Human Performance Issues and User Interface Design for Teleoperated Robots

Jessie Y. C. Chen, Ellen C. Haas, and Michael J. Barnes

***Abstract*—In the future, it will become more common for humans to team up with robotic systems to perform tasks that humans cannot realistically accomplish alone. Even for autonomous and semiautonomous systems, teleoperation will be an important default mode. However, teleoperation can be a challenging task because the operator is remotely located. As a result, the operator's situation awareness of the remote environment can be compromised and the mission effectiveness can suffer. This paper presents a detailed examination of more than 150 papers covering human performance issues and suggested mitigation solutions. The paper summarizes the performance decrements caused by video images bandwidth, time lags, frame rates, lack of proprioception, frame of reference, two-dimensional views, attention switches, and motion effects. Suggested solutions and their limitations include stereoscopic displays, synthetic overlay, multimodal interfaces, and various predicative and decision support systems.**

***Index Terms*—Human–robot interaction, multimodal displays and input controls, robotic control, teleoperation, user interface design.**

I. Introduction

TELEOPERATED robots have been used in a variety of situations, ranging from extraplanetary exploration (e.g., the National Aeronautics and space Administrations (NASA) Mars Rovers), military missions (e.g., surveillance/reconnaissance or detecting/removing hazardous materials), undersea operations, search and rescue activities (e.g., searching for survivors at the World Trade Center (WTC) after September 11, 2001), to robotic surgery [1]–[4]. Robots can be teleoperated through a wide variety of control media, ranging from hand-held devices such as personal digital assistant (PDA) systems [5] and cellular phones [6] to multiple-panel displays with control devices such as joysticks, wheels, and pedals [7] (see [8] for a brief overview of vehicle teleoperation interfaces). Typical control stations include panels displaying: 1) sensor view and/or data transmitted from the robots; 2) plans and commands issued to the robots; 3) health status of the robots; 4) status of the tasks; and 5) map displays to maintain the operator's situation awareness and to facilitate navigation. Smaller user interfaces such as the PDA, on the other hand, frequently employ touch-based interactions (e.g., stylus or finger) and multimodal systems such as natural language and visual gesturing [5], [9]. The sizes of the unmanned vehicles range from just a few inches in dimension to multiton vehicles [10]. Robotic systems of various sizes and capabilities will play an integral role in the U.S. Army's Future Combat Systems (FCS) [11]. The scope of proposed robotic systems will require the human controller to operate in diverse environments, with diverse systems under difficult and, in some cases, stressful situations.

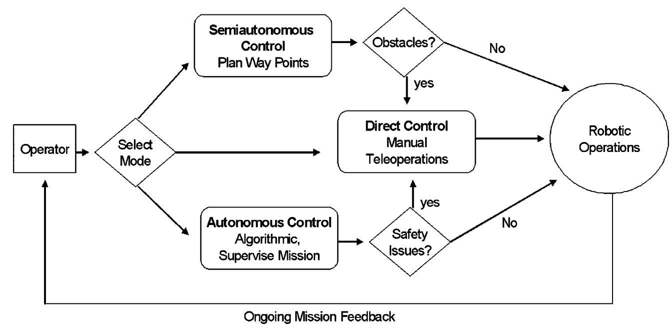

Fig. 1. Control options with decision feedback loops for teleoperations based on safety and obstacle criteria.

While many of the FCS robotic systems will be semiautonomous, teleoperations will be an essential element of these systems for the foreseeable future (either by design or when semiautonomous robots need human intervention). For example, teleoperations will be necessary when semiautonomous systems encounter particularly difficult terrain including natural or manmade obstacles [12]. Even for systems that purport to be fully autonomous, teleoperations will be a default mode for robotic systems that have significant military roles or safety issues. There is even the possibility of using adaptive automation algorithms that return control of robotic systems to the human operator under specified environmental conditions such as a requirement for a tactical maneuver or safe operations [13]. Fig. 1 illustrates varied control options including feedback decision loops for teleoperations under specified conditions indicating that the operator will be in the loop for the foreseeable future, either as a planner, a controller, or a supervisor. This paper examines human performance issues related to teleoperation, especially focusing on remote perception and navigation, and also surveys potential user interface solutions to enhance teleoperator performance.

Human performance issues involved in teleoperating unmanned systems generally fall in two categories, i.e., remote perception and remote manipulation, which includes both navigation and manipulation tasks (e.g., grasping, pushing, and payload management) [14]. Teleoperation tends to

Manuscript received November 2, 2005; revised June 7, 2006 and October 2, 2006. This work was supported by the Army Technology Objective: Robotics-Collaboration of the U.S. Army. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. This paper was recommended by Associate Editor P. Sanz.

The authors are with the U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD 21005–5425 USA (e-mail: jessie.chen@us.army.mil; ehaas@arl.army.mil; michael.barnes@hua.army.mil).

Digital Object Identifier 10.1109/TSMCC.2007.905819

be challenging because operator performance is "limited by the operator's motor skills and his ability to maintain situation awareness. . .difficulty building mental models of remote environments. . . distance estimation and obstacle detection can also be difficult" [15, p. 699]. In the teleoperating environments, human perception is often compromised because the natural perceptual processing is decoupled from the physical environment. This decoupling affects people's perception of affordances in the remote scene and often creates problems in remote perception such as scale ambiguity [16]. Simple tasks could be challenging due to the lack of motion feedback in remote visual processing as well as the mismatching of viewpoints, which could result from placing the camera at a height that does not match the normal eye height [17].

Poor perception has a detrimental effect on situation awareness and, therefore, teleoperating tasks. For remote navigation, the operators often need to estimate the absolute sizes of objects or terrain characteristics so that they can decide whether it is safe for the robot to maneuver in the remote environment (e.g., without getting stuck in a depression) [17], [18]. Expert operators of bomb disposal devices complained that the monochrome and monoscopic video that they had to use made their telemanipulation tasks very difficult, especially when "dealing with small objects outdoors, or in bright sunshine and shadow conditions" [9, p. 19]. People's spatial orientation and object identification in the remote environment also tends to be degraded [20]. For example, studies on rescue robots (e.g., robots for search and rescue at the site of the WTC after September 11, 2001) showed 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 [21]. In real world operations, operator performance is sometimes degraded even further due to robotic system failures. According to Carlson and Murphy [22], generally, reliability of the unmanned ground vehicle (UGV) performance in the field tends to be low (i.e., between 6 and 20 h between failures). Among the causes cited for low reliability, limited bandwidth (which can exacerbate human performance issues associated with teleoperation) was especially an issue for the military due to the military rules on allowed frequencies.

The following section discusses in detail how remote perception and manipulation is affected by factors such as limited field of view (FOV), orientation, camera viewpoint, depth perception, degraded video image, time delay, and motion. Innovative techniques and technologies designed to enhance operator performance and ameliorate those potential performance degradations are also reviewed (summaries are presented in Table I). Particularly, multimodal displays and input controls have great potential in improving teleoperation performance and will be discussed in detail in Section III.

II. Factors That Affect Remote Perception and Manipulation and User Interface Solutions

A. Limited FOV

The use of cameras to capture the environment in which the robot is navigating sometimes creates the so-called "keyhole" effect [16]. 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 situation awareness comparable to direct viewing. In real world operations, such as the WTC rescue effort, the operators often have to rely on the video from the robot eye view to diagnose problems encountered by the robot when automatic proprioception information is not available [2]. For example, in the WTC 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 due to the limited views.

A restricted 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 [20]. Thomas and Wickens [23] demonstrated that operators tended to show "cognitive tunneling" when viewing the remote environment using an immersive display (such as the ones typically used for ground robots) compared to displays with exocentric frame of reference (FOR) [similar to views from an unmanned aerial vehicle (UAV)], which had a greater FOV. Furthermore, important distance cues may be lost and depth perception may be degraded when FOV is restricted [24].

Related research in driving performance with restricted FOV shows that the effectiveness of remote driving can be compromised because of the limited view. With a reduced FOV, drivers have more difficulty in judging the speed of the vehicle, time-to-collision, perception of objects, locations of obstacles, and the start of a sharp curve [25]. Peripheral vision is important for lane keeping and lateral control, and drivers, when driving on curved roads, rely on the "tangent point" on the inside of the curve [25], [26]. A restricted FOV might hinder the turning task since this tangent point has to be determined 1–2 s before the bend. Drivers with a limited FOV often initiate their control actions earlier than optimal [25]. Oving and Van Erp [27] 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, both Oving and Van Erp [27] and Smyth *et al*. [28] showed that the HMD might induce greater motion sickness compared to other viewing conditions.

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 [29]. Wider FOV can be achieved by employing multiple cameras or by using wide-angle cameras. Voshell *et al*. [30] developed a multicamera display *Perspective Folding* to increase the FOV for the teleoperator. Perspective Folding uses five cameras (three horizontal, one pointing 45° upward, and one pointing 45° downward) and provides a "wrap around" effect. However, with increasing FOV, especially when using wide-angle cameras, the speed of travel tends to be perceived as increased because of the scene compression, and drivers usually respond by reducing

TABLE I
SUMMARY OF THE FINDINGS

Factor	Effects	Suggested solution	Ref.#
Field of View (FOV)	Erroneous speed & distance judgments; peripheral vision loss; degraded remote driving	· Increase FOV (e.g., Perspective Folding); changeable FOV can be considered · *Caveats:* perceived speed increases & motion sickness	[25], [29]-[31]
Orientation & Attitude of the Robot	Orientation in the environment; North-up vs. Track-up map; mismatch between actual & perceived attitude of robot; unawareness of (polymorphic) robot's inclination & shape	· Map - Track-up map for navigation; North-up map for tasks involving integration of spatial relations in the environment (e.g., recon, surveying, planning tasks). · Gravity referenced view (GRV) - better awareness of robot's attitude, better route selection, and faster completion of route. · Polymorphic views - operator less likely to tip the robot or have it caught on objects.	[32]-[39], [44], [46]
Multiple Cameras	Attention switching; change blindness; perceptual registration	Auditory alerts; multimodal solutions & visual momentum techniques	[15], [23], [48]-[54]
Camera Viewpoint & Frame of Reference (FOR)	Egocentric - cognitive tunneling; exocentric - loss of immediacy & true ground view; integration of info from different FORs may be challenging for operator; saliency effect	Dual mode & inserts of other views (e.g., Sensory Ego-Sphere); peripheral cues for egocentric	[23], [51]-[53], [55], [71]
Depth Perception	Underestimation of distance & size; degraded navigation, driving, & telemanipulation	· Stereoscopic Displays (SDs) - improved depth perception, obstacle avoidance, arm manipulation, important for difficult terrain & remote arm manipulation; inter-camera distance should be less than inter-ocular distance. · *Caveats:* limited use - benefits mainly for difficult tasks; may induce motion sickness & perceived stress; hyper-stereo SD have multiple negative effects	[18], [19], [29], [61]-[64]
Video Image/ Frame Rate	Degraded motion perception & spatial orientation; degraded target identification & latency	· Min frame rate: 10HZ · Augmented reality/ Synthetic overlay · SDs (see above)	[69], [72], [75]-[79]
Time Delays	Task dependent: negative effects range from 170 ms to over 1 sec; degraded driving, tracking, & telemanipulation; over-actuation when delay is variable; robot-to-operator delay more detrimental than the other direction; motion sickness; degraded telepresence	· Minimum -170 ms for driving like tasks; other minimums task dependent · Predictive displays (e.g., Ecological Display) - navigation faster & more accurate. · *Caveats:* disturbances in remote environment may make prediction model unreliable.	[84], [86], [88], [91], [99], [100]
Motion	Degradation on accuracy & latency; sometimes severe motion sickness	Multimodal user interfaces; tailor interface to vibratory & motion effects, possible medical remediation	[7], [76], [101]-[104]

their speed [31]. In addition, Smyth *et al.* [28] suggests that the decreased resolution and increased scene distortion associated with scene compression may increase cognitive workload for driving and object localization tasks, 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.* 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 peripheral cues for direct viewing. They concluded that for indirect vision driving, optimal performance might be achieved by employing unity vision display with the capability to electronically change FOV.

