; Wightman, F. L.; Foster, S. H. A virtual display system for conveying three-dimensional acoustic information. *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (pp. 86-90), Santa Monica, CA, 1988.

Williamson, D. T.; Barry, T. P. The design and evaluation of a speech interface for generation of air tasking orders. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 3762-3765), Santa Monica, CA, 2000.

Witmer, B. G.; Kline, P. B. Judging perceived and traversed distance in virtual environments. *Presence* **1998**, *7* (2), 144-167.

Witmer, B. G.; Sadowski, W. J. Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors* **1998**, *40*, 478-488.

Woods, D. D.; Tittle, J.; Feil, M.; Roesler, A. Envisioning human-robot coordination in future operations. *IEEE Transactions on Systems, Man & cybernetics: Part C: Special Issue on Human-Robot Interaction.* Retrieved April 21, 2004 from http://csel.eng.ohio-state.edu/woods/papers/woods_hri_IEEE.pdf, 2004.

Yanco, H. A.; Drury, J. L. A taxonomy for human-robot interaction. *Proceedings of AAAI Fall Symposium on Human-Robot Interaction*, AAAI Technical Report FS-02-03 (pp. 111-119). Retrieved August 5, 2004, from http://www.cs.uml.edu/~holly/papers/yanco-drury-taxonomy-fss02.pdf, 2002.

Zelek, J. S.; Asmar, D. A robot's spatial perception communicated via human touch. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 454-461), 2003.

Appendix A. Comparison of Microphone Technologies for HRI Systems

Manufacturer	Product	Comments
	Throat Microphone Systems	
Blue Kangaroo Technologies U.S. Contact: Glen Thomson Glen@BlueKangarooTech-nologies.com 801-400-0415 (Utah) AUSTRALIA BlueKangarooTechnologies.com	**Noise Terminator Microphone** Piezo-electric microphone	• Microphone rests on the neck. Compatible with many two-way radios and mobile phones. • Currently incorporating a VOX switch (as an alternative to PTT-switch) in their systems. • Not yet tested with any ASR systems. Local rep Glen Thomson said he asked the Company for such testing.
Holm Co 011-030-617-800 Berlin, Germany www.HolmCo.De	**Dynamic Throat Microphone** Series 71-02	• 300- to 3400-Hz Frequency Range, 16-dB signal-to-noise ratio with a 115-dBA noise level. For use in high ambient noise.
Sytronics 4433 Dayton-Xenia Road Dayton, OH 45432 937-431-6100 1-800-699-1466 www.Sytronics.Com	**VHIC** Voice-Head Input Controller, a system being developed for the Air Force.	• This is the only "system" in this table, consist-ing of sensor hardware and speech recognition software. • Speech and head-motion input are integrated in a wearable computer, with head tracker and speech-enabled input system, using available DragonNaturallySpeaking Software for ASR. • System uses a finite subset of computer application manipulating commands. • Software not robust in noisy environments.
Pryme Radio Products 80 Apollo St. #E Brea, CA 92821 Phone: 714-257-0300 www.pryme.Com	**Throat Microphone SPM-500** Dual Electret Condenser microphone	• Used in Sytronics VHIC design. Can be operated while users are wearing gloves. • SPM-500 discontinued, a newer version SPM-1500 series, undergoing development. • No product available at the moment.
Tactical Command Industries 1872 Verne Roberts Circle Antioch, CA 94509 (888)-990-1600 www.TacticalCommand.Com	**Tactical Throat Microphone Headset** with TCI Tactical PTT	• Manufacturer recommends this headset primarily for operations employing gas masks and respirators because throat microphone is very compatible with both. • Can be worn under any helmet, mask, or hood and does not interfere with peripheral hearing or weapon positioning.
Radio Accessory Headquarters 6119-A 28th St Sacramento, CA 1-888-438-7427 www.RAHQ.Com	**SR56i Throat Microphone**	• Waterproof and dust tight; PTT button can be placed in large number of locations; remote PTT capability; offers interfaces for many radios. • Frequency response is 300 to 3000 Hz.
	Bone Conduction Systems	
New Eagle Hwy 24 & Madore St PO Box 250, Silverlake, KS 66359 800-850-8512 www.NewEagle.Com	**Enforcer Series I and II** Special Operations Version	• Compatible with most two-way radios, helmets, and gas masks; single or dual vibration versions, several PTT switch types, in-line disconnects, volume control, and other accessories. • 250- to 4000-Hz frequency range; recommends BC receivers be used with standard (acoustic) microphones--for better response in high noise environments because the movement of a BC microphone relative to the body, in vigorous activity is considered unsatisfactory. • Supplied systems to a large number of law enforcement and military customers, with testimonials from them on their web site.

