DE GRUYTER OPEN — Paladyn, J. Behav. Robot. 2015

Research Article **Open Access**

Katherine M. Tsui*, James M. Dalphond, Daniel J. Brooks, Mikhail S. Medvedev, Eric McCann, Jordan Allspaw, David Kontak, and Holly A. Yanco

Accessible Human-Robot Interaction for Telepresence Robots: A Case Study

Abstract: The quality of life of people with special needs, such as residents of healthcare facilities, may be improved through operating social telepresence robots that provide the ability to participate in remote activities with friends or family. However, to date, such platforms do not exist for this population. **Methodology**: Our research utilized an iterative, bottom-up, user-centered approach, drawing upon our assistive robotics experiences. Based on the findings of our formative user studies, we developed an augmented reality user interface for our social telepresence robot. Our user interface focuses primarily on the human-human interaction and communication through video, providing support for semi-autonomous navigation. We conducted a case study (n=4) with our target population in which the robot was used to visit a remote art gallery. **Results**: All of the participants were able to operate the robot to explore the gallery, form opinions about the exhibits, and engage in conversation. **Significance**: This case study demonstrates that people from our target population can successfully engage in the active role of operating a telepresence robot.

Keywords: accessible user interface, augmented reality, teleoperation, assistive robotics, social telepresence robot, computer-mediated communication, remote presence, embodied video conferencing

1 Introduction

Assistive technology is impacted by the state of social norms and benefits directly from the consumer electronics market. Commercial telepresence robots are being sold as mediums for impromptu, mobile, embodied video conferencing; see Kristoffersson et al. [41] and Tsui and Yanco [75] for an overview. Telepresence robots provide interactive two-way audio and video communication over the Internet. Additionally, these telepresence robots can be controlled independently by an operator, which means that the person driving can explore and look around as he or she desires. These robots are at the intersection of physical and social presence, called *copresence* [34, 60].

Our research focuses on the use case in which people with special needs take the active role of operating telepresence robots; our target population is people with cognitive and/or motor impairments. It should be noted there has been considerable research already done in the use case of the passive role in which the person with special needs is visited by a healthcare professional, family member, or friend operating a telepresence robot (e.g., [6, 20, 21, 27, 84, 85]). The active role is depicted as the green person in Fig. 1a, and the passive role as the blue person in Fig. 1b.

Hassenzahl [30] describes a "user experience" as the answers to three questions: *why*, *what*, and *how*. *Why* speaks to the motivation to use the device, particularly the needs and emotions forming the experience and their meaning. We believe that telepresence robots can be used to support social engagement for people who reside at medical institutions, for example, in recreating the closeness one would have if he or she were physically present with his or her family. For some people, the telepresence robot may be used exclusively as a conversation tool. Other people may want to check on their

---

***Corresponding Author: Katherine M. Tsui:** University of Massachusetts Lowell (now at Yale University), E-mail: ktsui@cs.uml.edu
James M. Dalphond: University of Massachusetts Lowell, E-mail: jdalphon@cs.uml.edu
Daniel J. Brooks: University of Massachusetts Lowell, E-mail: dbrooks@cs.uml.edu
Mikhail S. Medvedev: University of Massachusetts Lowell (now at IBM), E-mail: mmedvede@cs.uml.edu
Eric McCann: University of Massachusetts Lowell, E-mail: emccann@cs.uml.edu
Jordan Allspaw: University of Massachusetts Lowell, E-mail: jallspaw@cs.uml.edu
David Kontak: Crotched Mountain Rehabilitation Center, E-mail: david.kontak@crotchedmountain.org
Holly A. Yanco: University of Massachusetts Lowell, E-mail: holly@cs.uml.edu

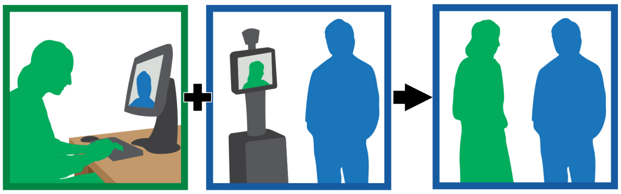


Fig. 1. Social telepresence definition [75]. The robot's user (green person in (**a**)) operates a telepresence robot in a remote environment (left side of image (**b**)). In this interpersonal communication use case, the user converses with an interactant (blue person in (**b**)). The degree to which the user feels telepresent with the interactant in the remote environment and vice versa (**c**) is dependent upon the quality of both the user's human-computer interaction (a) and the interactant's human-robot interaction (b).

family and observe them, while still others may wish to attend an art exhibit opening or tour a museum [6, 16]. Others may simply want to be present in a space to feel more included in an activity, such as attending high school via a telepresence robot [46, 58].

What lists the function(s) that people can do with a device [30]. Telepresence robots support "calls," which allow you to connect with another person. Once in a call, the robot acts as the caller's physical avatar. We believe that telepresence robots have the potential to recreate the desired closeness better than a telephone or video chat. Hassenzahl [30] provides insight as to why:

> "We have all experienced the awkward silence when we have run out of stories to tell while not wanting to hang up on our loved one. This is the result of a misfit between the conversational model embodied by a telephone and the psychological requirements of a relatedness experience. [30]."

Finally, *how* describes the design of the device and its interface [30]. User experience designers must consider individual components of a system and interactions between them in concert with an end goal. There are six individual components common to all social telepresence use cases [75]:

1. the robot itself (herein referred to as the *telepresence robot*),
2. the robot's user (herein referred to as the *user*, which is the active role),
3. the unit with which the user controls the telepresence robot (herein referred to as the *interface*),
4. the user's environment,
5. the robot's environment and the objects in it (herein referred to as the *remote environment*), and
6. the people in the remote environment who are physically co-located with the robot and may interact directly or indirectly with the user (herein referred to as *interactants* and *bystanders*, which are the passive roles, respectively).

Telepresence robot designers must consider three main interactions. First, there is the human-computer interaction between the user and the robot's interface (Fig. 1a), which allows the user to operate the robot in the remote environment; this interaction is often also considered human-robot interaction (HRI) by the research community (e.g., [8, 22, 32, 35, 47]). Second, there is the HRI in the remote environment between the interactants and the telepresence robot itself (Fig. 1b); the interactants converse with the user through his or her telepresence robot embodiment. Finally, there is the interpersonal human-human interaction (Fig. 1c); if these first two interactions are successful, then robot mediation will be minimized [69], and the experience of telepresence (i.e., the user's sense of remote presence and the interactants' sense of the user being telepresent) is maximized [15, 45].

To date, telepresence robots, their user interfaces, and their navigation behaviors have not been designed for use by people with special needs to be the robot operators. Designing a social telepresence robot system to maximize the user experience, for our target population, requires expertise in the domains of HRI and assistive robotics and the understanding of how to balance the two. In Section 2, we describe several key insights about commanding telepresence robots. We describe the remote art gallery environment in Section 3. Our social telepresence robot research platform is summarized in Section 4, including its movement behaviors. Our research has synthesized user interface (UI) design guidelines and principles from the domains of HRI, human-computer interaction (HCI), and assistive technology. In Section 5, we enumerate the design principles key in facilitating the development of telepresence robot interfaces for use by our target population; we demonstrate these design principles at work in Section 5.2. We present a case study (n=4) in Section 6, which demonstrates this research as a first critical step towards having our target population take the active role of the telepresence robot operator. Finally, we discuss the generalizability and scalability of our system in Section 7.

2 Related Work

Designing HRI systems for people with cognitive and/or motor impairments (our target population) is difficult. There are several approaches, originating from the research areas of HCI, human factors engineering, and assistive and rehabilitation technology. Our approach in designing HRI systems has been an iterative process which involves the target population (primary stakeholders), caregivers (secondary stakeholders), and clinicians from the beginning, formative stages through the summative evaluations, which is similar to the processes used by Amirabdollahian et al. [2], Cooper [12], and Schulz et al. [64]. We utilized this approach while also drawing upon our experiences in the domains of assistive robotics and HCI.

Just prior to the emergence of the first generation of commercial telepresence robots in 2010, we conducted a series of feasibility studies using prototype telepresence robots at Google in Mountain View, CA, with remote office workers and their teams. We developed several guidelines for the next-generation design of telepresence robots [14, 71]. Two key insights resulted from this early work. First, a wide field of view is needed to operate a telepresence robot, both horizontally and vertically. Second, some level of autonomous navigation is required. Many contemporary commercial telepresence robots are controlled by keyboard or mouse and directed to move forward/backward, turn left/right, or stop; see Kristoffersson et al. [41] and Tsui and Yanco [75]. This direct teleoperation is impractical due to inherent network latency and the movement of people in the remote environment. Able-bodied novices had difficulty driving telepresence robots straight down a corridor [14, 71], as the latency often caused the robot to turn more than the user intended and thus zig zag down the hallway.

The second insight was reinforced in a preliminary evaluation with our target population operating an alpha-version commercial telepresence robot using arrow keys [72]. We found that direct teleoperation, i.e., continuous robot movement, was an issue with our target population's mental model of the robot due to the latency between issuing the commands, the robot receiving the commands, the robot executing the command, and the video updating to show the robot moving. We believe that autonomous and semi-autonomous navigation behaviors are necessary for a person with special needs to use telepresence robots. Autonomous navigation behaviors can free the user from the details of robot navigation, making the driving task easier; consequently, the user can focus on the primary communication task or exploring the remote environment.

To date, there are few examples of people with special needs using telepresence robots in the real world. Only the PEBBLES robots, developed in 1997 by Telebotics, the University of Toronto, and Ryerson University, had been used by students with disabilities to attend their regularly scheduled classes during their hospitalizations [17, 18, 33, 63]; as of June 2006, there were forty PEBBLES robots on loan to hospitals [3]. One robot was placed in the child's classroom and the other robot was with the child. The primary function of PEBBLES is to provide a window into the classroom, and the child can "look around" his or her class room as PEBBLES's head can move left/right and up/down. However, it should be noted that PEBBLES robot was a passive mobile system, and an attendant was required to push the robot from one location to another.

Thus, it was important to understand how members of our target population conceptualize a remote environment and what they expected a telepresence robot to be able to do in terms of navigation in the given space. We conducted two formative evaluations regarding autonomous robot navigation by investigating a speech interface; our intention in eliciting speech was to mitigate any associations that participants from our target population may have had with joystick controls, which are common, low-level input devices for operating power wheelchairs. First, we conducted a focus group (n=5) to investigate how members of our target audience would want to direct a telepresence robot through several scenarios in a remote environment using speech [80]. We then conducted a follow-on experiment in which participants (n=12) directed a telepresence robot or a human in a scavenger hunt task [79, 80]. We collected a corpus of 312 utterances (first hand as opposed to speculative) relating to spatial navigation. From this corpus, we found that all participants gave directives at the low-level (i.e., forward, backward, left, right, stop), mid-level (i.e., referring to information within the robot's camera view), and high-level (i.e., requests to send the robot to places beyond its current camera view). We drew a third key insight: users would command the robots using multiple levels of abstraction.

3 Art Gallery Built for Case Study

Beer and Takayama [4] found that seniors wanted to use the Texai telepresence robot to attend concerts or sporting events, and visit museums or theatres. A number of telepresence robots have been placed in museums, allowing remote visitors to see a given museum from the robot's perspective (e.g., [16, 43]). We chose an art gallery scenario, as going to a museum or art gallery can be an individual experience, and also a shared social one. In their 2010 visitor survey, the Museum of Science, Boston, found that 51.8% of respondents said that they visited the museum to spend time together as a group or family [44].