B. Orientation

In order to successfully navigate in the remote environment, the robotic operator needs to have a good sense of orientation, both globally and locally. Globally, the operator needs to know where the areas of interest are relative to the location of the robot; locally, the operator needs to negotiate local turns and avoid obstacles in order to navigate to the robot's destinations. To successfully navigate locally, the operator also needs to be aware of the robot's attitude (i.e., pitch and roll) in order to avoid roll-over accidents. This section will discuss issues, and design solutions in the areas of orientation in the remote environment as well as awareness of robot's attitude.

1) Orientation in the Remote Environment: Navigation with a traditional (north-up) map can be challenging at times because of the demand of mental rotation. Studies comparing human performance using north-up (world-referenced; fixed viewpoint) versus track-up (ego-referenced; rotating viewpoints) maps consistently show that track-up maps are better for local guidance (i.e., navigation) and north-up maps are better for global awareness [32]–[37]. User interface design guidelines generally recommend making both north-up and track-up maps available [38]. It is also recommended that when in a route planning mode, the default should be north-up; during navigation, the default should be track-up. There are two general types of track-up maps: with moving icon or with icon remaining centered and stationary. So far, there has not been convincing evidence as to which one is superior [35].

Casner [33] examined pilots' navigation performance using a paper (north-up) map vesus a moving (track-up) map and found that navigation accuracy was significantly better for the track-up group. However, when the same groups of pilots were asked to complete the flight again without the maps, the track-up group performed significantly worse this time. A follow-on study [39] demonstrated that the loss of navigational awareness by the track-up group was due to the pilots' passive role in navigation when using the track-up map. A greater involvement in the navigation task for the track-up group (e.g., pointing out geographical features) was introduced in the follow-on study, and the performance was at about the same level as the north-up group. The author cautioned that the pointing technique had its obvious limitation in operational environment, and other more practical techniques need to be developed. However, the Casner studies did demonstrate the vulnerability of using track-up displays.

2) Attitude of the Robot: 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 hard to accurately assess the attitude of their robotic vehicles [40]. In fact, misperception of attitude was cited as the only problem in an egocentric teleoperation accident at Sandia [41]. Essentially, the operators were not aware that their robotic vehicles were on a grade until they rolled over. Other near roll-over incidents have also been reported and it was determined that insufficient awareness of the attitude of the teleoperated vehicle has caused the incidents [42]. In the WTC search and rescue efforts, the operators also had similar problems and were not aware of the orientation of the surface until their robots flipped or rolled (Murphy, as cited in [43]).

Lewis *et al.* [44] developed a gravity referenced view (GRV) display 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, however, cautioned that the conditions favoring the use of GRVs may be limited to those involving confusing environments and stressful operations, which are precisely the ones that the military will encounter [45].

Drury *et al.* [46] developed a user interface that showed the physical configuration (i.e., pose) of the robot, which was polymorphic. They found that, by providing the pose information (inclination and shape of the robot), the operators were less likely to tip the robot or have the robot caught on objects in the environment. However, the operator's high workload did not seem to be ameliorated by the "pose display." The authors recommended automated assistance that could change the shape of the robot automatically.

C. Camera Viewpoint (Context)

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 [47]. 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 eyesight of the operator, remote perception (e.g., position estimation) may be degraded due to the unnatural viewing angles for the human [21].

Multiple camera viewpoints are usually employed to enhance remote perception (especially object identification) [2]. Hughes and Lewis [48] 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 [48] suggest a two-screen approach, where one screen is under human control and the other screen is sensor-driven (i.e., sensor would direct the operator to a particular viewpoint of interest). However, it was suggested that the differences between the eye point and the camera viewpoint may induce motion sickness [25]. In addition, when handling multiple robots, it can be challenging for the operator to acquire the different contexts rapidly when switching among the robots [15], [49]. The user has to remember, e.g., the surroundings for each robot and what tasks have been and have not been performed [2]. Moreover, the literature in change blindness suggests that information in one scene may not be encoded sufficiently to be compared/integrated when accessed subsequently [23], [50]. Therefore, some changes may go undetected when viewpoints are changed. It is even more challenging when the robots are heterogeneous with different capabilities.

Future warfare employing the FCS may need to integrate information from multiple platforms, potentially from both 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. The ideal view depends on the task; overall awareness and pattern recognition are optimized by exocentric views whereas the immediate environment is often viewed better egocentrically. For example, using a simulated ground robot with a direct view (egocentric) provided faster and safer performance for a rescue mission scenario in a cluttered urban environment compared to the "God's eye" view from an aerial asset but the soldier participants still wanted the exocentric view for an overall situation [51].

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 [23]. Research has shown that integrating information across egocentric and exocentric views can be challenging for the operator [52], [53]. Essentially, dual displays with both frames of references require effective scanning of the displays and integrating information from two different perspectives to form an accurate assessment of the situation. Furthermore, operators may be susceptible to saliency effect and anchoring heuristic/bias [23]. In other words, 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 are not properly attended to and integrated. In Thomas and Wickens [23], participants were found to tunnel their attention into the egocentric view to the exclusion of information from the exocentric view. Olmos *et al.* [52] recommended several user interface designs to address these cognitive tunneling issues. By incorporating auditory alerts and techniques such as visual momentum (which was originally used in film editing to produce smooth transitions between different views of the same environments [54]), Olmos *et al.* [52] demonstrated that the problems of inadequate distribution of attention to separate displays were addressed and that the UAV operators' teleoperation performance was improved. Johnson *et al.* [55] demonstrated the use of a sensory ego–sphere interface that displayed multiple sensor data from both frames of references. Although statistically significant results were not found, teleoperators' mental workload was lower while their situation awareness (SA) was higher using this system when compared with other traditional interfaces.

D. Degraded Depth Perception

The use of monocular cameras and its effects on teleoperator's depth perception have been investigated in various contexts. Basically, projecting 3-D depth information onto 2-D display surface results in compressed or "foreshortened" depth perception [23]. The compression is worse with the ground robots because of their low viewpoints than with the aerial robotic vehicles. Degraded depth perception affects the teleoperator's estimates of distance and size and can have profound effects on mission effectiveness [18]. It is well documented that humans underestimate distances more in virtual environments (VEs) than in the real world [56], [57]. Both Lampton *et al.* [56] and Witmer and Kline [57] found that, while distance underestimation happened in both VE and real-world environment, the underestimates in the VE were more extreme. In field-testing environments, such as in McGovern's experiment [41], operators of teleoperated UGVs also consistently underestimated the distances from obstacles and landmarks.

In typical teleoperation environments, 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 the depth of the remote scene [47]. In unfamiliar or difficult terrain, such as the rubble pile at the WTC scene where objects are disorganized and deconstructed, depth perception is extremely challenging due to lack of apparent size cues [21]. In a usability test of a mixed-initiative robotic system, Marble *et al.* [58, p. 451] observed that "most participants indicated a desire for the interface to overlay the video with a depth indicator, especially in teleoperated mode."

Stereoscopic displays (SDs), which rely on various techniques to present a binocular image to the user, appear 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) [19], [29]. According to Dumbreck *et al.* (as cited in [19]), 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, however, generally report that SDs might be useful in only certain circumstances. For example, Drascic [19] found that the benefits of SDs, while longer lasting for tasks that required binocular depth cues (i.e., using a robot to place an object between two "bombs" separated by 8 cm), did not last as long for tasks that did not require much binocular depth perception (i.e., same task with "bombs" separated by 64 cm). Generally, participants quickly learned how to use the monocular cues available in the monocular displays to accomplish those tasks.

Draper *et al.* [59] had his participants perform Fitts' Law tapping tasks (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 [60]) and reported that SDs were only useful for more difficult tasks and only for inexperienced participants. He suggested that SDs would be useful when the image quality, task structure and predictability, user experience, and manipulator dexterity were suboptimal. Rosenberg [61] 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 intercamera distance was less than the interocular distance (i.e., 2–3 cm versus 6 cm). Green *et al.* [62], on the other hand, did not find significant benefits of using SDs for their tasks (teleoperating shipboard cranes to place cargo) in terms of time and accuracy of task performance and depth perception. As for user preference, a consistent finding from various studies is that teleoperators generally prefer SDs over monocular displays [62], [63]. However, as noted in Scribner and Gombash [29], artificially induced binocular stereovision may increase motion sickness and perceived stress.

More realistic tests were conducted by researchers from the U.S. Army who investigated the ability of humans to detect obstacles in static and moving video terrain with 3-D and hyperstereo (ocular distance artificially increased to accentuate depth cues) displays. The results indicated improved detection of negative terrain and mobility obstacles for the 3-D conditions versus 2-D conditions [64], [65]. The hyperstereo conditions did not improve the performance possibly because such images distort the scene, causing objects to appear smaller and closer, to give conflicting convergence and accommodation cues, and appear to be out of the frame for some frontal scene objects. The researchers attempted unsuccessfully to train the operators to use 2-D cues to replicate their 3-D performance, suggesting that the 3-D gains were robust for navigating in complex terrain for both night (night vision goggles) and day conditions [64]. Preliminary field demonstrations of actual Army systems suggest that the 3-D performance gains will extend to teleoperated systems. The researchers caution that SDs have definite perceptual and physical limitations (as already mentioned); however, SD should be an optional mode for complex terrain (especially where depth perception is crucial) and for arm manipulations and other tasks where normal 3-D cues are unavailable [65].

E. 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 [66]. 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 [25]. Teleoperation is often prone to poor spatial awareness of the remote environment due to the impoverished representations from video feeds, which could leave out essential cues for building teleoperator's mental models of the environment [17], [67].