Sensory Devices Inc 205 Main Street New Eagle, PA 15067 724-258-5353 Wpiroth@SensoryDevices.com Harold Holsopple, President HHolsopple@SensoryDevices.com	**Radio Ear** PiezoElectric polymer film based Microphone contacting the head	• Has a large frequency range of 300 to 4000 Hz. • Originally developed for U.S. Navy SEALs. Demonstrated in annual MockPrisonRiots (2000-2003), and Fire Chiefs Conference. Tested to MIL-STD-810 (an environmental spec - temp, humidity, shock...)
Temco Temco Communications Inc. 13 Chipping Campden, South Barrington, Illinois 60010, USA Phone: +1-847-359-3277 Fax:+1-847-359-3743 www.Temco-j.co.jp	**VoiceDucer** Bone Conduction microphone & receiver combination	• In a hearing-aid type package. Has another version where the microphone is attached to the head. • Allows simultaneous talk/listen operation. • Has "equalizer" circuitry built into the amplifier to produce sound with good clarity. • Have ongoing studies as to performance, with an ARL team.
Tactical Command Industries 1872 Verne Roberts Circle Antioch, CA 94509 (888)-990-1600 www.TacticalCommand.Com	**Tactical Assault Bone Conduction Headset**	• Uses bone conduction for microphone and receivers; compatible with gas masks and respirators; has PTT device
PerCom P.O. Box 15437 New Lynn, Auckland 1007 NEW ZEALAND 011-64-9-827-7667 www.percom2000.Com	**Miniature Inertial Transducers** Series 17 (microphone & receiver) Series 31 (receiver) TearDrop (receiver)	• Used in direct contact with the user's neck or head. Head and headband mountable. Interface amplifiers are available to compensate for placement on different positions on the head. • Series 17 has a frequency range of 300 to 7000 Hz. Tear Drop has best freq range (500 to 14 kHz) but is used mainly for hearing aid devices.

Appendix B. Comparison of Speech Synthesis and Recognition Technologies

Product & Manufacturer	Description	Comments
	Speech Recognition	
ShortTalk AT&T Bell Labs Government Markets Stephen Robinson (703) 691-5522 www.research.AT T.com	• TTS • Uses terse dictation-specific commands	• ShortTalk is a new method for composing text by speech. This spoken command language is carefully designed to be rewarding to use, right from the beginning. In contrast to so-called "natural language technology" of available dictation systems, ShortTalk can be fluently interspersed with dictation. There are no cumbersome phrases such as "go to the beginning of the line." Instead, ShortTalk codifies natural and universal editing concepts that can be combined in command phrases, typically consisting of only two syllables.
	Speech Recognition/Speech Synthesis	
WHISPER (Recognition) **WHISTLER** (Synthesis) MicroSoft Microsoft Corporation One Microsoft Way Redmond, WA 98052-6399 (425) 882-8080 www.Research.Mi crosoft.com	• Speaker-independent continuous speech recognition • Trainable TTS engine • Uses Waveform concatenation for low level synthesis	• Microsoft Speech Recognition engine, code named WHISPER, offers state-of-the-art speaker-independent continuous speech recognition. The WHISPER speech engine has been shipped by Speech.Net as part of the SAPI SDK, which in turn has been shipped in Microsoft Phone and Microsoft Agent, Microsoft Encarta, Windows 2000, Office XP and Windows XP. • Microsoft's Speech Synthesis engine, code named WHISTLER, is a "trainable" TTS engine which was released in 1998 as part of the SAPI 4.0 SDK, and then as part of Microsoft Phone and Microsoft Encarta and Windows 2000 and Windows XP operating systems. You type words on your keyboard, and the computer reads them back to you almost immediately. Although it still has that distinct machine sound, it's a big improvement in the flat, robotic voices of the past, particularly when large voice inventories are used.
	Speech Synthesis	
Mixed Excitation Linear Prediction (MELP) Embedded Systems (Software) Vocal Technology 200 John James Audubon Pkwy Buffalo, NY 14228 716-688-4675 www.Vocal.com	• Robust in high-noise environments • Selected by DoD Digital Voice Processing Consortium • Developed by Texas Instruments based on research at Georgia Tech, under a DoD contract. Computational efficiency results in low power consumption, advantageous for portable systems • Uses Formant Synthesis	• The MELP Vocoder is based on the traditional LPC parametric model, but also includes four additional features. These are mixed excitation, aperiodic pulses, pulse dispersion, and adaptive spectral enhancement.
MBROLA TCTS Lab Faculté Polytechnique de Mons	• Low level speech synthesizer based on the concatenation of diphones, together with prosodic information (duration of phonemes and a piecewise	• MBROLA is a speech synthesizer based on the concatenation of diphones. It takes a list of phonemes as input, together with prosodic information (duration of phonemes and a piecewise linear description of pitch), and produces speech samples on 16 bits (linear), at the sampling frequency of the diphone database used (it is