To give users an interesting environment in which to explore through a telepresence robot, we built an Artbotics gallery with kinetic, interactive exhibits. Artbotics was created through a collaboration between the University of Massachusetts Lowell and The Revolving Museum in Lowell, MA. In one of its offerings, Artbotics is taught as an undergraduate course in which students from various majors work to build robotic art installations [90]. Each Artbotics exhibit uses sensors such as infrared (IR) distance sensors or buttons to allow a person to interact with the art. When the sensors are triggered, the actuators (motors, servos, lights) react in some manner based on how the exhibit was programmed by its creator.

Five exhibits were chosen from the final projects from the Artbotics class: "Sunflower," "Vincent," "Face," "Monkey," and "Music," shown from left to right in Fig. 2. Each of these exhibits was selected because of its unique content and range of movements. All five exhibits were mounted on, or in front of, a 2.44 × 2.13 m (8 × 7 foot) plywood wall, which was painted to match the theme of the corresponding exhibit. Three IR distance sensors were placed, centered, under a kick plate in front of each exhibit (see Fig. 3). The area directly in front of each sensor is defined as an exhibit *hotspot* – a place where the user could interact with the exhibit; exhibits were programmed to react differently when approached from each hotspot as well as at a close and far distance. Each exhibit had a fourth hotspot on its rightmost side, called *info*. At this location, a placard with the exhibit's title and a brief description outlined in a black border was displayed, centered 45.7 cm (18 in) inward from the wall's right edge.

The five exhibits were configured in a "U" shape such that none of the hotspots overlapped between exhibits. As shown in Fig. 2, the Face exhibit was centered furthest back, and there were two exhibits on either side perpendicular to the Face exhibit. The outer dimensions of the space were 7.32 × 4.88 m (24 × 16 feet), leaving a 7.01 × 4.27 m (23 × 14 foot) interior area for the robot to move around.

Fig. 2. The art gallery contained five interactive Artbotics exhibits: (left to right) Sunflower (yellow background), Vincent (red background), Face (purple), Monkey (green background), and Music (not shown, blue background).

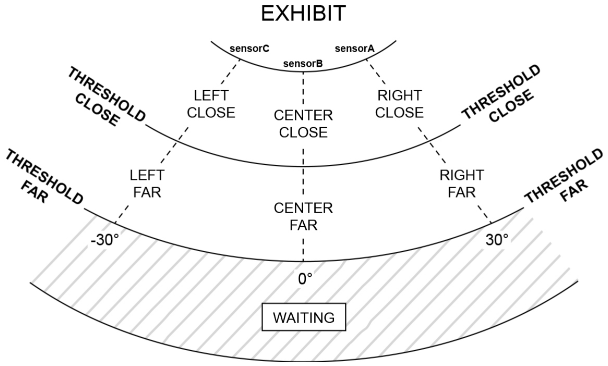


Fig. 3. Three sensors were placed in front of and below each exhibit. The *center* faced straight out while the *left* and *right* were aimed 30°in either direction with respect to the center. The area in front of each sensor is defined as an exhibit hotspot.

4 Telepresence Robot

4.1 Hardware

An augmented VGo Communications telepresence robot was used for this work; our robot design over three iterations is detailed in [81]. The standard VGo robot [86] is shown in Fig. 4d; it measured 0.35 m (14 in) in width and 1.22 m (48 in) in height. The VGo base contained its differential drive system, casters, motor controllers, headlights, bumper sensor, IR cliff and distance sensors, power system, and subwoofer. Other than the subwoofer, the remainder of VGo's audio and video communication system was contained in the robot's "head," held 0.9 m (36 in) above the base by two parallel vertical stalks. The VGo head contained a 15 cm (6 in) video screen, camera, tweeter, and directional microphone array. Using the VGo App software (Fig. 4c), an operator can tilt the camera 180°, which allows the flexibility to be either pointed at the ground while driving or forward during conversation. The standard VGo is teleoperated using the keyboard or mouse cursor with an on-screen widget (not shown).

In addition to necessary modifications for ROS compatibility [74], our augmentations included additional sensing and processing power that enabled us to implement autonomous behaviors (see Section 4.2) and our augmented reality UI (Section 5). A Hokuyo LIDAR and Lenovo laptop were added to the base of our robot to facilitate advanced navigation behaviors. The robot must be able to map, navigate, and move through a remote location commanded by the user. Three additional cameras (two Logitech c910 webcams and one Asus Xtion range camera) were rigidly mounted above the VGo's tilt camera, in a container called the "hat" (see Fig. 4a).

4.2 Movement Behaviors

Users are given high-, mid-, and low-level control of our telepresence robot's movement, as per the third key insight. High-level control is provided by allowing the user to change exhibits. Once at a specified exhibit, mid- and low-level control is provided by allowing the user to adjust his or her vantage point around the exhibit.

We designed two distinct behaviors for the telepresence robot's movements in the gallery to allow users to move in manner similar to in-person visitors [89]. The first corresponds to the high-level control. When the user specifies a change of exhibit, the robot should move such that it turns away from the current exhibit and towards the next exhibit. As it moves towards the next exhibit, the robot should move through the center of the gallery space.

In the second behavior, when the robot is at an exhibit and the user specifies a movement to another vantage point at the same exhibit, the robot should move such that the exhibit (i.e., the object of interest) always remains within the camera's field of view. The VGo robot has a differential drive system with two drive wheels and two passive rear casters, which is similar to a wheelchair. Moving closer to or further from a given exhibit is trivial. Low-level control is restricted forward and backward translation only. Changing our robot's vantage point is more complex otherwise, as translating sideways to the left or right is not possible. In general, changing the robot's vantage point can be completed using a series of commands to turn in place, move forward, and turn in place, or alternatively, perform a point turn by partially turning while backing up and then moving forward while completing the turn. In the first maneuver, the movement and orientation of the robot's base appear "invisible" to the user if the cameras are able to pan to stay focused on the exhibit while the robot adjusts its position. However, if a telepresence robot has a fixed forward facing camera, both maneuvers yield movements that appear contrary to the user's goal (i.e., turn away, back away, respectively); the user's sense of being telepresent may decrease since the non-human telepresence robotic embodiment becomes visible. Given the three hat cameras are fixed (i.e., unable to pan or tilt), we employ the alternative maneuver as our mid-level control, keeping the exhibit always within view.

Like many other planning systems, our navigation software was divided into a global path planner and a local planner [74]. The global planner performed a search over a graph of precomputed waypoints made up of an entry pose, hotspots, alignment poses, and retreat poses for each exhibit. The hotspots are navigation poses that the user can select. Alignment poses and retreat poses are used to transition between the hotspots, while satisfying the movement behavior requirements.

4.2.1 Named Exhibit Poses

Each exhibit has an entry pose *main*, shown in Fig. 5. There are four hotspots – *left*, *center*, *right*, and *info* – from which the robot can view an exhibit. The *left*, *center*, and *right* hotspots correspond to the exhibit's three IR distance sensors, which trigger different behaviors in each exhibit. The *info* hotspot corresponds to the exhibit's placard. The entry pose and each hotspot has a corresponding alignment point, indicated as dotted circles in Fig. 5. Each exhibit also has two retreat poses: *left retreat* and *right retreat*. The retreat poses and alignment poses are positioned such that the exhibit will remain within the view of the robot's camera if the robot is orientated along the vector from the retreat point to the current hotspot. These twelve named exhibit poses were calculated using a location for the base of the exhibit, its orientation in the gallery, and several parameters such as the distance from the hotspot to the exhibit and distance between hotspots and alignment points. Each of the five exhibits had its own corresponding set of name exhibit poses.

4.2.2 Path Planning

There were two basic routing scenarios: intra-exhibit and inter-exhibit. First, the intra-exhibit path planning

(a) (b) (c) (d)

Fig. 4. (a) Margo's three cameras in the hat are stitched together in a single vertical panoramic video stream; our cameras and corresponding portion of the view are shown. **(b–c)** The robot is located at the left hotspot of the Sunflower exhibit. Our interface is shown on the left, and the VGo App on the right. **(d)** The standard VGo robot's 180°tilt camera.

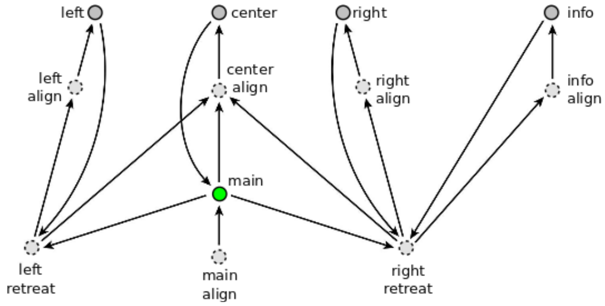


Fig. 5. Twelve named exhibit poses. Connected graph of the movement within an exhibit via its entry pose *main* and hotspots (*left*, *center*, *right*, and *info*).

considered only the graph of waypoints for the current exhibit (Fig. 5). The global planner used a breadth first search to plan a route from the pose closest to the robot's position to the destination hotspot. The local planner would then cause the robot to drive backwards if the next waypoint was behind the robot, and forward if the next waypoint was in front of the robot. Fig. 6 depicts the movement from the exhibit's *center* hotspot to its *right* hotspot. The global planner would first have the robot move backwards to the *right retreat* (see 6, frames 1–4), then move forward through the *right align* pose (frames 5–9), and stop on the *right* hotspot. The two *retreat* poses were positioned such that the robot would always drive backward to these poses and never be at an angle where the exhibit was out of view. We were thus able to obtain the desired behavior of always keeping the exhibit in view while changing hotspots.

The desired inter-exhibit behavior was slightly more complicated. The robot needed to first back away from the current exhibit, then turn and drive forward through the center of the gallery to the next exhibit. Upon nearing the destination exhibit, the robot needed to drive to its *main* entry pose via its alignment point. The global planner accomplished this navigation behavior in two stages. First, it computed the exit path from the closest named pose from the robot's current location to a corresponding retreat pose at the current exhibit. Next, the entry path from the alignment point to the main entry pose of the destination exhibit was computed. The final path was the concatenation of the exit and entry paths.

When the robot was positioned at one exhibit, all other exhibits were available for the user to select. Additionally, the user could select a new exhibit even while the robot was executing its current plan.

5 An Augmented Reality UI

It is not necessary for a user to have a fully immersive experience in a remote environment for effective social interaction. There is a need to bridge the gap between what is needed for effective movement of the robot and what is needed for an effective conversation [75]. As previously stated, the degree to which the user

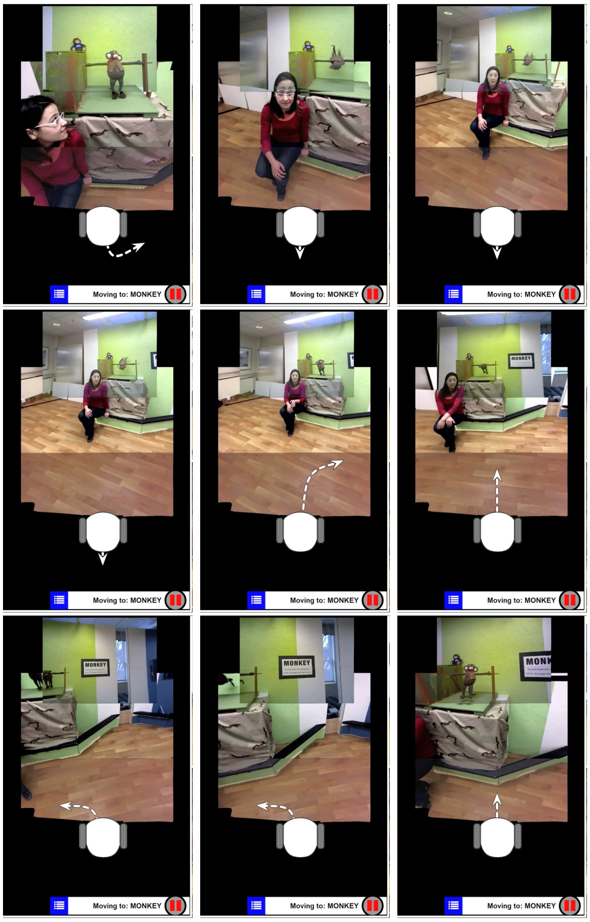


Fig. 6. Intra-exhibit movement behavior. When moving from monkey's *center* hotspot to its *right* hotspot, the robot first backs away to the *right retreat* pose and then moves forward to its alignment point *right align*, and finally arrives at the hotspot *right*. The Monkey exhibit remains in the robot's field of view for the duration of this transition.

feels telepresent with the interactant in the remote environment and vice versa is dependent upon the quality of the user's HRI (Fig. 1a) and the interactant's (Fig. 1b). Our goal was to design an "invisible-to-use" [69] telepresence robot UI, which would allow the user to focus on exploring the exhibits in a remote art gallery without drawing his or her attention to the mechanics of the UI itself or the telepresence robot embodiment.