Common forms of video degradation caused by low bandwidth include: reduced frame rate (FR) (frames per second or fps), reduced resolution of the display (pixels per frame), and a lower grayscale (number of levels of brightness or bits per frame) [47]. The product of FR, resolution, and grayscale is bandwidth (bits per second), and it is important to determine how to tradeoff these three variables with a given bandwidth so that the operators' performance can be optimized [67]. Generally, for applications in VE, it is recommended that 10 Hz be the minimum FR to avoid performance degradation [69]. Darken *et al.* [20] demonstrated that people had difficulty maintaining spatial orientation in a remote environment with a reduced bandwidth. Participants were asked to view streaming videos with FRs degraded to different degrees (lowest FR being 1.43 fps), and their task was to track the position of the traveling camera on a given floor plan as well as look for targets along the way. The participants had great difficulty in identifying objects in the remote environment and had significantly more errors in reporting the camera's locations, especially in the extremely slow FR condition (i.e., 1.43 fps).

Van Erp and Padmos [25] suggest that speed and motion perception may be degraded if image update rate is below 10 Hz. Massimino and Sheridan [70] demonstrated that teleoperation ("peg-in-hole" placement task) was significantly affected with a 5–6 fps rate and became almost impossible to perform when the FR dropped below 3 fps. Chen *et al.* [71] found that with a 5 Hz FR, participants' target acquisition performance was significantly degraded. French *et al.* [66] showed that reduced FRs (e.g., 2 or 4 fps) affected the teleoperator's performance in navigation duration (time to complete the navigation course) and perceived workload. However, no significant differences were found among different FRs (i.e., 2, 4, 8, and 16 fps) for navigation error, target identification, and situation awareness. The authors recommended that no less than 8 fps should be employed for teleoperating UGVs. It appears that increasing the FR to higher than 8 fps might not greatly enhance the teleoperation performance [41]. McGovern [41] evaluated operators' performance in teleoperating a small UGV, and found that reducing the FR from 30 to 7.5 fps did not significantly affect the number of obstacles hit by the operators. Spatial resolution was found to be a more determining factor for obstacle avoidance than did temporal resolution.

Thropp and Chen [72] reviewed more than 50 studies and summarized them in the areas of psychomotor performance, perceptual performance, behavioral effects, and subjective perception. They found that, generally, psychomotor performance improves at higher FRs and lower standard deviations of FR. Experimental results have suggested a minimum threshold of 17.5 Hz for successful placement performance [73]. Tracing performance may also require more than 10 Hz. Higher FRs such as 16 Hz are suggested to aid in navigation and target tracking. Overall, there seems to be a strong support for a threshold of around 15 Hz for many tasks, including those that are psychomotor and perceptual in nature. Less impressive, yet acceptable performance may be accomplished at around 10 Hz for many tasks. SDs are likely to enable a performance advantage over monoscopic displays, and when possible, they should be used to assist operators in compensating for lower FR presentations. If displays are stereoscopic, FRs as low as 7 Hz may be adequate, but monoscopic displays may only be useful for FRs as low as 14 Hz [74].

Several researchers have looked into ways to augment video image with other information [9], [75]–[79]. Keskinpala and Adams [9] developed a user interface that integrated the streaming video with sensor data (sonar and laser-range finder data), which was overlayed graphically on the video screen. They compared this display with a video-only interface and a sensor-data-only interface and reported that it took longer for the operators to finish the navigation task using the integrated display. Furthermore, the operators perceived higher workload with the integrated interface. However, the authors noted that the performance was slower due to the screen processing delay rather than the use of integrated display. Another study by Nielsen and Goodrich [79] provided more convincing evidence of the usefulness of integrated displays. They compared an integrated representation of video and (3-D perspective) map with a side-by-side representation, (video + 2-D map) and found that the video and map information tended to compete for the operator's attention when placed side by side; when presented in an integrated fashion, they tended to complement each other and resulted in an improved overall performance. Calhoun and Draper [76] recommended the use of a synthetic vision overlay to augment a degraded video feed. More specifically, the overlay can include the following information: maps and other synthetically generated symbology, photo-imagery, terrain elevation, and updates via networked communication with other sources. Collett and MacDonald [77] also developed an augmented reality-based visualization system, which overlayed onto video feed such critical navigation-related information as laser range scans, past robot path, potential future robot paths, and other vital statistical data. Nevertheless, neither Calhoun and Draper [76] nor Collet and MacDonald [77] reported any human-in-the-loop experimental data to demonstrate the usability of their systems. Additionally, according to Tufano's [80] and Wickens's [81] work on head-mounted displays, overlaying information on video feed can potentially lead to cognitive tunneling, as the operator's attention can be captured by the overlaid data while important elements/developments in the video might be overlooked. The tradeoff between adding information to the video and cognitive tunneling needs to be more systematically evaluated.

F. 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 transmitting information across a communications network [82]. Studies on human performance in VE show that people are generally able to detect latency as low as 10–20 ms [83]. Generally, when system latency is over about 1 s, the operators begin to switch their control strategy to a "move and wait" one, instead of continuously commanding and trying to compensate for the delay [84]. Sheridan and Ferrell [85] 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 [86] recommended that supervisory control and predictor displays be used to ameliorate the negative impact of time delays on teleoperation.

Several researchers have been investigating the human performance degradation in interactive systems caused by time delays less than a second (compared to several seconds in the Sheridan and Ferrell study [85]). In a study of target acquisition using the classic Fitts' law paradigm, MacKenzie and Ware [82] demonstrated that movement times increased by 64% and error rates increased by 214% when latency was increased from 8.3 to 225 ms. A model of modified Fitts' law (with latency and difficulty having a multiplicative relationship) was proposed based on the experimental results. Other studies found that latencies as short as 300–320 ms would significantly affect operator's compensatory pursuit tracking performance or even make the teleoperator decouple his or her commands from the robotic system's response [84], [87]. In a simulated driving task, the driver's vehicle control was found to be significantly degraded with a latency of only 170 ms [88].

On the other hand, some studies did not find performance degradations with latencies under 1 s. For example, Lane *et al.* [84] did not find any performance degradation in a 3-D tracking task (i.e., simulated free flight control) until the latencies were 1.5–3 s, although the authors also reported that it did take the participants significantly longer to complete the more difficult position (i.e., extraction and insertion) task when the latency was over 500 ms. Depending on the tasks, the effects of latencies can vary widely. For example, in Lane *et al.* [84], the same 3 s delay caused only 60% more completion time in a robotic manipulator maintenance task (i.e., extraction of a box) but a 213% completion time increase in a simulated "peg-in-hole" positioning task.

FR and latency are closely related and are sometimes manipulated at the same time [89], [90]. Ellis *et al.* [90] found that latency (ranging from 80 to 480 ms) was a more reliable and stronger negative influence than FR (20, 12, or 6 fps) on 3-D tracking performance. Similarly, Arthur *et al.* [89] also found latency (ranging from 50 to 550 ms) to be a more important factor than FR (30, 15, or 10 fps) for their participants' 3-D tracing task performance.

It was suggested that a short variable lag could be more detrimental than a longer fixed one [84]. In a study on latency effects on performance of grasp and placement tasks, Watson *et al.* [69] 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). Other researchers have shown that overactuation (e.g., oversteering and repeated command-issuing) is common when system delay is unpredictable [7], [11]. However, Luck *et al.* [91] did not show these differential effects between fixed (1 or 4 s) and variable latencies (variation was 50% in each direction) in their experiment, in which the participants were asked to navigate a small ground robot through a maze-like course. In fact, Luck *et al.* [91] participants made more navigation errors with the constant latencies than with the varied latencies, although they did take more time to complete the course with long variable latencies compared with long fixed latencies. The authors suggested that this discrepancy might be due to participants' increased caution when the latencies were variable. Latency direction was also found to affect navigation errors, with delays between robot-to-operator more detrimental than delays in the other direction.

Time delay has been associated with motion/cyber sickness, which can be caused by cue conflict (i.e., discrepancy between visual and vestibular systems) [92], [93]. For example, in Oving and Van Erp's [27] study on indirect driving of an armored vehicle, several participants in the HMD-driving condition (with a time delay) had to withdraw from the experiment due to motion sickness.

Studies have also shown that high latency lags tend to reduce perceived telepresence [90], [94]. Telepresence is defined as the perception of being present in a remote environment [95], and is usually measured subjectively by questionnaires such as the Presence Questionnaire [96]. Draper *et al.* [97] speculated that user interfaces that facilitate telepresence should improve teleoperator's performance, and Kaber *et al.* [94] subsequently demonstrated a positive link between telepresence and performance (using a telerobot in a simple pick-and-place task through a virtual reality interface). However, the relationship between telepresence and operator performance demonstrated so far has been largely correlational rather than causal. In other words, it is not clear if or how reduced telepresence affects robotic control performance. Woods *et al.* [16] suggested that achieving *functional presence* might be a more realistic goal for teleoperation user interface design. Functional presence occurs when the teleoperator receives sufficient perceptual cues to effectively conduct teleoperations.

A potential solution for time delay is the predictive (or predictor) displays. Predictive displays, using the teleoperator's control inputs, "simulate the kinematics without delay and immediately display graphically the (simulated) system output, usually superposed on the display of delayed video feedback from the actual system output" [86, p. 108]. Some predictive displays employ VE, in which the "phantom robot" reacts to the teleoperator's commands in real time [98]. Various techniques such as augmented reality, visual tracking, and image-based rendering have been used for VE-based predictive displays [47], [99], [100]. 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 performance time by 50%–150% [68]. Ricks *et al.* [100] 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), compared with a standard interface (with only maps, streaming video, and status panels). The ecological display presented spatial range information using 3-D graphics and a tethered perspective (viewpoint being a little above and behind the robot). The participants also preferred the ecological display over the standard interface.

Although somewhat task dependent, in general, time delays degraded performance and, for some tasks, made operators prone to motion sickness. For example, time delays as short as 170 ms affected driving performance. If these delays cannot be engineered out of the system, it is suggested that predictive displays or other decision support be provided to the operator.

G. Motion

As planned for the FCS of the U.S. Army, the operators will sometimes need to control their robotic assets from a moving vehicle (e.g., command and control vehicle). It has been demonstrated that performing computerized tasks or simulated teleoperation tasks on moving platforms is difficult, and can make the operators report motion sickness effects [101], [102]. Vibration of the moving vehicle makes viewing the visual displays and manual control/input more challenging for the operator. The FCSs Lead System Integrator performed a demonstration for the Concept and Technology Development phase, in which the operator teleoperated robotic vehicles from a moving command vehicle [7]. The results showed that motion made all tasks harder, 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 due to the difficulty experienced by the operators in stabilizing their hand movements. The operators also tended to oversteer their robotic vehicles when their own vehicle was turning one way but the robot needed to turn the other way. Besides perceptual and psychomotor aspects, motion also made cognitive tasks more challenging [103]. Schipani *et al.* [103] evaluated soldiers' cognitive performance while in a moving vehicle, and found that they were less accurate (7%–46% degradation) and slower (7%–40% decrease). They found degradations in areas such as time sharing, selective attention, inductive reasoning, memorization, and spatial orientation.