1, Copernic Ave, B-7000 Mons, Belgium tel : +32-65-374733 http://tcts.fpms.ac.be/synthesis/mbrola.html	linear description of pitch) • Handles several languages with a built-in structure to add more • Used by many other high level synthesizers, accepted as a standard • Uses Waveform concatenation for low level synthesis	therefore NOT a TTS synthesizer since it does not accept raw text as input). • This synthesizer is provided free for non-commercial, non-military applications only.
Embedded System/Speech Synthesis		
InterSound Rev 2.0 Intel Corp. 2200 Mission College Blvd. Santa Clara, CA 95052 800-628-8686 http://appzone.intel.com/pcadn/product.asp?productid=473	• Smooth synthesized speech in embedded systems • Works with many operating systems • Computer-based training, intelligent information terminals, toys, GPS in automobiles, military systems • Makes chipsets for other system developers • Uses Waveform concatenation for low level synthesis	• InterSound CN Rev2.0 Speech Synthesis System is the newest product developed by iFLYTEK, which can provide smooth synthesized speech on embedded devices. Employing a high efficient voice library compressing technology and text analysis technology, this system performs much better than its previous version InterSound CN Rev1.0 and retains a smaller voice library.
Embedded System/Speech Recognition + Speech Synthesis		
Interactive Speech™ integrated circuits (IC) chipset for embedded system development Logic Plus 1125 Garden Street San Luis Obispo, CA 805-783-2550 www.Logic-Plus.com	• A research environment for development of general multi-lingual speech synthesis techniques • TTS with application programming interface (API) interface • English/Spanish/Welsh TTS • Externally configurable language-independent modules • Diphone based, residual excited LPC • MBROLA database support • Portable UNIX® distribution, free and unrestricted	• A premier developer for Sensory's voice recognition technologies. Sensory develops highly integrated, low cost speech recognition IC and embedded software technology. Their Interactive Speech™ line of ICs offers industry-leading accuracy for small vocabulary C2 applications. • Logic Plus has achieved a toy industry focus and premier status with their most recent accomplishments in area of electronic toys. From concept to completion, their projects include electronics hardware design, embedded systems, software/firmware for clients such as Mattel®, DSI Toys, Fisher-Price® and more. Their projects include the domestic and international versions of the Diva Starz for Mattel, Cube It Up! for Toy Biz, and eBrain for DSI Toys.
Text-to-Speech		
SPRUCE University of Essex Eric Lewis or Mark Tatham Colchester, U.K. 44-117-928-7954 http://www.cs.bris.ac.uk/~eric/research/spruce97.html	• High level synthesizer, designed to work independently with any low level synthesizer, formant or waveform-concatenation systems • TTS synthesis • Is a research project not yet transitioned to a commercial product	• Can drive both forms of low level synthesizers TTS synthesis allows a computer to read text aloud without the direct use of recordings of human speech. Even when there is an indirect use of recordings (as in waveform concatenation), an essential property of the system is that it should be able to speak sentences which have not been recorded. SPRUCE is a high-level TTS synthesis system, which has these properties.