It is not always obvious to a user as to *how* to operate a robot system beyond low-level directives (i.e., forward, backward, left, right, stop). We believe that the discovery of the robot's autonomous capabilities should be facilitated by the HRI interface presentation and system feedback. It is unrealistic to present every possible interaction in a single interface, as the user would be overwhelmed [51]. Human mobility follows a power-law distribution (e.g., by vehicle [25], walking [57], activities of daily living at home [1]); we posit that users' intentions and, consequently, directives will also follow this pattern, which we observed in the third key insight in our formative assessments [80].

Nielsen's 10 usability heuristics for designing user interfaces are the gold standard in HCI [51]. However, Bergman and Johnson [5] note that the majority of HCI design focuses on the typical user: that is, users without physical or cognitive impairments. Further, they state that in Nielsen's book *Usability Engineering* [51], "disabilities are only mentioned in a few brief sentences in the entire book" [5]. Further, they state that accessibility "has not been generally recognized in standards texts or in work that is not explicitly focused on disability issues. Users with disabilities are simply not 'on the radar screen' of mainstream HCI" [5]. Fortunately, there have been several groups focusing on disability issues and accessibility in HCI. The W3C group [87, 88] and the Nielsen Norman Group [56] have focused on web accessibility, and Bergman and Johnson [5] have addressed these issues in computing applications. When designing for people with disabilities, we must consider a wide range of disabilities and abilities pertaining to vision, motor, behavioral, perception, cognition, and social skills; Vanderheiden and Vanderheiden [83] provide an introduction to disabilities.

We surveyed the design guidelines and heuristics that have been developed for both general-purpose UIs and interfaces for assistive technologies. These guidelines are provided in Appendix A in Tables 9 through 13; note that the guidelines are a selected subset, and concepts may be repeated with slightly different wording or emphasis from their authors. Specifically, we drew from HCI interface design guidelines and selected 8 of Nielsen's usability heuristics [51, 53] (Table 9). We also drew from guidelines for accessible web design and selected 8 Web Content Accessibility guidelines [87] by W3C (Table 10). We also selected 16 of Kurniawan and Zaphiris's guidelines for older adults [42] (Table 11). Finally, we selected 4 guidelines from Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [83] (Tables 12 and 13).

When designing our interface, we found the most relevant guidance to be the following:

1. Ensure a match between the system and the real world (N-2),
2. Provide visibility of system status (N-1),
3. Prevent errors (N-5),
4. Facilitate recognition rather than requiring recall (N-6), and

5. Aid perception (W-1).

The first four are Nielsen's usability heuristics, as denoted by their IDs which correspond to the first, second, fifth, and sixth rows, respectively, in Table 9 in Appendix A. The fifth is by W3C (first row in Table 10). We highlight a number of these guiding heuristics and their manifestation in our augmented reality telepresence robot UI.1 It should be noted that while our interface was designed for a remote art gallery scenario, these design principles can be applied to any telepresence robot UI.

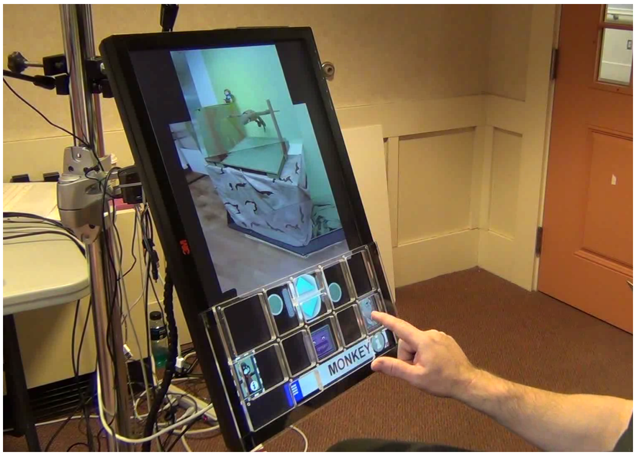


Fig. 7. A custom clear acrylic keyguard is shown over the lower third of the interface with cutouts for the exhibit buttons, step buttons, area around the robot base icon, menu button, and iButton. The 55.9 cm (22 in) monitor (shown with keyguard) was mounted on an Ergotron cart with an adjustable swing arm, which allowed it to be placed in a manner that was comfortable for the user to see and manipulate.

5.1 Interface Overview

Our interface enabled the users to navigate through an art gallery.2 A telepresence robot utilizes a video stream to provide "virtual presence" to a user; as such, our interface provided a video-centric, first person view. User controls corresponding to the semi-autonomous robot navigation behaviors were integrated into the interface design. The interface was touch based, and users would touch and release elements to move the robot in the remote environment.

Our interface was designed for a 55.9 cm (22 in) touchscreen monitor in portrait orientation with a resolution of 1050 × 1680 pixels (Fig. 7). Given the large size of the monitor, all buttons were placed in the bottom half of the touchscreen to facilitate our target population's access. The 55.9 cm (22 in) monitor used was mounted on an Ergotron cart with an adjustable swing arm, which allowed it to be placed in a manner that was comfortable for the user to see and manipulate. For users with limited dexterity in their hands, we designed an optional clear acrylic keyguard3with cutouts for the elements of our interface. The interface also supported any access method that emulated a mouse cursor (e.g., computer mouse, RollerBall2 Joystick).

Our UI was implemented as a web-based application (HTML5 and javascript). A web-based application can run cross platform, that is to say, it can run on any modern operating system and browser combination. A user can control the robot from any computer available to them without needing to install any application specific software ahead of time [14]. Communication with the robot was provided through roslibjs [70]. This allowed the interface to communicate with the robot bidirectionally, both receiving system updates and sending commands to be executed.

5.2 Heuristics and Guidelines At Work

Our telepresence user interface is rife with subtle design choices, which balanced accessibility and usability based on our research [73, 76, 77]. To create an effectively "invisible-to-use" interface [69], our design approach was to use a minimalist aesthetic (N-8) and

---

[1] For clarity, an ID has the form of <author>-<row_entry>. **N-#** refers to Nielsen's usability heuristics [51, 53] listed in Table 9. **W-#** refers to the Web Content Accessibility guidelines [87] by W3C listed in Table 10. **KZ-#** refers to Kurniawan and Zaphiris's guidelines for older adults [42] listed in Table 11. Finally, **V-#** refers to Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [83] listed in Tables 12 and 13. Tables 9 through 13 are given in Appendix A.

[2] http://www.youtube.com/watch?v=rd5pADte4bk

[3] A keyguard is a plastic or metal plate that sits above the keys on a standard keyboard. It is specially designed for computer users with limited motor skills, as it enables people with tremors or difficulty with finger isolation to type more accurately. A keyguard has holes directly above the keys, or, in this case, the regions of the touchscreen on which the interface places buttons (Fig. 7). The user is able to stabilize his or her hand on the keyboard surface while typing. The holes in the keyguard's surface also make using other alternative access tools such as mouth sticks, head sticks, and other pointers easier to use.

simple layout (V-4) with overlays on the camera view like an augmented reality technique. Our interface was displayed fullscreen (KZ-6), and we opted for a single point "touch and release" interaction, although the screen supported multiple touch points and gestures. Since a touchscreen has no physical feedback itself, both auditory cues and shadowing were employed to provide feedback to the user regarding their touch selection (KZ-2, W-1b). When the user released, for example an exhibit button, the robot immediately began moving without an additional confirmation or cancellation dialog box. While violating Nielsen's heuristic for user control and freedom (N-3), our interface employs Vanderheiden and Vanderheiden's guideline which suggests to "use selection techniques where the person need only make a single, simple, non-time-dependent movement to select" (V-4) [83]. This design choice supports accessibility and maintains the user's focus and sense of being telepresent in the remote environment.

5.2.1 Match Between System and the Real World

The two primary examples of this Nielsen's heuristic (N-2) were the camera view and the to-scale robot base icon. These two elements were the basis of our interface.

Camera View. Video information is critical for telepresence robots to allow user conversation with interactants and for robot navigation. Design of the interface must account for both the need to drive the robot system and the need to be present in the remote environment for conversations, just as a user would have if actually walking through the environment. The robot should have a wide field of view (FOV) both horizontally and vertically [75]. Chen et al. [11] found that a 45 degree FOV was regarded as too small for remotely driving a vehicle, and it is an open research question as to what a suitable FOV for operating a social telepresence robot is.

To give the user a wider field of view, we created a single camera view from a vertical panoramic video stream by stitching together the individual output of the three hat cameras (Fig. 4a). The images from each of the Logitech c910 webcams were 864 × 480 px, and the Asus Xtion was 640 × 480 px. The individual video streams were rectified to remove lens distortion [31, 59] prior to tiling the images together; the resulting video stream displayed on the interface averaged 1.6 MB/s compressed at 11 frames per second (V-4). The parameters for the two Logitech c910 webcams were tuned to match the Asus Xtion, which included the color temperature, white balance, contrast, and saturation. This vertical panoramic video, herein known as the camera view, is the foundation of our interface and accounts for 63% of the screen (KZ-3, KZ-9).

Fig. 4b shows our interface with its vertical panoramic camera view and the VGo App (4c), which also has a video-centric interface. Looking at a snapshot of the two interfaces, it is difficult to determine the robot's location relative to the Sunflower exhibit from the VGo App. Additionally, the Sunflower exhibit is not fully visible without additionally manipulating the robot's tilt camera, which is the video source for the VGo App.