There are several potential user interface design solutions to address the motion issues. For example, it has been proposed to make the displays adaptive to the environment (i.e., vibration) by enlarging the size of the fonts and soft-keys, enhancing the conspicuousness of critical information, canceling vibration effects for joysticks, and replacing visual with audio input/output [104]. Multimodal user display and voice input control interfaces present great potential, and will be discussed in detail in the following sections.

In summary, this section reviewed several factors that can affect teleoperator's performance in remote perception, navigation, and manipulation. These factors are usually investigated separately in experiments, and future research should combine these factors and examine the accumulated effects and interactions among the factors. Research in the field settings suggests that perception and situation awareness is still the major bottleneck in teleoperation missions [105]. According to Murphy and Burke, during missions, it is not unusual for robotics operators to spend half their time on determining what they are looking at instead of teleoperating. The majority of the issues reviewed in this section are related to building teleoperator's situation awareness. More research on these issues and development of user interfaces that can address these issues will greatly benefit robotics users in the field settings.

III. Multimodal Displays and Controls

Teleoperation requires advanced immersive human–computer interaction. Researchers have shown that enhancing teleoperator feedback plays a role in decreasing task difficulty [106] and creating a greater sense of operator immersion in a teleoperation environment [107]. Early robotic teleoperation systems used unimodal visual feedback. As awareness that the use of additional modalities could supplement the visual channel when it was heavily loaded, bimodal feedback systems (visual-auditory, visual-haptic) were developed. Currently, some teleoperation feedback systems use all three modalities. Although additional modalities such as olfactory feedback are available, they are less developed than the others at the present time [108]. This section reviews multimodal user interfaces developed for robotics display and control. A summary of research findings discussed in this section is presented in Table II.

A. Visual and Audio Displays

Until very recently, the majority of past multimodal teleoperation research dealt with combinations of visual and auditory displays [109], at least in part because the advent of digitization made it easy to design and synthesize audio cues. Designers discovered that audio cues were a useful supplement to visual feedback because they can increase awareness of surroundings, cue visual attention, and convey a variety of complex information, especially when the visual channel is heavily loaded [110]. Chong *et al.* [111] examined the use of audio and visual displays for multitelerobot operations in which several robots were controlled by multiple remote operators physically distant from each other. They found that by using audio and visual feedback, operators could detect the possibility of collision and were able to coordinate the conflicting motions between two telerobots. Nagai *et al.* [112] used an auditory feedback system to reduce operator workload during simulated space robot teleoperation. Nagai *et al.* [112] found that audio cues were a powerful tool in helping operators make decisions, and recommended that they would be helpful in preventing accidents during actual space operations.

Audio cues can also be presented spatially, allowing the listener to perceive and localize cues that sound as if they originate outside the head at veridical locations. Spatial audio displays have been shown to increase situation awareness in the operation of unmanned aerial displays [113]. Tachi *et al.* [114] created a humanoid robot with spatial audio and visual displays in order to permit operator control of the robot from a remote location while experiencing control as if he or she were inside the robot itself. They showed that presence lent by spatial audio and visual cues created more intuitive robotic control and facilitated the operator's sense of presence in teleoperation tasks.

B. Visual and Haptic Displays

Haptic displays are recent promising technologies for teleoperation feedback. Haptic displays generate skin-based as well as proprioceptive (body position, orientation, and movement) information. Tactile displays refer to a type of haptic display

TABLE II
MULTIMODAL DISPLAYS AND CONTROLS

Type	Findings	Citations (Ref. #)
Audio Display	Useful supplement to visual feedback - increase awareness of surroundings, cue visual attention, and convey complex info; can reduce operator workload; spatial audio displays can increase situation awareness	[109]-[114]
Tactile Display	Effective cueing mechanism - can be used to provide warning info and communication info regarding orientation and direction as well as user position and velocity; especially useful in noisy environments requiring long periods of vigilance	[115]-[118]
Haptic Input	Can provide continuous, proportional force feedback info; can improve teleoperator performance; ideal for surgery tasks requiring fine manipulation	[119]-[127]
Audio & Haptic Display	Providing different modalities in combination may not be more advantageous than presenting the modalities separately; virtual fixtures can provide guidance against certain directions of motion or forbidden regions and can improve operator accuracy; visual and audio fixtures were found to be more effective than tactile fixtures in terms of operator speed and accuracy	[106], [129]-[135]
Voice Input Controls	Useful when manual input is not effective (e.g., in a moving vehicle) or when both of the operator's hands are busy; multiple commands can be consolidated into single "macro" commands; can reduce operator fatigue during demanding procedures	[104], [136]-[141]
Gesture Input Controls	Easy to use, can be used anywhere in the field of view of a camera, does not require special hardware, and allows a wide variety of gestures since it is software-based; generally oriented to teleoperation tasks that leave the hands of the operator free	[142]-[146]
Voice & Gesture Input Controls	Provide a large range of interactions natural to humans. System interpretation and of multiple-modality input may be difficult; success depends on efficient, effective integration and delivery strategies.	[147]-[151]

that uses pressure or vibration stimulators that interact with the skin [115]. As with auditory displays, haptic displays have been used mainly to supplement visual information without taxing the visual system. Tactile displays have been used to provide warning information and communicate information regarding orientation and direction [116], as well as user position and velocity [117]. Calhoun *et al.* [118] found that tactile displays can significantly improve detection of faults in UAV control tasks, and can serve as an effective cueing mechanism. They suggested that tactile alerts may be advantageous in noisy task environments requiring long periods of vigilance, where both audio and visual channels are taxed.

Haptic displays have been used in specialized telerobotic applications. Sitti and Hashimoto [119] designed a scanning system for nanoscale operations, using a telenanomanipulator with haptic feedback. Kortschack *et al.* [120], who designed a nanohandling microrobot with a mobile platform, manipulator, and end effectors, recommended that haptic feedback is next in importance to visual feedback in the teleoperation of a micromanipulation control station.

Visual-haptic systems have also been employed in space-based teleoperation systems. Aleotti *et al.* [121] used glove actuators to communicate contact between a robotic gripper and a remote object, while vibration intensity was used to return information regarding distance from the object of interest. A wearable haptic glove was used to provide continuous proportional force feedback information for a telemanipulation task by Murray *et al.* [122], who found that the use of correlated amplitude and frequency signals to simulate force substantially improved teleoperator performance (pick-and-place manipulation task and weight-sorting task).

Haptic displays have also been used in simulated surgical applications. Visual and haptic force-feedback was used in the Robot Assistance Micro Surgery system developed by the Joint Physics Laboratory and NASA [123]. Kennedy *et al.* [124] used haptics and vision in robotic cardiac surgery, and Kitagawa *et al.* [125] noted that haptic feedback significantly enhances the execution of cardiothoracic surgery tasks requiring fine suture manipulation and knot-typing tasks. Kragic *et al.* [126] found that haptic force-feedback can be useful in retinal microsurgery, in which minimally invasive surgical tasks require micrometer-scale accuracy. In retinal microsurgery simulation, surgical tools are mounted on a robotic arm end effector that also contains a force/torque sensor. Surgical tools are manipulated by the surgeon applying force to a handle attached to a force sensor, with the robot moving in proportion to the applied force. To provide end effector motion assistance and enhance surgical accuracy, Kragic *et al.* [126] recommended that the use of haptic "virtual fixtures," which provide haptic force feedback to "stiffen" the hand-held guidance mechanism against certain directions of motion or forbidden regions of the workspace. Rosenberg [127] found that user performance can increase as much as 70% with fixture-based guidance. More on virtual fixtures will be presented in the following section.

C. Visual, Audio, and Haptic Displays

Researchers compared different modalities to each other by exploring the use of all three separately or in different combinations. Lathan and Tracey [106] explored remote operator teleoperation with different combinations of visual, auditory, and vibrotactile feedback, and found that providing different modalities alone or in combination did not significantly affect user response time in a telerobotic navigation task. However, they did find a significant interaction between operator spatial ability and feedback condition, where spatial ability was defined by subject performance on a test spatial recognition and manipulation test battery. Gunn *et al.* [128] used multimodal displays to communicate threats in a UAV control task, where threat signals led observers to perform a subsequent manual target acquisition. They found that visual, spatial-audio, and haptic cues used separately enhanced target acquisition performance over no cueing in the target acquisition phase of the task, and did so to a similar level.

Ng *et al.* [129] used forearm-mounted tactors and audio signals to provide information on operating room physiological variables in a simulated operating room. They compared tactile with audio and audio/tactile alarms in a physiological monitoring system, and found that the tactile alarm was as easy to learn and had a higher alarm identification rate than an auditory alarm alone. They also found that tactile display used alone provided greater alarm identification accuracy when compared to a combination tactile and auditory alarm. Massimino and Sheridan [130] examined the capabilities of audio and tactile cues to provide force-feedback information in a space teleoperation system. They found that the use of either of these cues to provide force feedback improved operator performance, and that the auditory and tactile displays compared favorably to traditional force-feedback displays that used force reflection.

Other researchers explored the simultaneous use of multiple modalities. Some of these applications were web-based teleoperation systems in which remote objects such as robotic arms or mobile robots are controlled through an Internet network by means of a browser. Chou and Wang [131] designed a multimodal interface for Internet-based teleoperation, in which live video images, audio, and force information were organized and presented in a predictive display. They found that presenting multimodal information reduced operator mental workload. Elhajj *et al.* [132] used video, audio, and haptic information to perform real-time robotic teleoperation via the Internet. They noted that challenges to multimodal telerobotic operations via the Internet include random communication time delay, which causes instability, loss of transparency, and desynchronization. Asynchronous (delayed visual) feedback can affect localization accuracy of stimuli, completion time of manipulation tasks, and perceived telepresence (see Section II on latency).

Multimodal feedback has also been used in virtual fixtures in teleoperation tasks. In this application, audio, visual, and tactile information are used as perceptual overlays to provide guidance information regarding direction of motion or forbidden regions of the user's workspace [133]. Virtual fixtures have been found to improve operator speed and accuracy, since they can alert the user for changes in the environment and support hand-eye coordination for object manipulation tasks [134]. Aleotti *et al.* [135] compared virtual fixtures of different modalities to denote an insertion acceptability zone for a virtual peg-in-hole insertion tasks with varying levels of difficulty. They found that visual and audio fixtures were more effective than tactile fixtures in terms of operator speed and accuracy. They also found that as task difficulty increased, virtual fixtures reduced error rates but not task performance time because users tended to spend more time trying to achieve successful task execution rather than trying to reduce task completion time. Abbott *et al.* [133] suggested that the level of guidance provided by the virtual fixture is key to task time and accuracy. They found that having a high level of guidance increases performance time and error for tasks that require off-path motions, although it significantly improves both time and error during path-following.