Festival Version 1.4.3 (Jan 2003) Univ of Edinburgh Cntr for Speech Tech Research Edinburgh EH8 9LW Tel: +44 131 651 1767 www.cstr.ed.ac.uk/projects/festival	• A research environment for development of general multi-lingual speech synthesis techniques • TTS with API interface • English/Spanish/Welsh TTS • Externally configurable language-independent modules • Diphone based, residual excited LPC • MBROLA database support • Portable UNIX® distribution, free and unrestricted	• Festival is a general multi-lingual speech synthesis system developed at CSTR. It offers a full TTS system with various APIs, as well as an environment for development and research of speech synthesis techniques. It is written in C++ with a scheme-based command interpreter for general control.
Integrated Suite *(Speech Recognition + Speech Synthesis + TTS + Speech-to-Text)*		
Galaxy MIT Marcia Davidson, Spoken Language Systems MIT Computer Science and Artificial Intelligence Laboratory Cambridge, MA (617) 253-3049 www.sls.csail.mit.edu/Galaxy.html	• A suite of speech-based applications • Recognition, understanding, information retrieval, language generation, and synthesis • Data retrieval from several domains of knowledge to answer queries • Uses waveform concatenation for low level synthesis	• The GALAXY system is a project in the Spoken Language Systems group attempting to leverage recent advances in conversational systems to provide a spoken language interface for on-line information. GALAXY differs from current spoken language systems in a number of ways. • It is distributed and decentralized. GALAXY uses a client-server architecture to allow sharing of computationally expensive processes (such as large vocabulary speech recognition), as well as knowledge intensive processes. • It is multi-domain, intended to provide access to a wide variety of information sources and service while insulating the user from the details of database location and format. • It is extensible, new knowledge domain servers can be added to the system incrementally.
SPHINX-2, and -3 Open Source Software Carnegie Mellon University Sphinx Group, Kevin A. Lenzo Pittsburgh, PA http://www.speech.cs.cmu.edu/	• DARPA-funded long term research for the creation of speech tools and applications and to advance the state-of-the-art in speech recognition, dialog systems, and speech synthesis • Various components of SPHINX feature speech recognition, synthesis, pronunciation dictionary, dialog system, VoiceXML browser, V-Mail for dictation • A UNIX® version of SPHINX is downloadable freely from a CMU site	• Uses Formant Synthesis Sphinx-2, a real-time, large vocabulary, speaker-independent speech recognition system is free software under the Apache-style license. Sphinx-2 is the engine used in the Sphinx Group's dialog systems that require real time speech interaction, such as the implementation of the DARPA communicator project, a many-turn dialog for travel planning. The pre-made acoustic models include American English and French in full bandwidth, and telephone-bandwidth communicator models; Sphinx-2 is a decent candidate for hand-held, portable, and embedded devices, and telephone and desktop systems that require short response times.

SpeechWorks 6.5SE **DirectoryAssistance** **RealSpeak** ScanSoft (Belgium) 9 Centennial Drive Peabody, MA 01960 978-977-2000 www.ScanSoft.com	• Interactive voice-response technology • Multiple language support (French, Spanish, Cantonese, Mandarin, German, Korean, Japanese, Portuguese, and various flavors of English) • Used for automated handling of customer calls at United Airlines, FedEx®, America Online®, ...	• SpeechWorks 6.5SE (Second Edition) is a comprehensive software product for building network-based speech recognition services. The product is based on award winning ScanSoft® technology, which is powering leading speech services worldwide at corporations such as America Online®, FedEx®, TD Waterhouse Australia, United Airlines, and WorldCom among many others. SpeechWorks 6.5SE supports a range of widely available hardware platforms and scales to thousands of phone lines.
Nuance - Speech recognition **Vocalizer** - TTS **Verifier** - Speaker authenticator Nuance 1005 Hamilton Court Menlo Park, CA 94025 650-847-0000 www.Nuance.com	• Speech recognition and speaker verification systems • Second-generation software (allows callers to speak freely when interacting with voice automation systems) • Used in automated call centers • Uses waveform concatenation for low level synthesis	• Came out of SRI • Commercialized the industry's first speech recognition engine and deployed the industry's first large-scale system working with Charles Schwab & Company. Verifier voiceprinting technology used around the world for security.
iCommunicator ™ Teltronics, Inc 7108 Fairway Drive Palm Beach Gardens, FL 33418 800-245-2133 www.MyiCommunicator.com	• Efficient real time translation • Speech-to-text, TTS, speech-to-VSL, text-to-VSL,… • Works with desktop and notebook PCs • Uses waveform concatenation for low level synthesis	• iCommunicator™ software program converts spoken language into sign language. This very powerful tool provides a multi-sensory, interactive communication solution for persons who are deaf or hard of hearing and other persons who experience unique communication challenges. It efficiently converts in real time: speech-to-text, speech- to-VSL, speech to computer-generated voice, text to computer-generated voice or VSL. • The iCommunicator™'s unique technological features provide end users with unparalleled opportunities to achieve efficient, effective communication in most natural environments. A simple point and click using the iText tool allows end users to simply point and click to have email, web pages, and documents created in other applications signed and/or spoken through the iCommunicator™ program. • ScanSoft®'s Dragon NaturallySpeaking® 7 software provides the speech engine for Teltronics' latest release of the iCommunicator™ V.4.0 technology, which is marketed by 1450, Inc. This truly revolutionary device offers people who are hard of hearing or deaf effective, independent two-way communication with the hearing world. By translating speech into text, sign language, and a synthesized voice, persons who are hard of hearing or deaf can communicate freely. Used in K-12 education, post-secondary educational institutions, corporate, government, and healthcare environments, as well as public access sites, the iCommunicator™ enables end-users to leverage speech technology to increase independence and fully participate in all types of communication situations.