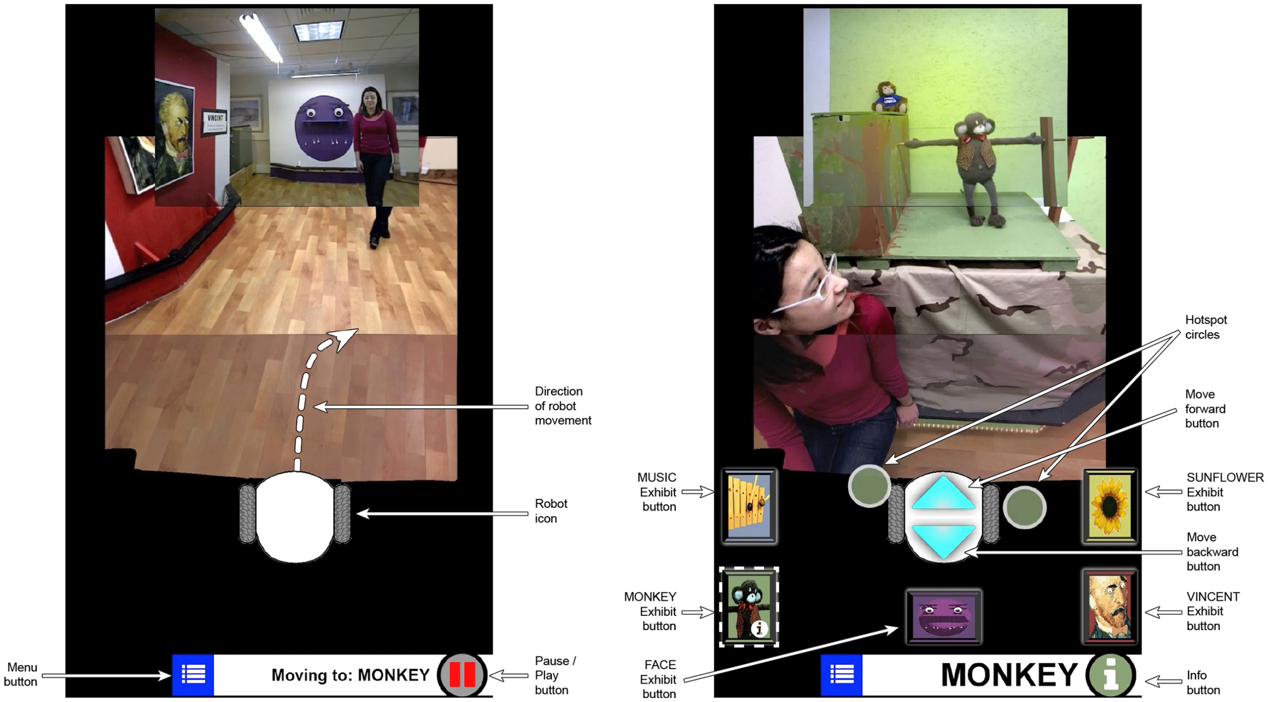
Scale Robot Base Icon. It is not yet possible for a user to experience the proprioception and kinesthesia of a telepresence robot embodiment [10]. However, users may still achieve a physical real-world frame of reference [50] in the remote environment through a sense of the location of parts of the robot [7].

To reinstate the proper sense of scale, we designed a to-scale icon of the robot's base. The contour of the robot's base provided an outline for the icon and included two wheels, one on either side. This depiction is an exaggeration of the robot's drive wheels, which were recessed under the base and therefore not visible. While the width may be exaggerated, it does, however, provide the user a grounded frame of reference (KZ-4).

5.2.2 Visibility of System Status

At times, it may be difficult to determine the status of a remote system. This may be due to lag inherent in data communications or an inconsistency between the user's expectations and the system state. In addition to seeing the robot's video update, we designed three levels of status indicators into the interface (N-1, V-2, V-4).

First, the robot base icon directly under the camera view provides two indicators of system status. The robot base icon is implemented as one of three interchangeable animated gifs. When the robot was not driving, the robot base icon's wheels were displayed as stopped. When the robot was driving forward or backing up, the wheels were animated to spin forward and backward, respectively. In addition to the rotation of the wheels, the robot base icon anchored the Animated Vector Indicator (AVI). Based on the direction of the local planner, the interface displayed a simplified version of the path as a dotted white arrow outlined in black. The user interfaces of several other telepresence robots, including the Giraff, Texai, and Beam robots, employ a



(a) (b)

Fig. 8. Our augmented reality graphical user interface design for visiting an art gallery via a telepresence robot. **(a)** To scale robot base icon overlaid on the bottom of the robot's video stream; the animated vector indicator (AVI, white dotted arrow) shows the robot's direction of travel towards selected exhibit. **(b)** Once at the selected exhibit, exhibit buttons are shown around the robot base icon, providing the user with high-level control over the robot's movement. Four hotspots can be overlaid on the robot's video stream indicate alternative view points to see the exhibit from the *left*, *center*, *right* and the exhibit information *info* via the far right hotspot. In this image, the robot is located at the *center* hotspot of the Monkey exhibit; adjacent *left* and *right* hotspots appear on either side of the robot base icon. Additionally, the forward and backward step buttons (i.e., two cyan triangle buttons) appear overlaid on robot base icon, providing low-level control.

similar feedback technique by showing a vector originating from the bottom center of the video on the interface [13, 24, 26, 28, 67]. A storyboard depicting the robot's movement is shown in Fig. 6. The AVI was designed to inform the user as to which direction the robot would drive to reach the next waypoint – forward or backward, and moving straight or turning to its left or right. For the backward AVI, there were three representations: one to indicate the robot moving straight back (see Fig. 6, frames 2, 3, and 4), one to indicate backing up to the right (see frame 1), and another backing up to the left. There were 4 magnitudes for forward translation to the right or to the left (>1.5 m, 1.2 m, 0.8 m, and 0.01 m) (see frames 5, 7, and 8). There were 3 magnitudes for straight forward motions (>1.5 m, 0.8 m, and 0.01 m) (see frames 6 and 9). Finally, the UI changed the AVI when the robot was turning in place right or left without any forward or backward translation; Fig. 9.

Fig. 9. Left and right rotation animated vector indicators (AVI), respectively.

Second, the iButton changes to relevant robot control as the user interacts with the interface (V-4). When the robot was moving, the iButton displayed a red Pause symbol in a grey circle (shown in Fig. 8a). Touching the iButton allowed the user to pause the robot's movement (N-3). While paused, the iButton toggled states to display a green Play button in the same grey circle. Touching the iButton again allowed the user the ability to resume their current path (N-3). This metaphor is consistent with how many media players work such as iPods, DVD players, and DVRs (N-4). When the robot was not in motion (e.g. when it arrived at an exhibit or hotspot), the background color of the circle reflected the color of the exhibit (shown in Fig. 8b). The iButton displayed an iconic white "i" in the center to indicate to the user that they may click for more information about the exhibit (KZ-3, KZ-4, N-4, W-1a).

Lastly, the title bar at the bottom right side of the screen gave the user explicit information about the status of the system (KZ-7, V-1, V-2, V-4, W-3a). While moving between the exhibits, the title bar displayed the text "Moving to **<exhibit_name>**", allowing the user to verify that their previous command had been accepted and the robot was moving the the correct exhibit; for example, as shown in Figs. 6 and 8a. When the robot reached the selected exhibit, the title text showed "**<exhibit_name>**" at full height (4.75 em) (Fig. 8b), and the interface played a short audioclip of an arrival sound (W-1b).

5.2.3 Recognition Rather than Recall

One of the primary design objectives for the interface was to be simple to use by the target population without the need for lengthy training exercises. In addition to this Nielsen heuristic (N-6), we employed the W3C's distinguishability guideline (W-1b), which stated, "Make it easier for users to see and hear content including separating foreground from background." To that end, much of the interface relied on the user's ability to recognize interaction points (KZ-12, KZ-14, N-6, V-1, W-1b). All of the buttons in the interface were given a visual affordance (described below) and auditory feedback. The auditory feedback for every button was the same. A downward press event (either touch or click depending on the user's access method) caused a sound to be played which was similar to the press down of a mouse button. Upon releasing a button, if the user's finger, pointing device, or mouse cursor were within the bounds of the same button, the action was selected and a sound similar to releasing a mouse button was played. If the release event occurred anywhere other than on the same button, it would be considered a cancel and the sound was not played. The sounds did not play if the user clicked on all areas outside of the buttons.

Movement. Users wanted multiple levels of control over the robot's movement. Based on these levels of control, the user interface was broken down into three discrete areas of interaction: global navigation, local navigation, and low-level control. All navigation buttons including the menu buttons, exhibit buttons, and hotspots were color matched to the exhibit that they represented (V-1, V-4). The visual affordances [23, 55] were different for each type of navigation button. We utilized the Gestalt principle of similarity [38].

To provide global navigation, five buttons were positioned around the robot base icon (Fig. 8b). Each button corresponded to one exhibit. These buttons provided the user the ability to navigate the gallery globally; that is, the user could send the robot to a new exhibit. The exhibit buttons were rectangular in shape (130 × 180 px, 36 × 51 mm), and filled with an icon of the corresponding picture of the exhibit itself (KZ-3 KZ-15, W-1a). If the user wanted to go to the "Vincent" exhibit, for example, he or she would be able to recognize both the image and color of the button. Buttons were also given ample room around them so that there was no confusion as to which button the user was targeting (V-3).

Exhibit buttons were given a three dimensional raised appearance, which was achieved using a grey and black double outset border. Additionally, a white translucent shadow acted as highlighting with the robot base icon as the origin of the light source. When an exhibit button received a downpress event, it was obscured by a semi-transparent grey layer, and the border changed to an inset one to give the button the appearance of deflection. In addition to the exhibit buttons located on the main screen, the corresponding buttons appear in a menu which can be opened at any time (Fig. 8a).

Every exhibit had four hotspots, positioned similarly around the exhibit and the wall information placard (N-4). Once at an exhibit, local navigation was represented as pulsing rings around the robot base icon. These pulsing rings represented the hotspots, which were real-world wayposes that the user could select. The hotspots were circular buttons (120 × 120 px, 33.6 × 33.6 mm) with a slow pulsing effect; they changed size from 50% to 100% at a rate of 1.5 seconds [82]. The hotspots were outlined with a 10 pixel white border, and the color of the center matched the exhibit color. The hotspots could appear overlaid on the camera view or in the space between the robot base icon and the exhibit buttons. To prevent occluding the camera view, the hotspots were translucent. When a downpress event occurred, the pulsing ceased and the opaque hotspot displayed at its maximum size until it was released.

When the robot arrived at an exhibit's *main* entry pose, coordinate frames were created for each of its four hotspots: *left*, *center*, *right*, and *info*. The UI converted the hotspots from the physical real-world coordinates (x, y in meters) to be displayed on the user interface (x, y in pixels). It should be noted that the visualizations of the hotspots in the user interface were an approximation of the hotspots' physical locations. It was necessary to visualize the hotspots such that (1) no hotspot overlapped the robot base icon, and (2) adjacent hotspots did not overlap each other. When determining overlap, we required an additional empty border of 30 px. The visualizations of the hotspots were pushed outside the robot base icon. Overlapping hotspots were iteratively separated by moving both equally from the midpoint of the line between their centers. If the robot was at a hotspot, only the adjacent hotspots were drawn. In Fig. 8b, the robot was located at the center hotspot, thus only the left and right hotspots were drawn.

Finally, low-level control of the robot was provided by two cyan arrow buttons, herein referred to as "step buttons" (Fig. 8b). The step buttons appeared on the robot base icon when the robot was positioned at any of the four exhibit hotspots. The step buttons allowed the robot to be moved forward and backward a maximum of two times each from the hotspot. Their default color was bright cyan and changed to dark cyan with a grey gradient when pressed, a similar shadowing technique as the exhibit buttons.

Icons. Additionally, we leveraged iconicity in several places. At each exhibit, there were three ways to access its description (KZ-16, N-3). Rather than remove specific elements from the screen and potentially confuse the user, for example, the actions corresponding to the exhibit button and iButton for the current exhibit were modified (V-4). The iButton also matched in color and featured a contrasting iconic "i" at its center (Fig. 8b). The exhibit button featured a dashed white outline and an "i" in a white circle (i.e., the inversion of the white info state of the iButton). The combination of colors and icons in this respect was chosen to provide the user the ability to recognize the button's functionality (KZ-12, N-4, N-6, V-4). Touching either of these buttons showed the title and short description of the exhibit on screen (KZ-7, V-1, V-2, V-4, W-3a); this exhibit information pop-up remained in the foreground until dismissed by the user when he or she was finished reading its contents (KZ-11, W-2a). Finally, the user could select the right-most hotspot in order to drive the robot to the description written on the gallery wall, to the right of each exhibit.