D. Voice Controls

Conventional teleoperation controls consist of keyboard, mouse, and joystick. However, voice input is a viable substitute for conventional controls, especially in environments where manual control is difficult (e.g., in a moving vehicle) or when both of the operator's hands are busy [104]. Draper *et al.* [136] compared the utility of manual and speech input for several UAV control tasks. Participants were asked to manually fly a high-fidelity simulated UAV and concurrently perform a series of data entry tasks using either manual (push button) or voice commands. Results showed that voice input enhanced not only participants' data entry performance (faster and more accurate) but also their performance of the UAV flight/navigation task. Speech commands were advantageous because they required fewer steps to complete; numerous sequential button presses could be replaced by one voice command. The authors noted that assigning more functions to buttons would only result in a slight performance enhancement because the number of functions to be controlled in UAV control stations will remain the same or increase, and that adding additional buttons is not as efficient as replacing several button presses with one voice command.

Voice commands have been used in telerobotic surgical systems. Downs [137] described an integrated remote neurosurgical system in which a remote expert surgeon could control a microscope by means of voice commands. Rininsland [138] described the advanced robotics and telemanipulator system for minimally invasive surgery (ARTEMIS), in which a surgeon uses voice commands to guide an endoscope effector while planning, training, and performing minimally invasive surgical procedures.

Speech displays have also been useful in teleoperation tasks where both of the operator's hands are busy. The ROBTET, a robotic system for maintaining live power lines, allows users to issue voice commands while using both hands to perform other control tasks [139]. Field tests indicated that time to perform teleoperation maintenance tasks such as changing insulator string was similar to the time spent by experienced users without teleoperation. Voice commands are also used with Robonaut, a NASA/DARPA mobile humanoid robot, where the user continuously employs both hands and arms to control the robot's dexterous five-finger hands when performing various space orbital and planetary operation tasks [140]. The authors found that voice commands were extremely important in reducing operator physical fatigue during demanding teleoperation procedures.

Speech input has been used in conjunction with other modalities in teleoperation tasks. Perzanowski *et al.* [141] performed a Wizard-of-Oz study to examine the extent to which speech recognition is used in conjunction with touch input (i.e., a touch screen) to guide a robot in a search task. Perzanowski *et al.* found that participants used the speech input most often, making an average of five times as many utterances as touch inputs only or combined speech and touch inputs in each search task. They found that participants making the most utterances tended to make the fewest touch inputs and vice versa. Most touch inputs were combined with utterances, and when isolated touch inputs were made, they were intended to be corrections of previous touch inputs. Future studies will involve the use of other input modalities, including a joystick.

E. Gesture Controls

Human gesture is another means of controlling robots in teleoperation. Gesture, which includes human body and arm pose, hand gesture, and facial expression, can be static (not moving) or dynamic (where meaning is dependent on gesture motion trajectory). Hu *et al.* [142] noted that the gesture interface is advantageous because it is easy to use, can be used anywhere in the field of view of a camera, does not require special hardware, and allows a wide variety of gestures since it is software-based. Generally, gesture interfaces are oriented to teleoperation tasks that leave the hands of the operator free [143].

In the last few years, several researchers worked in the area of gesture in teleoperation. Cohen *et al.* [144] developed dynamic arm gestures to control a pan-tilt robotic camera. Kuno *et al.* [145] developed a robotic wheelchair that includes gesture control, moving in the direction indicated by the user's face movements. Frigola *et al.* [143] described a vision system interface that tracks human gestures for relatively simple robot commands such as up-down, stop, turn, approach, and go. Adrizzone *et al.* [146] note that several methodological approaches to gesture interpretation have been proposed over the years, including camera tracking of the whole body or of the arm, using 3-D blob features to build a model of the human body, using a passive posture estimation method based on multiple camera input, and using motion, shape, and color analysis to localize facial expressions.

Hu *et al.* [142] noted that good gesture design should possess several important features. Gestures should be natural, consistent, and easy to demonstrate. The background against which the user stands should be simple but effective, to allow the camera to easily identify the gestures. In addition, image acquisition and processing time should be short so that the gestures can be identified in real time.

F. Voice and Gesture Controls

Several researchers recommended combining gesture with speech interfaces because they are a natural means of providing a full range of interactions that complement and augment speech [147]. Weimer and Ganapathy [148] developed a teleoperation system using hand gesturing and voice input, using a data glove to track the hand. Juster and Roy [149] developed "Elvis," a working robotic lighting system that translated speech and gesture commands into lighting changes. They found that Elvis was able to carry out multimodal interactions in real time with relatively high accuracy when interacting with familiar users. Yoshizaki *et al.* [150] described a service robot that used a vision-based interface to recognize gestures, as well as obtain real-world information about the objects mentioned in the user's speech. Finally, Mitsugami *et al.* [151] reported using gaze to operate multiple robots. With this system, the user first gives a voice command and then uses gaze to select a robot or specify the destination for the robot. This system can potentially free the operator's hands for other manual tasks but it also requires the operator to wear an HMD for eye-tracking purposes.

Zue [147] noted that the use of integrated multimodal input poses several challenges, including system interpretation. Inputs in different modalities must be understood in the proper context, so that when someone says, "What about that one" while pointing to an object in the environment, the system interprets the indirect referencing in the speech signal using the information in the visual channel. Zue stated that for a multimodal interface to be effective, it is important to develop a unifying linguistic formalism that can describe multimodal interactions (e.g., "move this one over here") as well as successful integration and delivery strategies.

IV. Conclusion

The burgeoning interest in robotic technology for both commercial and military use ensures that a diverse set of robotic applications will mature in the next decade [45]. The human role in robotic operations will be essential for the foreseeable future. Teleoperations in particular will be an important feature of almost all robotic operations. The irony of semiautonomous (and to some extent autonomous) systems is that teleoperations will be necessary under the most difficult situations when the underlying robotic intelligence and sensor capabilities require human intervention [12].

Teleoperation can be a challenging task because the operator is remotely located and has to operate the robots through video images, which tend to have a restricted FOV, provide limited depth information, and can be further degraded by bandwidth limitations. As a result, the operator's situation awareness of the remote environment can be compromised and the mission effectiveness can suffer. Teleoperation is also challenging in terms of operator's workload because he or she often has to switch among different camera views and/or maneuver the robots with a time delay due to technological limitations. With the U.S. Army's FCS, it is likely that the operator will have to control the robots from a moving vehicle, which will make the teleoperator's tasks even more difficult. This paper presented a detailed examination of these human performance issues and also reviewed some of the potential user interface solutions to these issues. U.S. Army research programs, such as the Army Technology Objective: Robotics-Collaboration, are starting to explore how to enhance operator performance by employing advanced technologies and user interface design concepts that will hopefully make operators' teleoperation tasks less challenging by reducing the workload and improving SA [11].

The purpose of this review was to establish preliminary constraints for human centered design of teleoperated systems and to point to areas where additional research is necessary. Table I is a summary of the findings, indicating perceptual and cognitive constraints, as well as the more obvious motor and tracking problem areas. As Murphy pointed out in her rescue work at the WTC, SA is a critical constraint in using teleoperated systems in complex environments [21], [105]. Fortunately, mitigating technologies, such as multimodal technology, predictive displays, improved interface design, and SD, are being developed to improve SA as well as operator control functions (Tables I and Tables II).

Multimodal user interfaces have been developed for a wide variety of robotic display and control tasks. Audio and haptic displays have been used in combination with visual displays to provide operator feedback in a wide variety of robotic teleoperation systems. Many of these systems employ bimodal feedback, specifically combinations of visual-audio and visual-haptic information. Some systems, such as Web-based applications, involve the simultaneous use of all three channels. Multimodal channels have been used to provide warnings information, communicate orientation, direct operator attention to events of interest, and provide virtual fixtures to guide or limit the direction of motion. Speech and gesture controls provide a wide range of natural interactions in the human–robot interface. As can be seen, multimodal controls and displays have great potential in decreasing task difficulty and creating a greater sense of immersion in robotic teleoperation tasks.

REFERENCES

[1] C. Bulich, A. Klein, R. Watson, and C. Kitts, "Characterization of delay-induced piloting instability for the Triton undersea robot," in *Proc. 2004 IEEE Aerospace Conf.*, 2004, vol. 1, p. 423.

[2] Casper and R. R. Murphy, "Human–robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," *IEEE Trans. Syst. Man, Cybern. B: Cybern.*, vol. 33, no. 3, pp. 367–385, Jun. 2003.

[3] L. A. Nguyen, M. Bualat, L. J. Edwards, L. Flueckiger, C. Neveu, K. Schwehr, M. Wagner, and E. Zbinden, "Virtual reality interfaces for visualization and control of remote vehicles," *Auton. Robots*, vol. 11, no. 1, pp. 59–68, 2001.

[4] J. Johnston, R. Wilson, and M. Birch, "Training for the operation of semi-autonomous systems—Interim report," Human Syst. Inform. Anal. Center, Wright Patterson AFB, OH, Tech. Rep. HSIAC-TR-2002–002, 2002.

[5] T. Fong, C. Thorpe, and B. Glass. (2003). PdaDriver: A handheld system for remote driving. presented at IEEE Int. Conf. Adv. Robot., [Online]. Available: http://imtsg14.epfl.ch/papers/ICAR03-TF.pdf

[6] A. Sekmen, A. B. Koku, and S. Zein-Sabatto, "Human robot interaction via cellular phones," in *Proc. 2003 IEEE Int. Conf. Syst., Man, Cybern.*, 2003, vol. 4, pp. 3937–3942.

[7] G. Kamsickas, "Future combat systems (FCS) concept and technology development (CTD) phase—Unmanned combat demonstration—Final report," Boeing Company, Seattle, WA, Tech. Rep. D786–1006102, 2003.

[8] T. Fong and C. Thorpe, "Vehicle teleoperation interfaces," *Auton. Robots*, vol. 11, pp. 9–18, 2001.

[9] H. K. Keskinpala and J. A. Adams, "Objective data analysis for a PDA-based human–robot interface," in *Proc. 2004 IEEE Int. Conf. Syst., Man, Cybern.*, 2004, vol. 3, pp. 2809–2814.

[10] A. A. Malcolm and J. S. G. Lim. (2003). Teleoperation control for a heavy-duty tracked vehicle. [Online]. Available: http://www.simtech. a-star.edu. sg/Research/TechnicalReports/TR0369.pdf.

[11] M. J. Barnes, J. Y. C. Chen, K. A. Cosenzo, and D. K. Mitchell, "Human robot teams as soldier augmentation in future battlefields: An overview," in *Proc. 11th Int. Conf. Hum.–Comput. Interact.*, vol. 11, 2005.

[12] S. P. Schipani. (2003). An evaluation of operator workload during partially-autonomous vehicle operations. presented at Proc. PerMIS. [Online]. Available: http://www.isd.mel.nist.gov/research_areas/research_engineering/Performance_Metrics/PerMIS_2003/Proceedings/Schipani.pdf

[13] M. J. Barnes, R. Parasuraman, and K. A. Cosenzo, "Adaptive automation for military robotic systems," in *Uninhabited Military Vehicles: Human Factor Issues in Augmenting the Force*, NATO Tech. Rep. RTO-TR-HFM-078, 2006, pp. 420–440.

[14] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proc. 2006 ACM Conf. Hum.–Robot Interact.*, 2006, pp. 33–40.