Embedded System/Integrated Suite (Speech Recognition + Speech Synthesis + TTS + Speech-to-Text)		
Aurix® **20/20 Speech** 20/20 Speech Ltd 1215 Jefferson Davis Hwy, #1102 Arlington, VA 22202 703-414-8160 www.Aurix.Com	• Designed and tested to U.K. military standards • Hands- and eyes-free operation of complex C2 • Military markets for speech technology, legal transcription • Uses waveform concatenation for low level synthesis	• 20/20 Speech's 'Speech in Media' speech processing tools automate the manipulation of speech-based content, providing improved flexibility and productivity for content generators and managers. The ability of this technology to detect when speech is present, what is being said and who is speaking can be of benefit wherever there is a need to analyze large volumes of spoken material, whether it be in call centers, broadcasting and media, or legal sectors.
FluentSpeech™ Sensory, Inc. 1991 Russell Avenue Santa Clara, CA 95054 408-327-9000 www.SensoryInc.com	• VoiceActivation™ for embedded speech recognition, TTS, AnimatedSpeech™ for synchronizing speech for animated characters • Makes chipsets and module level boards for original equipment manufacturers • Uses waveform concatenation for low level synthesis	• Recognition into consumer electronics with small to medium scale processing platforms. Requires only 40 MIPs and 100KB+ ROM. Applications include telephony (e.g., cordless handsets, telephone answering devices, cell phones), automotive (e.g., hands free kits, entertainment systems, navigation), and handheld (e.g., PDAs, music players, pagers).

INTENTIONALLY LEFT BLANK

Appendix C. Comparison of Tactile Display Technologies

Organization	Product	Description
		Research & Development Organization
Carnegie Mellon University Dr. Francine Gemperle CMU, Pittsburgh, PA 15213 Gemperle@CMU.edu	**Wearable Vibro-Tactile Vest**	• Uses an array of vibrating tactor motors. • Tested in navigation applications. • Suggests guidelines for tactor design requirements, placement of tactors on different areas of the body (clavicle, ribcage, forearm, pelvis, and shin), design of tactor arrays and tactile icons, and information coding.
MIT Dr. Lynette Jones 77 Massachusetts Avenue Cambridge, MA 02139 617-253-3973 www.MIT.edu lJones@MIT.Edu	**Wearable Tactile Display**	• Uses vibrating tactor motors mounted on a vest or arm band. • Also working with shape memory alloy actuators and thermal displays.
Johns Hopkins University Dr. Allison Okamura, Latrobe Hall 3400 North Charles Street Baltimore, MD 410-516-7266 www.JHU.edu	**General Haptics Research**	• Works with Telerobotics and VE for medical applications.
Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO)/Netherlands Dr. Jan Van Erp Kampweg 5, PO Box 23 3769 ZG Soesterberg The Netherlands Phone: +31 (0)346 356 211 www.tno.nl	**Tactile Torso Display** for helicopter hover tasks **Haptics Vest**, modified for automotive applications	• Designed and used tactile torso displays in many different applications to present simple and complex information. • Developing a car seat embedded with actuators that provide navigational information to the driver. • Developed guidelines for the design of tactile displays that include parameters such as actuator amplitude, location, frequency, information coding, and safety considerations for vibration and temperature.
University of Central Florida Dept. of Psychology Orlando, FL 32826 Dr. Richard Gilson 407-823-2755 gilson@mail.ucf.edu	**TACTICS** (Tactile Communication System) Wearable Belt	• 8 rugged waterproof tactors fitted around mid-waist • Sewn in tough stretchable fabric • Wireless PDA control • Covert messaging • Low interference with visual and auditory tasks • Superiority in degraded conditions
University of Central Florida Institute for Simulation & Training Orlando, FL 32826 Don Washburn 407-882-1433 dWashbur@IST.UCF.Edu	**HAMMER** Wearable Vest	• Haptic applications for multimodal environments research. • Uses 32 vibrators in 8 zones around the torso, in a sleeveless drysuit. • Also uses a CyberGrasp Force Feedback Glove by Immersion Technologies.
Stanford University Dr. Katherine Kuchenbaker	**Haptic Research**	• Haptic display of contact location in Telerobotics. • Thimble based mechanism attached to a PHANTOM robotic arm.