The menu button also utilized an icon: a list with four lines and dotted bullets. At any time, the user could open the exhibit menu by selecting the menu button; it was always visible on the interface (Fig. 8). If the robot was already moving, opening the exhibit menu will preempt the robot's movement to its current goal. The user can then pick a new exhibit and continue on to it (N-3).

5.2.4 Aid in Perception

There are several techniques to aid perception when viewing a user interface. First is to minimize the number of elements displayed (N-8). Each additional unit of information displayed increases a user's cognitive processing time (N-8). Thus, only the relevant elements should be shown. There is already a large amount of information in the camera view for a user to process since our interface is video-centric and the camera view accounts for 63% of the screen. We therefore minimized the number of user control buttons (V-2). At most, there are 11 buttons on the screen when the robot is at an exhibit. When the robot was moving, the exhibit buttons, hotspots, and step buttons were disabled. Rather than showing these disabled buttons on screen, buttons that were not active were hidden (V-2). Shown in Fig. 8a, only two buttons were displayed when the robot was moving (i.e., menu button and pause), which both caused the robot to momentarily stop (described further in the following section). Additionally, all interactions with the robot were accessible at the top level of the interface.

In order to create an interface usable for our target population, buttons needed to be large and spread out (KZ-1, V-1, V-3). Based on a Fitts' Law comparison, Micire [48] found that widgets on a touchscreen should have a minimum size of 30 mm. All buttons in our interface were a minimum of 100 × 100 pixels in size, which exceed Micire's findings.

Gestalt principles of grouping [38] were also leveraged as there were three interaction areas regarding the robot's navigation (V-2). The five exhibit buttons were placed in a "U" around the robot base icon; additionally, the layout of the exhibit buttons did not change and the exhibit menu utilized this same left to right ordering. The hotspots were positioned around the robot base icon emanating from it. The forward and backward step buttons were overlaid on the robot base icon (V-4), which also adhered to Shneiderman's theory of direct manipulation [65]. The shape of the buttons corresponded to functional groupings (V-4); exhibit buttons were rectangular, hotspots circular, and step buttons triangular.

Coloring and contrast were used to make the buttons highly visible [56, 83] and emphasize groupings [38]. As described previously, each exhibit had a background color, and all buttons relating to a particular exhibit utilized its color. The colors chosen were distinct and able to be perceived by users with red-green color blindness (KZ-14, V-1). The exhibit colors were arranged for maximum pairwise contrast. The information symbols were highly contrasted with the background colors of the exhibits. Although translucent, the hotspots' white borders created contrast against the camera view.

Several interface elements employed a static color profile: the interface's black background (KZ-13) and the robot base icon. The blue of the menu button and cyan of the step buttons were not used elsewhere in our interface.

Finally, we carefully designed motion cues into the interface. As previously discussed, the three motions of the updating camera view, AVI, and robot base icon tire rotation work harmoniously to show the robot's movement. We also utilized motion to draw the user's attention to the hotspots. If the user directed the robot to a new exhibit, the hotspots corresponding to that exhibit did not previously appear on the interface. Once the robot stopped, the hotspots appeared and slowly pulsed at a rate of 1.5 seconds. It should be noted that only one set of motion cues were used at a time. The perceived motion of the hotspots was only present when the robot was stopped.

5.2.5 Error Prevention

Our semi-autonomous navigation behaviors for moving between exhibits and within an exhibit can be likened to human proprioception and kinesthesia. The user is therefore free from the details of robot navigation and can focus on the primary communication task or exploring the remote environment. This type of assisted navigation control can increase telepresence, as the driving task is made easier (N-5), yet the user must still pay attention to the environment around the robot. The user had limited low-level control and could move the robot two steps forwards or backward from any exhibit hotspot.

In the event that the user's attention shifted away from the camera view as the foreground (e.g., by the opening of the exhibit menu), the robot and interface entered the pause state implicitly (N-5); that is, the robot halted. If the user selected a new destination exhibit, the menu closed and the robot changed course to the new exhibit and continued driving. If the menu was closed (i.e., no exhibit selection made), the robot and interface remained paused if the user had first explicitly paused the robot prior to opening the exhibit menu (KZ-16, N-3). Otherwise, the robot resumed its prior course if the pause was triggered implicitly (KZ-16, N-3). Pausing the robot while the menu was open allowed the user as much time as they needed to make a decision on where to go or to resume (KZ-11, W-2a).

6 Case Study: Exploring an Art Gallery

We believe that the user's understanding of a robot's autonomous capabilities should be facilitated by the HRI interface presentation and system feedback. We have developed an alternative augmented-reality graphical user interface that provides cognitive support for our target audience. Simple language and familiar real world analogies may allow robot operators to recognize how to use the interface rather than having to recall how to use it from training and/or their own experience [52]. Hints about the robot's autonomous navigation capabilities and the robot's local and global environmental knowledge have been overlaid on the robot's video. We conducted a usability case study with four users from our target audience to evaluate our end-to-end system. The goal for this case study was to understand what portions of our interface were used, how, and in what situation(s). We posit that each of our four participants would be able to have an interaction that was both highly interactive and personal and made the person feel as if he or she were actively present.

6.1 Experimental Design

In this study, four participants took the role of a telepresence robot operator and used our modified VGo robot and alternative user interface to explore a mock art gallery (remote from the participant's location). The gallery contained five robotic art exhibits: Sunflower, Vincent, Face, Monkey, and Music (Fig. 2, p. 4). Each participant visited the gallery twice using the robot and once in-person. During the first in-robot visit, the participant was alone in the gallery and explored three exhibits. For the participant's second in-robot visit, there was another person in the gallery, and two new exhibits were added. Finally, the participant visited the gallery in-person, accompanied by members of the research team.

6.1.1 Setup

A Windows 7 Dell XPS laptop ran the VGo Desktop App and our interface in the Chrome browser. The XPS laptop was connected to a 55.9 cm (22 in) 3M multi-touch monitor. We leveraged the VGo App for its bi-directional audio and video stream from the XPS laptop to the robot. The VGo App was launched and run in the background; it was configured with a maximum video bitrate of 384 kbps. A Microsoft LifeChat LX-3000 USB headset was set as its audio input and output device, and a Logitech c920 USB webcam as its video capture device. Participants operated the telepresence robot using our interface instead of the VGo App.

6.1.2 Recruitment and Participants

Inclusion/exclusion criteria. Potential participants selected for this study were members of the Crotched Mountain Community, including students at the school, inpatient clients from the Brain Injury Center, and participants in the residential program. They were between the ages of 7 to 75 and had a condition that significantly limited their ability to travel and maintain contact with important individuals in their "regular" environment. Their medical conditions included disabilities such as Cerebral Palsy, Spina Bifida, Spinal Cord Injury, Traumatic Brain Injury, or other conditions. We required that the participants were able to speak English fluently, but not necessarily as a native speaker. People with blindness, severe cognitive challenges, low arousal levels, or other conditions were unlikely to benefit from using a telepresence robot and thus were not included in the study. Students or clients with severe cognitive challenges were unlikely to have the conceptual ability to understand that the telepresence robot was a representation of themselves as opposed to a TV show or video game.

Participants. We recruited four native English speakers with cognitive and motor impairments who reside at the Crotched Mountain Rehabilitation Center (CMRC) and who have a compelling need to use telepresence robots as a means for social engagement beyond the CMRC facility. All participants were their own legal guardians and gave their consent to participate in this study. All four participants had intact literacy. All four participants used a manual wheelchair: one required the assistance of a caregiver (P3), and three were independently able to propel their wheelchairs (P1, P2, and P4).

Two participants had spinal cord injury; P2's injury occurred one year prior to this study, and P4's was less than one year prior. P2 was left hand dominant and able to use both of her hands dexterously; she used the keyguard for the first in-robot visitation to the gallery only, and it was removed for the second visit. P2 had typical cognition and functional vision corrected with glasses. P2 noted her technical background and worked with computer-aided drafting command line software.

P4 had typical cognition and no significant visual challenges. He was right hand dominant, and used the keyguard for both in-robot visitations to the gallery. P4 had the most experience with video games and reported playing for 20 hours per week.

Two participants had cerebrovascular accidents (strokes). P1, age 53, was 2 years post-stroke, and P3, age 46, was 1.5 years. P1 and P3 used the keyguard for both in-robot visitations to the gallery. P1 had mild cognitive challenges regarding memory. P1 had moderate visual neglect in his right eye and had difficulty seeing the lower right side of the interface. P3 had mild cognitive challenges regarding memory and problem solving. P3 had moderate visual loss in his left eye, and was right hand dominant. P1 and P3 wore glasses.

Both P2 and P4 were familiar with the concept of a telepresence robot, noting the episode of "Big Bang Theory" television sitcom featuring the Texai robot [68]. During the second gallery visitation, P2 discussed with the confederate her increased use of video conferencing (i.e., Skype) since her arrival at CMRC. P1 had participated in a prior study [79] and therefore had prior knowledge about our robot system.

6.1.3 Procedure

The total time for this study was approximately 4 hours per participant. The study was split into two sessions to maximize their attention.4 Participants were compensated $150 for their time at the study's completion.

In the first session, the experimenters obtained the participant's consent and administered an interview (demographic information and prior experiences) and a training exercise (T1). The experimenters provided a description of the robot and its custom user interface; training entailed (1) moving robot forwards and backwards near an exhibit, (2) viewing an exhibit from more

---

[4] The full procedure details can be found in [74], including the training scripts, the confederate's role, and the task descriptions.

than one viewpoint, and (3) moving from one exhibit to another. A short interview followed the training in which we asked the participant what was easy to do with the user interface, what was hard, what he or she would change about the user interface, and how.

Then the participant entered the gallery using the telepresence robot and explored for up to 20 minutes. It should be noted that the participant (via the telepresence robot) was alone in the gallery. The gallery contained three working exhibits (Vincent, Face, and Monkey); the other exhibits were covered. The experimenters instructed the participant to start at a specific exhibit. In the post-session interview, we asked the participant to describe his or her experience including his or her favorite and least favorite exhibits.

There was a 7- to 10-day break before the second session. The participant was again provided a description of the robot and its custom user interface, and given an opportunity to practice moving the robot in the remote location (T2).

The participant visited the gallery using telepresence robot a second time and explored for up to 45 minutes. The increased time limit was due to two changes in this second visitation. First, the Sunflower and Music exhibits were incorporated, for a total of five exhibits. Second, a person (a confederate) was physically present in the gallery and engaged the participant in conversation about the exhibits. The confederate was an occupational therapy intern at Crotched Mountain Rehabilitation Center who was familiar to the participants. She had prior interactions with P1, P2, and P4 outside of this study. The participant was instructed to start at a specific exhibit, which was his or her least favorite exhibit as noted from the first visitation. The confederate was already interacting with this exhibit when the participant entered the gallery using the robot. To foster conversation and movement around the gallery, the confederate offered that this exhibit was her favorite, later asked what the participant's favorite exhibit was, and informed the participant of the two new exhibits. The confederate excused herself from the gallery after visiting all five of the exhibits with the participant or approximately 30 minutes into the session (15 minutes remaining). In the post-session interview, we asked the participant to describe his or her experience.

Finally, we accompanied the participant into the gallery (in-person) and asked the participant to describe his or her experience in-person vs. via a telepresence robot. The experimenters debriefed the participant and answered questions about the study.

6.1.4 Data Collection

In addition to semi-structured interviews, we video and audio recorded the sessions (both of the gallery and of the participant interacting with our user interface). Recordings of the interviews and the interaction between the participant and confederate in the second gallery visit were transcribed using CastingWords [9]. We also noted our observations during the sessions. Collected data included the total time spent in the gallery (i.e., task completion), the number of exhibits visited, the number of UI interactions (e.g., button presses, clicks), and the robot's movement trajectories.