[15] T. Fong, C. Thorpe, and C. Baur, "Multi-robot remote driving with collaborative control," *IEEE Trans. Ind. Electron.*, vol. 54, no. 4, pp. 699–704, Aug. 2003.

[16] D. D. Woods, J. Tittle, M. Feil, and A. Roesler, "Envisioning human-robot coordination in future operations," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 2, pp. 210–218, May 2004.

[17] J. S. Tittle, A. Roesler, and D. D. Woods, "The remote perception problem," in *Proc. Hum. Factors Ergonom. Soc. 46th Annu. Meet.*, 2002, pp. 260–264.

[18] A. K. Kanduri, G. Thomas, N. Cabrol, E. Grin, and R. C. Anderson, "The (in)accuracy of novice rover operators' perception of obstacle height from monoscopic images," *IEEE Trans. Syst. Man, Cybern. A, Syst., Humans*, vol. 35, no. 4, pp. 505–512, Jul. 2005.

[19] D. Drascic, "An investigation of monoscopic and stereoscopic video for teleoperation," M.S. thesis, Univ. Toronto, Toronto, ON, Canada, 1991.

[20] R. P. Darken, K. Kempster, and B. Peterson, "Effect of streaming video quality of service on spatial comprehension in a reconnaissance task," in *Proc. 23rd Army Sci. Conf.*, 2001, pp. 391–400.

[21] R. R. Murphy, "Human–robot interaction in rescue robotics," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 2, pp. 138–153, May 2004.

[22] J. Carlson and R. R. Murphy. (2004). How UGVs physically fail in the field. [Online]. Available: http://crasar.csee.usf.edu/research/Publications/CRASAR-TR2004–16.pdf.

[23] L. C. Thomas and C. D. Wickens, "Effects of display frames of reference on spatial judgments and change detection," Univ. Illinois. Urbana-Champaign, Tech. Rep. ARL-00–14/FED-LAB-00–4, 2000.

[24] B. G. Witmer and W. J. Sadowski, "Nonvisually guided locomotion to a previously viewed target in real and virtual environments," *Hum. Factors*, vol. 40, pp. 478–488, 1998.

[25] J. B. F. Van Erp and P. Padmos, "Image parameters for driving with indirect viewing systems," *Ergonomics*, vol. 46, pp. 1471–1499, 2003.

[26] M. F. Land and D. Lee, "Where we look when we steer," *Nature*, vol. 369, pp. 742–744, 1994.

[27] A. B. Oving and J. B. F. Van Erp, "Driving with a head-slaved camera system," in *Proc. Hum. Factors Ergonom. Soc. 45th Annu. Meet.*, 2001, pp. 1372–1376.

[28] C. C. Smyth, V. Paul, A. M. Meldrum, and K. McDowell, "Examining alternative display configurations for an indirect vision driving interface," unpublished.

[29] D. R. Scribner and J. W. Gombash, "The effect of stereoscopic and wide field of view conditions on teleoperator performance," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, Tech. Rep. ARL-TR-1598, 1998.

[30] M. Voshell, D. D. Woods, and F. Phillips, "Overcoming the keyhole in human–robot coordination: Simulation and evaluation," in *Proc. Hum. Factors Ergonom. Soc. 49th Meet.*, 2005, pp. 442–446.

[31] C. C. Smyth, J. W. Gombash, and P. M. Burcham, "Indirect vision driving with fixed flat panel displays for near-unity, wide, and extended fields of camera view," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, Tech. Rep. ARL-TR-2511, 2001.

[32] A. J. Aretz, "The design of electronic map displays," *Hum. Factors*, vol. 33, pp. 85–101, 1991.

[33] S. M. Casner, "The effect of GPS and moving map displays on navigational awareness while flying under VFR," *J. Appl. Aviation Stud.*, vol. 5, no. 1, pp. 153–165, 2005.

[34] R. P. Darken and H. Cevik, "Map usage in virtual environments: Orientation issues," in *Proc. IEEE Virtual Reality Conf.*, 1999, pp. 133–140.

[35] M. C. Lohrenz, M. L. Gendron, S. S. Edwards, S. A. Myrick, and M. E. Trenchard, "Demonstration of a moving-map system for improved precise lane navigation of amphibious vehicles and landing craft," *Sea Technol.*, pp. 39–43, 2004.

[36] W. Wang, "Human navigation performance using 6 DOF dynamic viewpoint tethering in virtual environments," Ph.D., dissertation, Univ. Toronto, Toronto, ON, Canada, 2004.

[37] S. Werner, "Cognitive reference systems and their role in designing spatial information displays," *KI*, vol. 16, no. 4, pp. 10–13, 2002.

[38] U.S. Department of Transportation. (2006). Human factors design guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO). [Online]. Available: http://www.fhwa.dot.gov/tfhrc/safety/pubs/atis/ch05/

[39] S. M. Casner, "Mitigating the loss of navigational awareness while flying with GPS and moving map displays under VFR," *Int. J. Appl. Aviation Stud.*, vol. 6, no. 1, pp. 121–129, spring 2006.

[40] T. Heath-Pastore, "Improved operator awareness of teleoperated land vehicle attitude," SPAWAR, San Diego, CA, Tech. Rep. 1659, 1994.

[41] D. E. McGovern, "Experience and results in teleoperation of land vehicles," in *Pictorial Communication in Virtual and Real Environments*, S. R. Ellis, M. K. Kaiser, and A. C. Grunwald, Eds. London, U.K.: Taylor and Francis, 1991, pp. 182–195.

[42] W. A. Aviles, T. W. Hughes, H. R. Everett, A. Y. Umeda, S. W. Martin, A. H. Koyamatsu, M. Solorzano, R. Laird, and S. McArthur, "Issues in mobile robotics: The unmanned ground vehicle program teleoperated

vehicle (TOV)," in *Proc. SPIE* (Mobile Robots V), vol. 1388, pp. 587–597, 1991.

[43] M. Lewis, J. Wang, J. Manojlovich, S. Hughes, and X. Liu, "Experiments with attitude: Attitude displays for teleoperations," in *Proc. 2003 IEEE Int. Conf. Syst., Man, Cybern.*, 2003, vol. 2, pp. 1345–1349.

[44] J. Wang, M. Lewis, and S. Hughes, "Gravity-referenced attitude display for teleoperation of mobile robots," in *Proc. Hum. Factors Ergonom. Soc. 48th Annu. Meet.*, 2004, pp. 2662–2666.

[45] C. Shoemaker, "Robotics CTA," presented at the 2004 Collaborative Technol. Alliances Conf., Washington, DC.

[46] J. L. Drury, H. A. Yanco, W. Howell, B. Minten, and J. Casper, "Changing shape: Improving situation awareness for a polymorphic robot," in *Proc. 2006 ACM Conf. Hum.–Robot Interact.*, 2006, pp. 72–79.

[47] A. Rastogi, "Design of an interface for teleoperation in unstructured environments using augmented reality displays," M.S. thesis, Univ. Toronto, Toronto, ON, Canada, 1996.

[48] S. B. Hughes and M. Lewis, "Task-driven camera operations for robotic exploration," *IEEE Trans. Syst. Man, Cybern. A, Syst., Humans*, vol. 35, no. 4, pp. 513–522, Jul. 2005.

[49] D. R. Olsen and M. A. Goodrich. (2003). "Metrics for evaluating human-robot interactions," presented at Per MIS. [Online]. Available: http://www.isd.mel.nist.gov/research_areas/research_engineering/Performance_Metrics/PerMIS_2003/Proceedings/Olsen.pdf

[50] D. T. Levin and D. J. Simons, "Failure to detect changes to attended objects in motion pictures," *Psychonom. Bull. Rev.*, vol. 4, no. 4, pp. 501–506, 1997.

[51] J. Luck, P. L. McDermott, L. Allender, and A. Fisher, "Advantages of co-location for effective human to human communication of information provided by an unmanned vehicle," in *Proc. Hum. Factors Ergonom. Soc. 50th Annu. Meet.*, 2006, pp. 550–554.

[52] O. Olmos, C. D. Wickens, and A. Chudy, "Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering," *Int.J. Aviation Psychol.*, vol. 10, no. 3, pp. 247–271, 2000.

[53] L. C. Thomas and D. C. Wickens, "Visual displays and cognitive tunneling: Frames of reference effects on spatial judgments and change detection," in *Proc. Hum. Factors Ergonom. Soc. 45th Annu. Meet.*, 2001, pp. 336–340.

[54] D. D. Woods, "Visual momentum: A concept to improve the cognitive coupling of person and computer," *Int.J. Man-Mach. Stud.*, vol. 21, pp. 229–244, 1984.

[55] C. A. Johnson, J. A. Adams, and K. Kawamura, "Evaluation of an enhanced human-robot interface," in *Proc. 2003 IEEE Int. Conf. Syst., Man, Cybern.*, 2003, vol. 1, pp. 900–905.

[56] D. R. Lampton, M. J. Singer, D. McDonald, and J. P. Bliss, "Distance estimation in virtual environments," in *Proc. Hum. Factors Ergonom. Soc. 39th Annu. Meet.*, 1995, pp. 1268–1272.

[57] B. G. Witmer and P. Kline, "Judging perceived and traversed distance in virtual environments," *Presence: Teleoperators Virtual Environ.*, vol. 7, no. 2, pp. 144–167, 1998.

[58] J. L. Marble, D. J. Bruemmer, and D. A. Few, "Lessons learned from usability tests with a collaborative cognitive workspace for human-robot teams," in *Proc. 2003 IEEE Conf. Syst Man, Cybern.*, vol. 1, pp. 448–453.

[59] J. V. Draper, S. Handel, and C. C. Hood, "Three experiments with stereoscopic television: When it works and why," in *Proc. 1991 IEEE Int. Conf. Syst., Man, Cybern.*, 1991, vol. 2, pp. 1047–1052.

[60] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.*, vol. 47, pp. 381–391, 1954.

[61] L. B. Rosenberg, "The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks," in *Proc. 1993 IEEE Virtual Reality Annu. Int. Symp.*, pp. 27–32.

[62] J. Green, E. Dougherty, and E. Savacool, "Stereovision system for shipboard cranes," in *Proc. Interservice/Ind. Training, Simul., Educ. Conf.*, 2003.

[63] D. Drascic and J. Grodski, "Defense teleoperation and stereoscopic video," in *Proc. SPIE*, vol. 1915, pp. 58–69, 1993.

[64] J. O. Merritt, V. G. CuQlock-Knopp, M. Kregel, J. Smoot, and W. Monaco, "Perception of terrain drop-offs as a function of L-R viewpoint separation in stereoscopic video," in *Proc. SPIE*, vol. 5800, pp. 169–176, 2005.

[65] B. Vaughan. U.S. Army Research Lab., Aberdeen Proving Ground, MD, private communication, Mar. 2006.

[66] J. French, T. G. Ghirardelli, and J. Swoboda, "The effect of bandwidth on operator control of an unmanned ground vehicle," in *Proc. Interservice/Ind. Training, Simul., Educ. Conf.*, 2003.