Stanford, CA Katherine.Kuchenbaker@ Stanford.edu		
University of Wisconsin Medical School Dr. Paul Bach-y-Rita Madison, WI 53792 pBachyri@FacStaff.Wisc. Edu	**Tactile Vision Substitution Systems**	• Form perception with a 49-point electro-tactile array mounted on the tongue.
Sandia National Labs Ms. Arthurine Breckenridge ILAB, Sandia Labs Albuquerque, NM 87185 505-284-2001 www.Sandia.Gov	**High Density Tactile Array**	• Impulsive, vibratory. • 2x3 array of electromagnetic actuators. • Sandia Labs hosted the 7th PHANTOM Users Group Workshop in 2002.
U.S. Navy Naval Aerospace Medical Research Laboratory 51 Hovey Road, Pensacola, FL 32508- 1046 850-452-4496 CPT Angus Rupert, USN aRupert@namrl.navy.mil	**Tactile Situational Awareness System (TSAS)**	• Developed a vibro-tactile vest to give pilots feedback on spatial orientation in a 3-D space. A matrix of tactors was embedded in the vest, and transmits information in several modes. • An upgraded form of the vest was tested with Navy Special Forces in 2003, in high altitude, high opening parachute operations. Also on ground and underwater.
	Hardware & Software Vendors	
SensAble Technologies 15 Constitution Way Woburn, MA 01801 781-937-8315 www.Sensable.com	**PHANTOM** Haptic Devices and Toolkits	• Manufacturer of a line of PHANTOM® Haptic devices used by many researchers. These are used in manipulating virtual objects, and offer from 2 to 6 degrees of freedom (DOF) of movement. The PHANTOM Desktop™ and PHANTOM Omni™ devices offer affordable desktop solutions. PHANTOM Desktop delivers higher fidelity, stronger forces, and lower friction, while the PHANTOM Omni is an inexpensive cost-effective haptic device. • Phantom toolkit includes several versions of 3-D positioning device and associated software.
Comments: The SensAble Technologies PHANTOM product line of haptic devices makes it possible for users to touch and manipulate virtual objects. Different models in the PHANTOM product line meet the varying needs of both research and commercial customers. The PHANTOM premium models are high-precision instruments and, within the PHANTOM product line, provide the largest workspaces and highest forces, and some offer 6-DOF capabilities. The PHANTOM Desktop device and PHANTOM Omni device offer affordable desktop solutions. Of the two devices, the PHANTOM Desktop delivers higher fidelity, stronger forces, and lower friction, while the PHANTOM Omni is the most cost-effective haptic device available today.		
Immersion, Inc. 801 Fox Lane San Jose, CA 95131 408-467-1900 www.Immersion.com	**Haptic WorkStation, CyberForce, CyberGrasp, CyberTouch, CyberGlove, VirtualHand, HMD, …**	• Manufacturer of many varieties of "Hand-Centric" hardware and software solutions for Force Feedback devices, conveying realistic grounded forces to the hand and arm, and providing 6-DOF positional tracking that accurately measures translation and rotation of the hand in 3-D. • A leading developer and manufacturer of Force Feedback Haptic Systems.
TiNi Alloy Company 1619 Neptune Drive, San Leandro, CA 94577 510-483-9676 www.TiNiAlloy.com	**Displaced Temperature-Sensing System (DTSS)**	• Temperature feedback. • Temp range: 10-45 °C, resolution: 0.1 °C • E-mail communication indicates that the DTSS system is no longer in development. This company is working on an integrated thermal/tactile/pressure sensation package but plans

		no commercial version in the near future.
Cybernet Systems Corp 727 Airport Boulevard Ann Arbor, MI 48108 734-668-2567 www.cybernet.com	**Wearble OCU** Version	Developed a wearable OCU for man-portable robotic applications. It may be used to guide small, man-portable robots for tactical missions, such as reconnaissance. The robot relays video information back to the operator, and displays other status information such as direction, velocity, tilt,...and accepts commands from the operator to control the robot movements.

Comments:
Cybernet Systems Corporation is developing a wearable OCU for man-portable robotic applications. Typically, the OCU will be used to guide small, man-portable robots for tactical missions, such as reconnaissance in enclosed spaces like sewer tunnels. In military parlance, it is a “first man in” situation, where the robot relays video information back to the operator. The OCU also controls movement of the robot. Since this is a tactical situation, the OCU must allow the Soldier to be free to carry out other duties without undo hindrance. It is especially important that the Soldier remain free to perform battle tasks while operating the control system. The OCU displays video information and other status information (direction of travel, velocity, tilt angle, etc). It also accepts control commands from the Soldier and transmits the control data to the robot, commanding the robot to move left, right, or forward.

INTENTIONALLY LEFT BLANK

Appendix D. Glossary and Recommended Readings for the Multi-modal Display Technologies

1. *Speech Synthesis & Recognition Technologies.*

Glossary

diphone	A phoneme modified by the succeeding phoneme.
formant frequency	A distinguishing or meaningful frequency component of human speech.
grapheme	A grapheme designates a unit in written language.
phoneme	A basic, theoretical unit of sound.
pitch	Property of a musical tone measured by its frequency.
prosody	The intonation, stress pattern, and rhythm of speech.
tri-phone	A phoneme modified by the previous and succeeding phonemes.
voice authentication	A biometric used to verify a person's identity.