We categorized the transcribed utterances based on the content of the conversation (e.g., discussion of an exhibit, discussion of an aspect related to the exhibit, off-topic conversation; see Table 1). Cohen's kappa for inter-rater reliability between two raters was κ=0.73 excluding chance; both raters fully reviewed the categorization of the utterances.

6.2 Visiting the Gallery

6.2.1 Session 1

All four participants visited the three exhibits (Vincent, Face, and Monkey) during their first session. As shown in Table 2, P2 and P4 interacted with the exhibits more thoroughly than P1 and P3; both P2 and P4 revisited all of the exhibits in the same order that they had originally directed the robot through the gallery, counterclockwise and clockwise, respectively.

During his first session, P1 spent approximately 17 minutes (85.4%) viewing the exhibits, and 3 minutes (14.6%) moving from one exhibit to the next. P1 revisited two of the exhibits after his initial pass. He directed the robot to 6 of the 12 hotspots (Fig. 10a); however, his movement at each exhibit was uneven. The experimenter instructed him to begin at Vincent; he viewed the exhibit from only one hotspot (left). Then, P1 interacted with the Monkey from three hotspots; he briefly visited the right hotspot, and spent time moving the robot back and forth between the center and left hotspots. At the Face exhibit, he directed the robot to the two left hotspots both times.

P2 visited the gallery for 17m 15s, spending 79.5% of her time viewing the exhibits and 20.5% of her time moving between exhibits. As shown in Table 2, P2 revisited each of the three exhibits after her first pass during her first in-robot visitation to the gallery. She visited 11

Table 1. Coding categories for conversation between confederate and participant during participant's second gallery visit (κ=0.73)

Movement	Related to movement within or between exhibit (e.g., stand here, switch places, negotiating next exhibit)
System Visibility	Related to user interface visibility, audio (e.g., "Can you see me?" or how to do something on the interface)
About Exhibit	Statement directly related to exhibit (e.g., color, content, material(s), shape, interaction, motion, sound, information sign)
• Neutral statement	Neutral statement about the exhibit (e.g., factual description, questions, observations)
• Positive statement	Positive statement about the exhibit (e.g., "I like..." or "it's [fun, cool, powerful, creative, nice, neat, pretty]")
• Negative statement	Negative statement about the exhibit (e.g., "I dislike..." or "it's ugly")
• Neutral judgement	Neutral judgement about the other person's statement (e.g., "oh," "ok," "really")
• Positive judgement	Positive judgement about the other person's statement (e.g., "I think you're right...")
• Negative judgement	Negative judgement about the other person's statement (e.g., "I think you're wrong...")
Related to Exhibit	Sharing experience related to the content of the exhibit (i.e., giving new information, following this topic of conversation)
• Initiator	Participant or confederate; who initiated sharing of the experience, providing new information, etc.?
Unrelated to Exhibit	Off topic statement; not related to previous categories (i.e., exhibit directly, robot, user interface) (e.g., greeting, exit, banter, talking to self)
Miscommunications	Explicit break downs in the conversation between confederate and participant
• Simultaneous start	Confederate and participant start speaking at the same time
• Interruption	While confederate or participant is speaking, the other verbally interjects (e.g., "yeah," "mm-hmm")
No Code	There is no appropriate code for this utterance

Table 2. Summary of first in-robot gallery visitation (S1)

P#	Time	Time spent at exhibit in order of visitation					
P1	20m 29s	Vincent, 4m	Monkey, 2m 35s	Face, 5m 36s	Monkey, 3m 8s	Face, 2m 11s	
P2	17m 15s	Face, 2m 19s	Vincent, 1m 51s	Monkey 3m 2s	Face, 2m 11.5s	Vincent, 2m 17s	Monkey 1m 59s
P3	20m 45s	Monkey, 1m 27s	Vincent, 3m 26s	Face, 10m	Monkey, 3m 15s		
P4	18m 8s	Vincent, 2m 32s	Face, 2m 8s	Monkey, 4m 12s	Vincent, 1m 45s	Face, 2m 5s	Monkey, 1m 20s

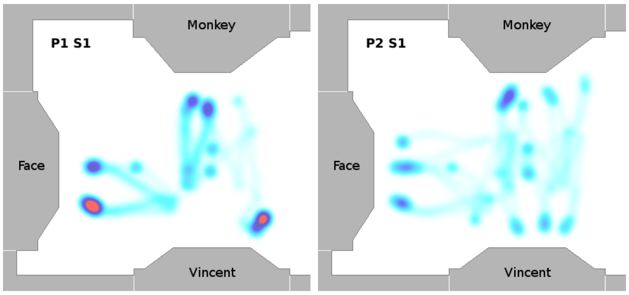
of the 12 hotspots corresponding to the three exhibits (Fig. 10b), only bypassing the information sign on the Face exhibit. P2 noted in her interview that the information on Vincent's sign ("The closer you get, the goofier he is") prompted her to move closer to the exhibit. She visited Face prior to Vincent; P2 visited the Monkey exhibit before making the same loop a second time.

P3 explored the gallery for the full amount of time, and spent 12.6% of his time moving between exhibits. He visited each once and returned to the first exhibit. P3 spent the most time at the Face exhibit and viewed it from two hotspots (Fig. 10c); he viewed the Monkey and Vincent primarily from a single hotspot.

Like P2, P4 also revisited each of the exhibits a second time during his first session (Table 2); he visited the gallery for 18 minutes and 8 seconds and spent 22.7% of his time moving between exhibits. P4 noted that when visiting his third exhibit, he selected the rightmost hotspot and was able to read the information sign directly in front of him. On his second pass, he moved to the side of the exhibit that he had spent the least amount of time (i.e., Vincent and Face right, Monkey left). P4 thoroughly visited all 12 hotspots (Fig. 10d). Like P2, he revisited the exhibits in the same order: Vincent, Face, and Monkey.

6.2.2 Session 2

All four participants visited the five exhibits during their second session; P1 and P3 explored the gallery for the full 45 minutes allocated (Table 3). All were able to engage in conversation with the confederate about the exhibits and other topics; none of the participants declined to converse with the confederate. The conversation between P2 and the confederate had the most utterances. The fewest utterances occurred between P4 and the confederate (n_{P4}=98), and P4 was the quickest



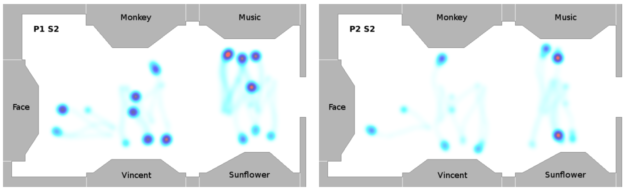
(a) P1 (b) P2

(c) P3 (d) P4

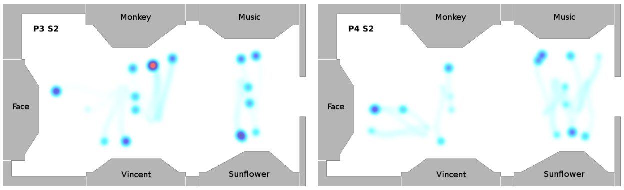
Fig. 10. Heatmap of the robot's movements within each exhibit area during the first in-robot gallery visitation (S1) for each participant; movement between exhibits is not shown. In each image, the Face exhibit is the leftmost, then the Monkey and Vincent exhibits moving in the clockwise direction. The heatmaps depict the amount of time the robot spent moving within each exhibit area, specifically from an exhibit's *main* entry pose and its 4 hotspots (*left*, *center*, *right*, and *info*).

to exit the gallery (t_{P4}=14m 43s). The conversations between the confederate and P1 and P3 had a similar number of utterances (n_{P1}=187 and n_{P3}=223, respectively). It should be noted that the confederate accompanied P1 to only four of the five exhibits, whereas she visited all five exhibits with P2, P3, and P4.

In his second session, P1 spent the full 45 minutes in the gallery. He met the confederate at the Monkey exhibit and engaged in conversation for 2m 13s; during this time, P1 did not move the robot. They next visited Vincent and continued the conversation; P1 moved to the left side of the exhibit. P1 and the confederate went to a new exhibit (Music), where he directed the robot to the left, center, and right exhibit hotspots (not the info hotspot). The confederate joined P1 at the Face exhibit before excusing herself. P1 completed the full session length, revisiting the Music, Monkey, and Vincent. He also explored the Sunflower exhibit from the left, center, and right exhibit hotspots; in total, P1 visited 6 of the 8 new exhibit hotspots (Fig. 11a). Over the course of visiting the nine exhibits, P1 spent 7 minutes (15.6%) moving from one exhibit to another.



(a) P1 (b) P2



(c) P3 (d) P4

Fig. 11. Heatmap of the robot's movements within each exhibit area during the second in-robot gallery visitation (S2) for each participant; movement between exhibits is not shown. In each image, the Face exhibit is the leftmost, then the Monkey, Music (new), Sunflower (new), and Vincent exhibits moving in the clockwise direction. The heatmaps depict the amount of time the robot spent moving within each exhibit area, specifically from an exhibit's *main* entry pose and its 4 hotspots (*left*, *center*, *right*, and *info*).

P2 and the confederate engaged in the most talkative visitation. Together, they visited each exhibit once, and as per protocol, the confederate excused herself and departed. P2 immediately concluded her second visitation, but did not explicitly provide a reason. P2 and the confederate began at the Face exhibit; the confederate sat on right side of the Face exhibit (from the participant's view) and P2 moved the robot to the left exhibit hotspot. The confederate and P2 proceeded to Vincent where P2 demonstrated her interaction with the exhibit from the exhibit's left, center, and right hotspots. Next, P2 chose to go to the new Sunflower exhibit. She explored the center, right, and info hotspots, but primarily remained at the right hotspot which engaged in an off-topic conversation with the confederate about wildlife sightings in the White Mountains. P2 and the confederate continued their counterclockwise movement around the gallery to the Music exhibit before finishing their visitation at the Monkey exhibit. Two robot failures occurred during this run, and the robot had to be restarted during the second; the audio stream from the VGo App remained, so the confederate and P2 continued their conversation during these two breaks. The total time P2 spent in her second in-robot visitation was 23 minutes and 26 seconds (Table 3), not including the time needed to restart the robot.

P3 and the confederate began their visit together at the Monkey exhibit. P3 moved the robot back and

Table 3. Summary of second in-robot gallery visitation (S2)

P#	Time	Time spent at exhibit in order of visitation								
P1	45m	Monkey, 2m 13s	Vincent, 5m 24s	Music, 5m 43s	Face, 48s	Music, 6m 40s	Sunflower, 3m 43s	Monkey, 1m 25s	Vincent, 4m 2s	Music, 3m 44s
P2	23m 26s	Face, 1m 43s	Vincent, 2m 27s	Sunflower, 5m 57s	Music, 5m 59s	Monkey, 1m 58s				
P3	45m	Monkey, 7m 7s	Face, 3m 42s	Sunflower, 5m 54s	Music, 5m 21s	Vincent, 5m 2s	Sunflower, 2m 5s	Vincent, 35s	Monkey, 8m 58s	
P4	14m 43s	Face, 1m 36s	Vincent, 40s	Music, 1m 20s	Sunflower, 2m 35s	Monkey, 46s	Face, 53s	Music, 1m 55s		

forth between the exhibit's right and info hotspot. They then went to the Face exhibit where P3 stationed the robot at the exhibit's right hotspot for the duration of the conversation. Of the two next exhibits, P3 and the confederate went first to the Sunflower exhibit and then to Music. P3 viewed the Sunflower exhibit from its right hotspot, and the Music exhibit from its center and right hotspots. The confederate visited the Vincent exhibit before departing; P3 moved between the info and right hotspots. P3 revisited the Sunflower exhibit viewing it from the right and center hotspots. He also revisited Vincent briefly before turning around to interact with the Monkey exhibit for the remainder of his second session. P3 visited 4 of the 8 hotspots around the two new exhibits (Fig. 11c); however, he visited an additional 3 hotspots that he had not selected during his first in-robot gallery visit.