[67] R. P. Darken and B. Peterson, "Spatial orientation, wayfinding, and representation," in *Handbook of Virtual Environment Technology*, K. Stanney, Ed. Mahwah, NJ: Erlbaum, 2002, pp. 493–518.

[68] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press, 1992.

[69] B. Watson, N. Walker, W. Ribarsky, and V. Spaulding, "Effects of variation in system responsiveness on user performance in virtual environments," *Hum. Factors*, vol. 40, pp. 403–414, 1998.

[70] M. Massimino and T. B. Sheridan, "Teleoperator performance with varying force and visual feedback," *Hum. Factors*, vol. 36, pp. 145–157, 1994.

[71] J. Y. C. Chen, P. J. Durlach, J. A. Sloan, and L. D. Bowens, "Human robot interaction in the context of simulated route reconnaissance missions," *Military Psychol.*, to be published.

[72] J. E. Thropp and J. Y. C. Chen, "Effects of slow frame rates on human performance," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, Tech. Rep. ARL-TR-3813, 2006.

[73] B. Watson, V. Spaulding, N. Walker, and W. Ribarsky, "Evaluation of the effects of frame rate variation on VR task performance," in *Proc. 1997 Virtual Reality Annu. Int. Symp.*, pp. 38–44.

[74] P. Richard, G. Birebent, P. Coiffet, G. Burdea, D. Gomez, and N. Langrana, "Effect of frame rate and force feedback on virtual object manipulation," *Presence: Teleop. Virtual Environ.*, vol. 5, no. 1, pp. 95–108, 1996.

[75] M. Baker, R. Casey, B. Keyes, and H. A. Yanco, "Improved interfaces for human-robot interaction in urban search and rescue," in *Proc. 2004 IEEE Int. Conf. Syst., Man, Cybern.*, vol. 3, pp. 2960–2965.

[76] G. L. Calhoun and M. H. Draper, "Multi-sensory interfaces for remotely operated vehicles," in *Human Factors of Remotely Operated Vehicles*, N. J. Cooke, H. L. Pringle, H. K. Pedersen, and O. Connor, Eds. Oxford, U.K.: Elsevier, 2006, pp. 149–163.

[77] T. H. J. Collett and B. A. MacDonald, "Developer oriented visualization of a robot program: An augmented reality approach," in *Proc. 2006 ACM Conf. Hum.–Robot Interact.*, pp. 49–56.

[78] M. Daily, Y. Cho, K. Martin, and D. Payton, "World embedded interfaces for human-robot interaction," in *Proc. 36th Hawaii Int. Conf. Syst. Sci.*, 2003, pp. 125–130.

[79] C. W. Nielsen and M. A. Goodrich, "Comparing the usefulness of video and map information in navigation tasks," in *Proc. 2006 ACM Conf. Hum.–Robot Interact.*, pp. 95–101.

[80] D. Tufano, Automotive HUDs: The overlooked safety issues. *Hum. Factors*, vol. 39, pp. 303–311, 1997.

[81] C. D. Wickens. "Attentional tunneling and task management," University of Illinois, Urbana-Champaign, IL, Tech. Rep. AHFD-05-23/NASA-05-10, 2005.

[82] S. MacKenzie and C. Ware, "Lag as a determinant of human performance in interactive systems," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 1993, pp. 488–493.

[83] S. R. Ellis, K. Mania, B. D. Adelstein, and M. Hill, "Generalizability of latency detection in a variety of virtual environments," in *Proc. Hum. Factors Ergonom. Soc. 48th Annu. Meet.*, 2004, pp. 2632–2636.

[84] J. C. Lane, C. R. Carignan, B. R. Sullivan, D. L. Akin, T. Hunt, and R. Cohen, "Effects of time delay on telerobotic control of neutral buoyancy vehicles," in *Proc. 2002 IEEE Int. Conf. Robot. Autom.*, 2002, vol. 3, pp. 2874–2879.

[85] T. B. Sheridan and W. R. Ferrell, "Remote manipulative control with transmission delay," *IEEE Trans. Hum. Factors Electron.*, vol. 4, no. 1, pp. 25–29, Sep. 1963.

[86] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*. Hoboken, NJ: Wiley, 2002.

[87] R. Held, A. Efstathiou, and M. Greene, "Adaptation to displaced and delayed visual feedback from the hand," *J. Exp. Psychol.*, vol. 72, pp. 887–891, 1966.

[88] L. H. Frank, J. G. Casali, and W. W. Wierville, "Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator," *Hum. Factors*, vol. 30, pp. 201–217, 1988.

[89] K. W. Arthur, K. S. Booth, and C. Ware, "Evaluating 3D task performance for fish tank virtual worlds," *ACM Trans. Inf. Syst.*, vol. 11, no. 3, pp. 239–265, 1993.

[90] S. R. Ellis, B. D. Adelstein, S. Baumeler, G. J. Jense, and R. H. Jacoby, "Sensor spatial distortion, visual latency, and update rate effects on 3D

tracking in virtual environments," in *Proc. Virtual Reality Conf.*, 1999, pp. 218–221.
[91] J. P. Luck, P. L. McDermott, L. Allender, and D. C. Russell, "An investigation of real world control of robotic assets under communication latency," in *Proc. 2006 ACM Conf. Hum.–Robot Interact.*, 2006, pp. 202–209.
[92] K. M. Stanney, R. R. Mourant, and R. S. Kennedy, "Human factors issues in virtual environments: A review of the literature," *Presence: Teleop. Virtual Environ.*, vol. 7, pp. 327–351, 1998.
[93] E. M. Kolasinski, "Simulator sickness in virtual environments," U.S. Army Res. Inst., Alexandria, VA, Tech. Rep. 1027, 1995.
[94] D. B. Kaber, J. M. Riley, R. Zhou, and J. Draper, "Effects of visual interface design, and control mode and latency on performance, telepresence and workload in a teleoperation task," in *Proc. Hum. Factors Ergonom. Soc. 44th Annu. Meet.*, 2000, pp. 503–506.
[95] J. V. Draper, D. B. Kaber, and J. M. Usher, "Telepresence," *Hum. Factors*, vol. 40, pp. 354–375, 1998.
[96] B. G. Witmer and M. J. Singer, "Measuring immersion in virtual environments," U.S. Army Res. Inst., Alexandria, VA, Tech. Rep. 1014, 1994.
[97] J. V. Draper, D. B. Kaber, and J. M. Usher, "Speculations on the value of telepresence," *CyberPsychol. Behav.*, vol. 2, no. 4, pp. 349–362, 1999.
[98] A. Kheddar, R. Chellali, and P. Coiffet, "Virtual environment-assisted teleoperation," in *Handbook of Virtual Environment Technology*, K. Stanney, Ed. Mahwah, NJ: Erlbaum, 2002, pp. 959–997.
[99] Z. Deng and M. Jagersand. (2003). Predictive display system for tele-manipulation using image-based modeling and rendering. [Online]. Available: http://www.cs.ualberta.ca/~jag/research/papers/iros03predSys.pdf
[100] B. Ricks, C. W. Nielsen, and M. A. Goodrich, "Ecological displays for robot interaction: A new perspective," in *Proc. IROS 2004*, vol. 3, pp. 2855–2860.
[101] P. S. Cowings, W. B. Toscano, C. DeRoshia, and R. A. Tauson, "Effects of command and control vehicle (C2 V) operational environment on soldier health and performance," NASA Ames, Moffet Field, CA, Tech. Rep. NASA/TM-1999–208786, 1999.
[102] S. G. Hill, C. Stachowiak, and R. A. Tauson, "Soldier performance in the enclosed compartment of a moving vehicle," presented at the NATO Appl. Veh. Technol. Panel Symp., Habitability Combat Transp. Veh.: Noise, Vibration Motion, Paper RTO-MP-AVT-110, 2004.
[103] S. P. Schipani, R. S. Bruno, M. A. Lattin, B. M. King, and D. J. Patton, "Quantification of cognitive process degradation while mobile, attributable to the environmental stressors endurance, vibration, and noise," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, Paper ARL-TR-1603, 1998.
[104] S. G. Hill and R. A. Tauson. (2005). Soldier performance issues in C2 'On the move'. presented at 10th Int. Command Control Res. Technol. Symp., [Online]. Available: http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/074.pdf
[105] R. R. Murphy and J. L. Burke, "Up from the rubble: Lessons learned about HRI from search and rescue," in *Proc. Hum. Factors Ergonom. Soc. 49th Annu. Meet.*, 2005, 437–441.
[106] C. E. Lathan and M. Tracey, "The effects of operator spatial perception and sensory feedback on human-robot teleoperation performance," *Presence: Teleoperators Virtual Environ.*, vol. 11, no. 4, pp. 368–377, 2002.
[107] G. Burdea, P. Richard, and P. Coiffet, "Multimodal virtual reality: Input-output devices, system integration, and human factors," *Int. J. Hum.–Comput. Interact.*, vol. 8, no. 1, pp. 5–24, 1996.
[108] G. V. Popescu, G. C. Burdea, and H. Trefftz, "Multimodal interaction modeling," in *Handbook of Virtual Environments*, K. Stanney, Ed. Mahwah, NJ: Erlbaum, 2002, pp. 65–92.
[109] G. Calhoun, M. Draper, H. Ruff, and J. Fontejon, "Utility of a tactile display for cueing faults," in *Proc. Hum. Factors Ergonom. Soc. 46th Annu. Meet.*, 2002, pp. 2144–2148.
[110] R. D. Shilling and B. Shinn-Cunningham, "Virtual auditory displays," in *Handbook of Virtual Environments*, K. Stanney, Ed. Mahwah, NJ: Erlbaum, 2002, pp. 65–92.
[111] N. Y. Chong, T. Kotoku, K. Ohba, H. Sasaki, K. Komoriya, and K. Tanie, "Multioperator teleoperation of multirobot systems with time delay. II—Testbed description," *Presence: Teleop. Virtual Environ.*, vol. 11, no. 3, pp. 292–303, 2002.
[112] Y. Nagai, S. Tsuchiya, and S. Kimura, "A study on audio feedback system for the operation of Engineering Test Satellite VII—Effects on telemetry monitor," in *Proc. Int. Sym. Space Technol. Sci.*, 2000, pp. 1226–1231.
[113] B. D. Simpson, R. S. Bolia, and M. H. Draper, "Spatial auditory display concepts supporting situation awareness for operators of unmanned aerial vehicles," in *Human Performance, Situation Awareness, and Automation: Current Research and Trends,* vol. 1, D. A. Vincenzi, M. Mouloua, and P. A. Hancock, Eds. Mahwah, NJ: Erlbaum, 2004, pp. 61–65.
[114] S. Tachi, K. Komoriya, K. Sawada, T. Nishiyama, T. Itoko, M. Kobayashi, and K. Inoue, "Telexistence cockpit for humanoid robot control," *Adv. Robot.*, vol. 17, no. 3, pp. 199–217, 2003.
[115] F. Gemperle, N. Ota, and D. Siewiorek, "Design of a wearable tactile display," in *Proc. 5th IEEE Int. Symp. Wearable Comput.*, 2001, pp. 5–12.
[116] R. W. Cholewiak and A. A. Collins, "The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode," *Percept. Psychophys.*, vol. 62, pp. 1220–1235, 2000.
[117] J. L. Rochlis and D. J. Newman, "A tactile display for international space station (ISS) extravehicular activity (EVA)," *Aviation Space Environ. Med.*, vol. 71, pp. 571–578, 2000.
[118] G. Calhoun, J. Fontejon, M. Draper, H. Ruff, and B. Guilfoos, "Tactile versus aural redundant alert cues for UAV control applications," in *Proc. Hum. Factors Ergonom. Soc. 48th Annu. Meet.*, 2004, pp. 137–141.
[119] M. Sitti and H. Hashimoto, "Teleoperated nano scale object manipulation," in *Recent Advances on Mechatronics*, O. Kaynak, Ed. New York: Springer-Verlag, 1999, pp. 172–178.
[120] A. Kortschack, A. Shirinov, T. Truper, and S. Fatikow, "Development of mobile versatile nanohandling microrobots: Design, driving principles, haptic control," *Robotica*, vol. 23, no. 4, pp. 419–434, 2005.
[121] J. Aleotti, S. Bottazzi, S. Caselli, and M. Reggiani. (2002). A multimodal user interface for remote object exploration in teleoperation systems. [Online]. Available: http://rimlab.ce.unipr.it/Publications/HUROIN2002.pdf
[122] A. M. Murray, R. L. Klatzky, and P. K. Khosla, "Psychophysical characterization and testbed validation of a wearable vibrotactile glove for telemanipulation," *Presence: Teleop. Virtual Environ.*, vol. 12, no. 2, pp. 156–182, 2003.
[123] S. Charles, H. Das, T. Ohm, C. Boswell, G. Rodriguez, R. Steele, and D. Istrate, "Dexterity-enhanced telerobotic microsurgery," in *Proc. Int. Conf. Adv. Robot.*, 1997, pp. 5–10.
[124] C. W. Kennedy, T. Hu, J. P. Desai, A. S. Wechsler, and J. Y. Kresh, "A novel approach to robotic cardiac surgery using haptics and vision," *Cardiovascular Eng.: Int.J.*, vol. 2, no. 1, pp. 15–22, 2002.
[125] M. Kitagawa, D. Dokko, A. M. Okamura, and D. D. Yuh, "Effect of sensory substitution on suture-manipulation forces for robotic surgical systems," *J. Thoracic Cardiovascular Surg.*, vol. 129, pp. 151–158, 2005.
[126] D. Kragic, P. Marayong, M. Li, A. M. Okamura, and G. D. Hager, "Human–machine collaborative systems for microsurgical applications," *Int.J. Robot. Res.*, vol. 24, no. 9, pp. 731–741, 2005.
[127] L. Rosenberg, "Virtual fixtures: Perceptual overlays enhance operator performance in telepresence tasks," Ph.D. dissertation, Dept. Mech. Eng., Stanford Univ., Stanford, CA, 1994.
[128] D. V. Gunn, J. S. Warm, W. T. Nelson, R. S. Bolia, D. A. Schumsky, and K. J. Corcoran, "Target acquisition with UAVs: Vigilance displays and advanced cuing interfaces," *Hum. Factors*, vol. 47, pp. 488–497, 2005.
[129] J. Y. C. Ng, J. C. F. Man, S. Fels, G. Dumont, and J. M. Ansermino, "An evaluation of a vibro-tactile display prototype for physiological monitoring," *Anesth. Analg.*, vol. 101, no. 6, pp. 1719–1724, 2005.
[130] M. Massimino and T. Sheridan, "Sensory substitution of force feedback for the human-machine interface in space teleoperation," presented at the 43rd Congr. Int. Astronaut. Fed., World Space Congress, Paper IAF/iAA-92–0246, 1992.
[131] W. Chou and T. Wang, "The design of multimodal human–machine interface for teleoperation," in *Proc. 2001 IEEE Int. Conf. Syst., Man, Cybern.*, 2001, vol. 5, pp. 3187–3192.
[132] I. Elhajj, N. Xi, W. K. Fung, Y-H Liu, Y. Hasegawa, and T. Fukuda, "Supermedia-enhanced Internet-based telerobotics," in *Proc. Inst. Elect. Eng.*, vol. 91, no. 3, pp. 396–421, 2003.
[133] J. J. Abbott, P. Marayong, and A. M. Okamura, "Haptic virtual fixtures for robot-assisted manipulation," in *Robotics Research: Results of the 12th International Symposium ISRR,* S. Thrun, R. Brooks, and H. Durrant-Whyte, Eds. New York, NY: Springer, 2007, pp. 49–64.
[134] C. P. Sayers and R. P. Paul, "An operator interface for teleprogramming employing synthetic fixtures," *Presence: Teleop. Virtual Environ.*, vol. 3, pp. 309–320, 1994.
[135] J. Aleotti, S. Caselli, and M. Reggiani, "Evaluation of virtual fixtures for a robot programming by demonstration interface," *IEEE Trans. Syst. Man, Cybern. A, Syst., Humans*, vol. 35, no. 4, pp. 536–545, Jul. 2005.