Recommended Readings:

Keller, E. (Ed.) (1994). Fundamentals of speech synthesis and speech recognition: Basic concepts, state of the art, and future challenges. Chichester: Wiley.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis.* Dordrecht: Kluwer.

2. Tactile Display Technologies.

Glossary

Actuator	Usually mechanical (hydraulic) or electric means used to provide force or tactile feedback to a user
Effectors	Interfacing devices used in VE for input/output, tactic sensation and tracking. Examples are gloves, HMD, headphones, and trackers
Force Feedback	An output device that transmits pressure, force or vibrations to provide the VR participant with the sense of resisting force, typically to weight or inertia. This is in contrast to tactile feedback, which simulates sensation applied to the skin
Haptic Interfaces	Use of physical sensors to provide users with a sense of touch at the skin level, and force feedback information from muscles and joints
Kinesthesis/Kinaesthesis	Sensations derived from muscles, tendons and joints and stimulated by movement and tension
Proprioception	The ability to sense the position and location and orientation and movement of the body and its parts
Tactile Displays	Devices that provide tactile and kinesthetic sensations
Tactor	A tactile output device

Recommended Readings:

Boff, K., Kaufman, L., & Thomas, J. (1986). *Handbook of perception and human performance*, Vol. 1, Sensory Processes and Perception. New York: Wiley.

Schiffman, H.R. (2000). *Sensation and perception.* New York: Wiley.

INTENTIONALLY LEFT BLANK

Glossary of Acronyms

3-D	three-dimensional
AAAI	American Association of Artificial Intelligence
API	application programming interface
ARL	Army Research Laboratory
ARV	armed reconnaissance vehicle
ASCII	American standard code for information interchange
ASR	automatic speech recognition
ATO	Army Technology Objective
BCT	brigade combat team
C2	command and control
C2V	C2 vehicle
CMU	Carnegie Mellon University
COTS	commercial off-the-shelf
CRASAR	Center for Robotic Assisted Search and Rescue
CVC	combat vehicle crewman
DARPA	Defense Advanced Research Projects Agency
DoD	Department of Defense
DOF	degrees of freedom
DSTL	Defense Science and Technology Laboratory
DTSS	Displaced Temperature Sensing System
FCS	Future Combat System
FOV	field of view
GPS	global positioning system
GRV	gravity-referenced view
HMD	head-mounted display
HMMWV	high mobility multipurpose wheeled vehicle
HRI	human-robot interaction
HRTF	head-related transfer function
IC	integrated circuit

ICV	infantry carrier vehicle
LPC	linear predictive coding
MBROLA	multi-band resynthesis overlap add
MELP	mixed excitation linear prediction
MIT	Massachusetts Institute of Technology
MOUT	military operations on urban terrain
NASA	National Aeronautics and Space Administration
NVG	night vision goggle
OCU	operator control unit
PDA	personal digital assistant
POF	probabilistic optimal filtering
PSOLA	pitch synchronous overlap add method
PTT	push to talk
RCTA	Robotics Collaborative Technology Alliances
SA	situational awareness
SBIR	Small Business Innovation Research
SD	stereoscopic displays
SDK	software development toolkit
SPL	sound pressure level
SRI	Stanford Research Institute
TACTICS	Tactile Communication System
TCTS	Circuit Theory and Signal Processing
TNO	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek
TSAS	Tactile Situational Awareness System
TTP	tactics, techniques, and procedures
TTS	text to speech
UAV	unmanned aerial vehicle
UCD	unmanned combat demonstration
UCF	University of Central Florida
UGV	unmanned ground vehicle
U.K.	United Kingdom
USAR	urban search and rescue

UV	unmanned vehicle
VE	virtual environments
VSL	video sign language
WTCU	wireless tactile control unit

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	US ARMY RSRCH DEV & ENGRG CMD SYSTEMS OF SYSTEMS INTEGRATION AMSRD SS T 6000 6TH ST STE 100 FORT BELVOIR VA 22060-5608
1	INST FOR ADVNCD TCHNLGY THE UNIV OF TEXAS AT AUSTIN 3925 W BRAKER LN STE 400 AUSTIN TX 78759-5316
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC IMS 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
2	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK T 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR M DR M STRUB 6359 WALKER LANE SUITE 100 ALEXANDRIA VA 22310
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MA J MARTIN MYER CENTER RM 2D311 FT MONMOUTH NJ 07703-5630
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MC A DAVISON 320 MANSCEN LOOP STE 166 FT LEONARD WOOD MO 65473-8929