Unlike the other participants' second sessions, P4 initiated the conversation with the confederate before the robot arrived at the first exhibit (Face). P4 also led the movement to each of the next exhibits. P4 and the confederate briefly visited Vincent before moving over to the new Music exhibit. P4 moved between the exhibit's left and center hotspot, remaining at the center hotspot for the majority of the interaction. Then the confederate and P4 turned around to the Sunflower exhibit, which P4 viewed from its left, center, and right hotspots. The confederate excused herself after accompanying P4 to the Monkey exhibit, as she had visited all exhibits with the participant. P4 returned briefly to the Face exhibit and then the Music exhibit. P4 moved to the Music exhibit's left hotspot to interact with the xylophone before ending his short visit.

Table 4. Time spent on training (seconds). Experimenter time is the time spent proctoring the script before free practice began.

Training Session	Experimenter Time (s)	Free Practice Time (s)
P1 T1	315	480
P1 T2	390	575
P2 T1	315	588
P2 T2	220	0
P3 T1	410	1043
P3 T2	280	600
P4 T1	345	120
P4 T2	320	170

6.3 Results and Discussion

6.3.1 Interface Ease of Use

Table 4 shows the participants' T1 and T2 trainings. P3 required more than 10 minutes during his first training session. Due to his visual impairment on his left eye, P3 had trouble targeting the hotspots and consistently touched the screen 2 inches to the right during his first training exercise. During the second training session (T2), two of the participants required little to no retraining, while the other two only required marginally more time to practice. P2 opted out of the free practice portion of the training exercise, saying "I still remember it from the other day." There were 10 days between P2's first and second in-robot visitations.

Overall, the four participants reported finding the user interface easy to use. P2 noted that the user interface was "simple to run" and "extremely basic;" similarly, P4 noted it as being "straightforward." P3 noted the ease of pressing the buttons and selecting an exhibit. After P3's first in-robot visitation, he reported that the interface was "very easy to understand. It only takes a couple minutes of instruction, and you're on your way with it. That's what's nice." P3 also commented on the interface's overall accessibility, he stated:

Table 5. Frequency count of the participants selecting user interface elements during their two in-robot visitations.

	P1		P2		P3		P4	
	S1	S2	S1	S2	S1	S2	S1	S2
Choose new exhibit								
• exhibit menu	4	4	4	3	1	1	1	1
• exhibit buttons	6	6	3	3	3	7	6	6
Choose new view (via hotspots)								
• left	6	6	5	4	0	0	4	3
• center	9	9	5	3	1	3	8	4
• right	5	5	3	3	5	13	7	3
• info	0	0	2	1	2	5	4	1
View closer/further (via step buttons)								
• forward	2	2	17	19	0	1	0	2
• backward	9	9	8	7	1	1	4	3
View exhibit information								
• hotspot	0	0	2	1	2	5	4	1
• info button	1	1	5	3	0	0	1	1
• exhibit button	1	1	1	0	0	0	1	4

> You don't have to have a bunch of dexterity. If you're handicapped or have limited movement, you can use all of the same buttons and it shows you the same thing. You're not missing out on anything because you can't move a button or a toggle switch as far as you'd like to. Your hand doesn't get tired. It doesn't matter if you're left or right handed, you can use the controls with both hands.

Choosing a new exhibit. There were two methods of directing the robot to a new exhibit. First, there were exhibit buttons around the robot base icon with the icon and background color of each exhibit. Second, the participant could also choose a new exhibit by pressing the menu button; the interface would then display a menu with the exhibit buttons arranged horizontally in a pop-up modal. The participants began each of their in-robot visitations with the exhibit menu opened, and selected the first exhibit in this manner.

Table 5 shows that the participants primarily used the first method to direct the robot to a new exhibit. P3 and P4 used the menu to make their initial selections for both in-robot visitations, then exclusively used the exhibit buttons around the robot base icon. P1 and P2 used both methods during both in-robot visitations with similar number of selections for each. P1 and P3 both noted that selecting an exhibit was easy to do with the user interface during interviews.

Accessing an exhibit's information. There were three ways of accessing the information for a given exhibit. First, when the robot arrived at a new exhibit, there was an information sign to the right of the exhibit; the participant could move the robot closer to the sign by selecting the rightmost hotspot *info*. The content of the sign was also incorporated into the interface. When the robot was stationary at an exhibit, pressing the iButton in the lower right corner would display a text box with the exhibit's title and the description in a pop-up modal. This information could also be accessed by pressing the exhibit button around the robot base icon.

Table 5 also shows the participants' distinct interaction styles. P1 never physically moved the robot to read the sign in the remote environment. For his first session, he accessed information for two of the three available exhibits, whereas in his second session, he only accessed exhibit information for two of the five exhibits. P2 viewed exhibit information frequently and used all three ways during her first session. She primarily accessed information using the iButton. Unlike P1, P3 always positioned the robot physically in front of the exhibit information sign.

P4 swapped interaction styles. In his first session, P4 noted trying to read the information signs on the exhibit walls while robot was moving. At his third exhibit, he figured out that selecting the rightmost hotspot would position the robot in front of the sign. P4 changed strategies for his second in-robot visitation and primarily viewed the exhibit information on the screen by pressing corresponding exhibit button around the robot base icon.

Movement near the exhibits. Table 5 shows the participants' selections of the hotspots for the first and second in-robot visitations of the gallery, which was described in detail above. The participants had a variety of feedback regarding the hotspots. In all three interviews, P1 specifically noted the hotspots as being easy to use. After his first in-robot gallery visitation, P1 stated:

> Just choosing the exhibit I wanted and have the robot go there, that was pretty easy. That got me into the neighborhood of where I wanted to be, and then I just move it around going backward or forward using the hotspots.

P3 described how he used the hotspots during his second in-robot gallery visit:

> I use the buttons [hotspots] to move it in closer. The buttons came up. I went in to change the angles this much. But this time [second visit] I could look once I started pressing the buttons to hit the angles... I watch when they come up. I would say, "OK. This one would be a good angle."

P4 noted going to the hotspots as his strategy for directing the robot's movements during his interview following the first gallery visitation. P4 noted that the simple appearance of the hotspots did not give him enough information about their function:

> It would be nice to know what they were going to do... It took me until the third one [exhibit] to figure out that that one all the way on the right would go and look at the placard for the description. The first two times [exhibits], I kept trying to read the sign when it was moving around. I never got to go and actually look at it [the sign]... [The hotspot needed] some indicator as to where it was going to go or what it was going to look at.

Overall, the hotspots' simple appearance was able to convey a real-world position in the remote environment. There was no indication of any associated orientation or function, which the participants learned through their experiences.

Each of the exhibits was programmed with a different behavior which would trigger when the robot was directed to the left, center, and right hotspots. The exhibits also behaved differently if the robot was directed to move forward at each of these hotspots, based on the values read from the distance sensors in front of each exhibit. As shown in Table 5, P2 frequently used the robot's forward and backward translations; in her interview after the first in-robot visitation, P2 noted, "Once I read that Vincent did things as you were in different positions or got closer, then I tried that out with the other exhibits. What it [an exhibit] would do if you were at different angles, or forward, or backwards." P1 primarily directed the robot to back up; in an interview, he noted this action as "trying to be a little bit more independent than just moving to hotspots. Get a closer look at stuff, and backing up to take a longer view."

P1 and P4 noted a desire for more precise movement over the robot. Even when an autonomous robot performs its behaviors with perfect reliability [54], a user may feel dissatisfied due to lack of involvement and/or lack of control (e.g., [37, 39]). P1 felt that that it was difficult to plan ahead for where he wanted to go in the gallery, hence the compensatory technique of moving the robot backwards himself. Prior to participating in the study, P2 and P4 thought they would have to drive the robot themselves. P4 suggested joystick or gamepad control and the ability to turn the robot in place (i.e., rotate). P1 suggested being able to create a new hotspot either by dragging from the robot base icon to the desired location or touching the screen at different location (e.g., over the robot's composite video).

6.3.2 Interface Transparency

If a telepresence robot system has been designed well – for the user, interactants, and bystanders – the technology should just disappear [15, 69]. With a well-designed system that provides the user with telepresence, the user will be able to focus on the remote environment through the interface, not the interface itself. The focus of both parties will be the communication and the interpersonal relationships, not the technology. The quality of an interaction depends upon both communication and telepresence. An overview of qualitative and quantitative performance measures for the audio signal, video signal, and human-human communication can be found in Tsui et al. [78].

Our interface allowed the participants to feel as though they were telepresent in the gallery. All participants were able to use our system to experience the gallery. When visiting the gallery for the first time, participants would laugh and make comments about the exhibits to themselves or to the experimenters. For both gallery visitations, every participant was to develop an informed opinion about their favorite exhibit and provide reasoning as to why they liked it. Table 6 shows that each of the participants spent time at all available exhibits during Sessions 1 and 2. Aside from P1 and P3 visiting their favorite exhibits during Session 2, participants spent less time at exhibits that they had seen previously.

Movement and Navigation Strategies. Our interface allowed participants to form and execute movement strategies for viewing the exhibits. P3 stated:

> I would look at it straight on and then I would start looking at the angles before I was going to leave whatever I was looking at just to see if there was something that maybe I did miss... It made you want to study the exhibit even more. It made it more interesting.

Similarly, by moving to different hotspots, P4 noted that the Sunflowers "don't all move in the same direction... there's no sunlight for them to follow."

Participants were also able to form and execute navigation strategies for exploring the gallery in a manner similar to being there in person. In both her visitations, P2 toured the gallery in a counterclockwise manner. P4 similarly toured the gallery in a clockwise manner for his first visitation. In his second visitation, he chose exhibits at random before returning to confederate's favorite, and finishing with Music, which he reported as his new favorite exhibit. P3 stated a general, open approach:

Table 6. Total time spent at each exhibit (seconds) by each participant during Sessions 1 and 2. Note that during S1, only the Face, Monkey, and Vincent exhibits were available for viewing. During S2, all exhibits could be viewed.

	P1		P2		P3		P4	
	S1	S2	S1	S2	S1	S2	S1	S2
Face	446.9	248.3	270.9	102.7	599.9	221.8	252.1	149.6
Monkey	342.8	277.7	301.8	117.6	282	965.3	332.6	46.4
Vincent	240	566	248	187	206.2	337.1	256.3	39.9
Music	n/a	967.3	n/a	358	n/a	320.5	n/a	194.5
Sunflower	n/a	214	n/a	357.2	n/a	469.5	n/a	154.7

> I tried to go to the back wall so while I was moving around, I could catch three or four that I wanted to check out. "OK. I'll check this out," but I don't want to go the farthest part of the room and then come back. Sometimes you'll walk by stuff and then later on you'll come back and be able to go, "Oh man, I missed that," because you slow down and look at things a little bit more.

Table 7 shows that the confederate and participants at times explicitly verbally negotiated their movements. The confederate was instructed to prompt the participants to move between exhibits. For her second visitation, P2 alternated picking exhibits with the confederate, noting that the confederate led first; then P2 chose her favorite (Vincent) and following that the new one closer to Van Gogh (Vincent), which was the Sunflower exhibit. However, P4 led the movement to each of the next exhibits. The confederate had to explicitly ask P4 where he was already in process of going to the next exhibit and catch up.