[136] M. Draper, G. Calhoun, H. Ruff, D. Williamson, and T. Barry, "Manual versus speech input for unmanned aerial vehicle control station operations," in *Proc. Hum. Factors Ergonom. Soc. 47th Annu. Meet.*, 2003, pp. 109–113.
[137] H. Downs, *Integrated Remote Neurosurgical System*. Charlottesville, VA: Virginia Neurological Inst., 1998.
[138] H. Rininsland, "ARTEMIS: A telemanipulator for Cardiac surgery," *Eur. J. Cardio-Thoracic Surg.*, vol. 16, pp. S106–S111, 1999.
[139] L. F. Penin, R. Aracil, M. Ferre, E. Pinto, M. Hernando, and A. Barrientos, "Telerobotic system for live power lines maintenance: ROBTET," in *Proc. 1998 IEEE Int. Conf. Robot. Autom.*, vol. 3, pp. 2110–2115.
[140] S. M. Goza, R. O. Ambrose, M. A. Diftler, and I. M. Spain, "Telepresence control of the NASA/DARPA robonaut on a mobility platform," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2004, pp. 623–629.
[141] D. Perzanowski, D. Brock, W. Adams, M. Bugajska, A. Schultz, J. Trafton, S. Blisard, and M. Skubic, "Finding the FOO: A pilot study for a multimodal interface," in *Proc. 2003 IEEE Conf. Syst., Man, Cybern.*, vol. 4, pp. 3218–3223.
[142] C. Hu, M. Q. Meng, P. X. Liu, and X. Wang, "Visual gesture recognition for human–machine interface of robot teleoperation," in *Proc. 2003 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 1560–1565.
[143] M. Frigola, J. Fernandez, and J. Aranda, "Visual human machine interface by gestures," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2003, pp. 386–391.
[144] C. J. Cohen, L. Conway, and D. Kiditschek, "Dynamical system representation, generation, and recognition of basic oscillatory motion gestures," in *Proc. 2nd Int. Conf. Face Gesture Recog.*, 1996, pp. 60–65.
[145] Y. Kuno, S. Nakanishi, T. Murashima, N. Shimada, and Y. Shirai, "Robotic wheelchair with three control modes," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1999, pp. 2590–2594.
[146] E. Adrizzone, A. Chella, and R. Pirrone, "An architecture for automatic gesture analysis," in *Proc. Working Conf. Adv. Vis. Interfaces*, 2000, pp. 205–256.
[147] V. Zue, "Eighty challenges facing speech input/output technologies," in *Proc. From Sound Sense: 50+ Years Discovery Speech Commun.*, 2004, pp. B179–B195.
[148] D. Weimer and S. K. Ganapathy, "A synthetic visual environment with hand gesturing and voice input," in *Proc. Comput.–Hum. Interact.*, 1989, pp. 235–240.
[149] J. Juster and D. Roy, "Elvis: Situated speech and gesture understanding for a robotic chandelier," in *Proc. Int. Conf. Multimodal Interfaces*, 2004, pp. 90–95.
[150] M. Yoshizaki, Y. Kuno, and A. Nakamura, "Human–robot interface based on the mutual assistance between speech and vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2002, pp. 1308–1313.
[151] I. Mitsugami, N. Ukita, and M. Kidode, "Robot navigation by eye pointing," in *Proc. 4th Int. Conf. Entertainment Comput.*, 2005, pp. 256–267.

Jessie Y. C. Chen received the B.A. degree in linguistics from the National Tsing-Hua University, Hsinchu, Taiwan, R.O.C., in 1987, the M.A. degree in communication studies from the University of Michigan, Ann Arbor, in 1989, and the Ph.D. degree in applied experimental and human factors psychology from the University of Central Florida, Orlando, in 2000.

She is currently a Research Psychologist with the U.S. Army Research Laboratory (ARL), Orlando, FL. Prior to joining ARL, she was a Postdoctoral Fellow with the U.S. Army Research Institute for Behavioral and Social Sciences, Arlington, VA. Her research interests include human–robot interaction, robotics controller performance issues, and user interface designs for robotics control.

Ellen C. Haas received the B.A. degree in psychology from the University of Colorado, Boulder, in 1976, the M.S. degree in industrial psychology from California State University, Long Beach, in 1980, and the Ph.D. degree in industrial and systems engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1993.

She has been a Research Engineer with the U.S. Army Research Laboratory (ARL), Aberdeen Proving Ground, MD, for over 20 years, and heads the ARL Auditory Controls and Displays Laboratory. She has written and published extensively about auditory warnings and displays, multimodal displays, and speech recognition controls. She has coedited the book *The Ergonomics of Sound* (Human Factors and Ergonomics Society, 2003).

Dr. Haas has been a member of the International Ergonomics Association (IEA), and is the founder and Chair of the IEA Auditory Ergonomics Technical Committee. She has been a member of the Human Factors and Ergonomics Society for over 25 years, and is a member of the American Psychology Association Division 21 (Applied Experimental and Engineering Psychology).

Michael J. Barnes received the B.A. degree in psychology from the University of Arizona, Tucson, in 1972 and the M.A. degree in experimental psychology from New Mexico State University, Las Cruces, in 1975.

He is currently with the U.S. Army Research Laboratory (ARL), Aberdeen Proving Ground, MD, where he is the ARL Lead for the Army Technology Objective (ATO): Collaborative Robotics. He was a Researcher with the Naval Weapons Center, China Lake, CA, and the Naval Air Development Center, Warminster, PA. He was also the Aegis Combat System human factor's Unit Manager, General Electric Cooperation, Moorestown, NJ. His current interests include cognitive research in intelligence analysis, language acquisition tools, unmanned aerial vehicles, and visualization tools for risk management.