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MD T COOK BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1	COMMANDANT USAADASCH ATTN ATSA CD ATTN AMSRD ARL HR ME MS A MARES 5800 CARTER RD FT BLISS TX 79916-3802
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MI J MINNINGER BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MM DR V RICE BLDG 4011 RM 217 1750 GREELEY RD FT SAM HOUSTON TX 78234-5094
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MG R SPINE BUILDING 333 PICATINNY ARSENAL NJ 07806-5000
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MH C BURNS BLDG 1002 ROOM 117 1ST CAVALRY REGIMENT RD FT KNOX KY 40121
1	ARMY RSCH LABORATORY - HRED AVNC FIELD ELEMENT ATTN AMSRD ARL HR MJ D DURBIN BLDG 4506 (DCD) RM 107 FT RUCKER AL 36362-5000
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MK MR J REINHART 10125 KINGMAN RD FT BELVOIR VA 22060-5828
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MV HQ USAOTC S MIDDLEBROOKS 91012 STATION AVE ROOM 111 FT HOOD TX 76544-5073
5	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MY M BARNES 2520 HEALY AVE STE 1172 BLDG 51005 FT HUACHUCA AZ 85613-7069

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MP D UNGVARSKY BATTLE CMD BATTLE LAB 415 SHERMAN AVE UNIT 3 FT LEAVENWORTH KS 66027-2326
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR M DR B KNAPP ARMY G1 MANPRINT DAPE MR 300 ARMY PENTAGON ROOM 2C489 WASHINGTON DC 20310-0300
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MJK MS D BARNETTE JFCOM JOINT EXPERIMENTATION J9 JOINT FUTURES LAB 115 LAKEVIEW PARKWAY SUITE B SUFFOLK VA 23435
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MQ M R FLETCHER US ARMY SBCCOM NATICK SOLDIER CTR AMSRD NSC SS E BLDG 3 RM 341 NATICK MA 01760-5020
10	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MT DR J CHEN 12350 RESEARCH PARKWAY ORLANDO FL 32826-3276
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MS MR C MANASCO SIGNAL TOWERS RM 303A FORT GORDON GA 30905-5233
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MU M SINGAPORE 6501 E 11 MILE RD MAIL STOP 284 BLDG 200A 2ND FL RM 2104 WARREN MI 48397-5000
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MF MR C HERNANDEZ BLDG 3040 RM 220 FORT SILL OK 73503-5600
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MW E REDDEN BLDG 4 ROOM 332 FT BENNING GA 31905-5400

NO. OF COPIES	ORGANIZATION
1	ARMY RSCH LABORATORY - HRED ATTN AMSRD ARL HR MN R SPENCER DCSFDI HF HQ USASOC BLDG E2929 FORT BRAGG NC 28310-5000
1	ARL-HRED LIAISON PHYSICAL SCIENCES LAB PO BOX 30002 LAS CRUCES NM 88003-8002
1	DIRECTOR UNIT OF ACTION MANEUVER BATTLE LAB ATTN ATZK UA BLDG 1101 FORT KNOX KY 40121
1	DIR FOR PERS TECHNOLOGIES DPY CHIEF OF STAFF PERS 300 ARMY PENTAGON 2C733 WASHINGTON DC 20310-0300
1	CODE 1142PS OFC OF NAVAL RSCH 800 N QUINCY STREET ARLINGTON VA 22217-5000
1	CDR USA AEROMEDICAL RSCH LAB ATTN LIBRARY FORT RUCKER AL 36362-5292
1	US ARMY NATICK RD&E CTR ATTN STRNC YBA NATICK MA 01760-5020
1	PEO STRI 12350 RSCH PARKWAY ORLANDO FL 32826-3276
1	GOVT PUBLICATIONS LIB 409 WILSON M UNIVERSITY OF MINNESOTA MINNEAPOLIS MN 55455
1	HUMAN FACTORS ENG PROGRAM DEPT OF BIOMEDICAL ENGNG COLLEGE OF ENGINEERING & COMPUTER SCIENCE WRIGHT STATE UNIVERSITY DAYTON OH 45435

NO. OF
<u>COPIES</u> <u>ORGANIZATION</u>

NO. OF COPIES	ORGANIZATION
1	DIR AMC-FIELD ASSIST IN SCIENCE & TECHNOLOGY ATTN AMC-FAST FT BELVOIR VA 22060-5606
	<u>ABERDEEN PROVING GROUND</u>
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL CI OK (TECH LIB) BLDG 4600
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL CI OK S FOPPIANO BLDG 459
1	DIRECTOR US ARMY RSCH LABORATORY ATTN AMSRD ARL HR MR F PARAGALLO BLDG 459