The confederate also encouraged the participants to interact with the exhibits, and asked the participants to switch sides with her. P2 and the confederate were quite verbal in these negotiations. Analyzing the position between the confederate and the participant was beyond the scope of this study [40, 66]; however, P2 reported developing a strategy for engaging with the exhibit and the confederate simultaneously:

> I went opposite of whatever [the confederate] was, to interact with her in the exhibit... If she was on the left side, I went to the right side. One, I was facing her, but also the exhibit, and I wasn't invading the space she was in... I could interact with her, face her as well.

Quality of Communication. Audio is critical for carrying the content of a communication between two parties. Rosenberg noted that as the audio fidelity increases, the length of a conversation also increases [61]. We chose the VGo robot as our base platform for its sophisticated audio and video communication system. We utilized the VGo App for the bidirectional audio communication for this study. The synchrony between its audio and our interface supported the conversations between the participants and the confederate; the delay for both was approximately 1 second. Speech intelligibility can be quantified in terms of the number of echoes, feedback occurrences, and cutouts (e.g., [29, 49]). Only P4 noted that he could "hear the [robot] speaker say what I said a little delayed from when I said it."

A human-human communication is difficult to directly measure given the inherent involvement of interpersonal relationships, and there are a number of scales that investigate different types of relationships and situations (see [62] for an overview). We developed a series of open-ended questions for the Session 2 interview to measure the quality of the communication. We incorporated four items from Yarosh and Markopoulos's Affective Benefits and Cost of Communication Technologies (ABCCT) questionnaire, which focuses on connectedness between two parties, the engagement and expressiveness supported by a communication technology, and potential unmet expectations relating to the response time and attention levels using a communication technology. We rephrased the questionnaire items into an open-ended format (Table 8). Like Kiesler et al. [36], we asked the participants to recall a particular topic of conversation – the confederate's favorite exhibit.

All four participants reported that the conversation between themselves and the confederate was normal; none of the participants felt obligated to converse with the confederate. P2 and P3 felt as though they were co-present with the confederate, and P1 likened their encounter to touring a gallery with a curator. All participants were able to discern the confederate's mood after spending time with her in the gallery and recall her favorite exhibit. P3, who had no prior social interactions with the confederate, noted that she was a good conversational partner and even initiated a joke with her ("Is that big spider next to your shoulder supposed to be there?").

Table 7. Number of utterances volleyed between confederate (Conf) and participant during participant's second gallery visit (S2)

	P1	Conf$_{P1}$	P2	Conf$_{P2}$	P3	Conf$_{P3}$	P4	Conf$_{P4}$
Movement	8	24	33	38	12	25	7	8
System Visibility	7	2	9	8	17	8	0	0
About Exhibit								
• Neutral statement	9	21	22	20	18	33	9	21
• Positive statement	10	24	39	33	16	25	7	14
• Negative statement	1	0	0	2	0	1	1	1
• Neutral judgement	7	6	5	5	1	2	1	4
• Positive judgement	4	5	28	28	8	17	10	11
• Negative judgement	3	0	0	0	0	0	0	0
Related to Exhibit	21	26	14	15	4	9	10	8
Unrelated to Exhibit	12	11	48	45	26	38	10	13
Miscommunications								
• Simultaneous start	3	4	3	3	6	6	3	3
• Interruption	5	6	8	26	0	4	0	0

Table 8. Rephrasings of the Affective Benefits and Cost of Communication Technologies (ABCCT) questionnaire items [91]

ABCCT Item	Open-ended Rephrasing
I felt sad because **X** took too long to respond when I tried to contact **X** using the medium. (unmet expectations)	Was there any time in today's session that you <*confederate's_name*> took too long to respond to something you said? **Yes** **No** How did it make you feel? Describe the situation.
I felt sad because **X** didn't pay enough attention to me when we used the medium. (unmet expectations)	Was there any time in today's session that you felt <*confederate's_name*> didn't pay enough attention to you when you were using the robot? **Yes** **No** How did that make you feel? Describe the situation.
I could tell over the medium how **X** was feeling that day. (emotional expressiveness)	How was <*confederate's_name*> feeling today? What was his or her mood?
I had to talk to **X** using the medium even if I didn't want to. (feeling obligation)	Was there any time in today's session that you felt like you had to talk to <*confederate's_name*> when you didn't want to? **Yes** **No** Describe the situation.

Like Fish et al. [19], we looked at the content of the conversations between the participants and confederate in their second in-robot visitation; see Table 7. We discerned three salient categories: discussion of an exhibit, discussion of an aspect related to the exhibit, and off-topic conversation. As with face to face conversations, there were instances of crosstalk in which the participants and the confederate began to speak at the same time, or verbally interjected (usually backchanneling "mm-hmm").

Comparing In-robot vs. In-person Visitations. The participants visited the gallery in-person at the conclusion of this study, and saw each of the five exhibits and the telepresence robot they had operated. They noted that the gallery seemed largely the same as how they first experienced it using the telepresence robot. Participants were allocated 30 minutes to examine the gallery in-person and discuss the similarities and differences with the experimenters. It should be noted that the participants concluded their in-person visit quickly: $t_{P1} = 4\text{m } 10\text{s}$, $t_{P2} = 8\text{m } 30\text{s}$, $t_{P3} = 6\text{m } 35\text{s}$, and $t_{P4} = 7\text{m}$.

Participants were not given a map on the remote environment. There were no comments regarding the layout of the gallery. Two of the four participants (P1 and P2) felt that the gallery space seemed larger via robot than in person. P2 noted that the exhibits were mounted at a lower height in the gallery than expected. The participants did not comment on the robot's height or size.

P1 and P4 found that the colors appeared brighter in the space; P4 noted that Van Gogh (Vincent) was brighter and more vivid in person. Participants also noticed details in the exhibits when visiting in person. Specifically, P2 saw that the Face exhibit had washers, not lightbulbs like she had previously thought. P3 found that the Monkey looked less "beat up" and that the Face appeared to move less in person. P4 found that the purple Face's sounds from its movements were less annoying in person.

7 Conclusion and Future Work

Telepresence is a multifaceted continuum of user, task, system, and environmental factors [75]. The degree to which human operators can achieve telepresence in teleoperation varies largely given that the experience is dependent upon user perception and psychology, system design characteristics, and the fidelity of the medium for presenting the remote environment. We investigated our interface's ease of use and its transparency to understand the degree to which the participants experienced remote social interaction and the remote environment itself in this case study. The quality of an interaction via a telepresence robot can be measured both quantitatively and qualitatively, decomposed into the quality of a communication from a technical standpoint (audio and video), and the quality of a human-human communication through a telepresence robot.

Towards this end, we instantiated these two communication concepts for the case study indirectly as open-ended interview questions. We measured the degree to which the participants experienced the remote environment by asking them to describe their most favorite and least favorite exhibits after each in-robot visitation. Specifically, we asked them (1) to recall the exhibit's color and what it did (e.g., motions, sounds), (2) why it was his or her most/least favorite, and (3) how interacting with the exhibit made him or her feel. In order for participants to form these opinions and answer the questions, our interface must have provided the ability to experience the remote environment. The camera view had sufficient fidelity and responsiveness to display an exhibit's color and motion. Additionally, when the participants visited the gallery in-person at the conclusion of this study, they noted that the gallery seemed largely the same as how they first experienced it using the telepresence robot. There were no comments regarding the layout of the gallery. Although 30 minutes were allocated for this in-person visit, they concluded quickly with P2 spending the longest (t_{P2} = 8m 30s).

To measure the quality of the communication between the participants and the confederate in Session 2, we incorporated four items from Yarosh and Markopoulos's ABCCT questionnaire. We asked the participants to recall a particular topic of conversation [36] – the confederate's favorite exhibit. Again, these questions indirectly measured the degree to which the participants experienced remote social interaction. In order for participants to form these opinions and answer the questions, sufficiently clear and responsive audio was required; we utilized the VGo App for this purpose. Additionally, our interface had sufficient synchrony with the audio to sustain the conversation between the participants and the confederate.

The case study demonstrated that the issue of latency between commanding the robot, and the robot's subsequent movements can be addressed by the combined use of autonomous robot navigation and multiple methods of providing visual feedback (i.e., video itself, animated vector indicator, tire rotation, etc.). Our user interface consisted of integrated presentation of system status and feedback, alongside simple controls for all three levels of interaction, at top level. These features prevented the interface from being cluttered, and allowed the users to focus on interactions in the remote environment. It should be noted that this case study represents the best case scenario with respect to its content and environment [74]. First, the content of each of the gallery's exhibits was dynamic and changed when the participants approached from different angles. The placement of exhibits' sensors were designed to be triggered by the robot. Additionally, the layout of the gallery was quite simple given the number of exhibits (n=5). Second, there was only one other person in the gallery with the participant (in-robot); that is, the gallery was a closed environment, devoid of other visitors, staff, and docents. Further, the participants were already familiar with the confederate, who was an occupational therapy intern.

We look ahead to scaling our system to function in these larger, less structured public settings, and there a number of challenges which must be addressed. The concept behind our UI can be scaled to homes, schools, and museums. The content of the buttons could be changed to reflect a home, school, or work environment, and can be customized with visual support photographs, images, and icons. As the size of an environment increases in scale however, it will no longer be possible show all waypoints marked in the environment. It will also become difficult for the user to keep track of the robot's current location. We believe that an overhead map representation of the iconified environment may become necessary.

The movement of a social telepresence robot in a remote environment should be safe for people physically present with the robot, and no damage should occur to the environment or the robot [75]. Our robot system did not employ collision avoidance, which is necessary if it is to be used in dynamic environments with people in them. It is not yet known though how a telepresence robot should weigh its user's movement commands versus its own safety in an environment crowded with people. Until telepresence robots become pervasive, it can be expected that a robot will draw the attention of a crowd. Instead of the user being able to direct the robot as he or she chooses, the user may have to wait until the crowd begins to dissipate or try to circumvent the crowd.

In the near future, people will be able to go anywhere, rather everywhere, using telepresence robots. By designing for people with physical and/or cognitive impairments to be the robot operators, telepresence robots and their user interfaces will increase in the ease of use for all people. Our goal was to maximize the number of users from our target population who could perceive, understand, and operate our telepresence robot [83]. The side effect is that typically abled users with and without compromised circumstances (e.g., are in a low-lighting environment; are tired or distracted) can also operate our telepresence robot. Vanderheiden and Vanderheiden [83] note "more accessible designs are also usually easier to use by everyone all the time – but only if the ease of use is directly built in."

Social robotic telepresence is in its commercial infancy. There is a real opportunity to shape its design and capabilities in the near future in a way that will create the next must-have technology. We hope that our research contributes to the next generation of telepresence robots that people with disabilities (and without) will be able to use easily to visit remote people and places.

Acknowledgements

This research has been funded in part by the National Science Foundation (IIS-0546309, IIS-0905228, and IIS-1111125). We would thank Jill Drury of UMass Lowell and the MITRE Corporation and Thomas Ryden of VGo Communications.

References

[1] R. Aipperspach, E. Cohen, and J. Canny. Modeling human behavior from simple sensors in the home. *Pervasive Computing*, pages 337–348, 2006.

[2] F. Amirabdollahian, S. Bedaf, R. Bormann, H. Draper, V. Evers, J. G. Pérez, G. J. Gelderblom, C. G. Ruiz, D. Hewson, N. Hu, et al. Assistive technology design and development for acceptable robotics companions for ageing years. *Paladyn, Journal of Behavioral Robotics*, 4(2):94–112, 2013.

[3] Associated Press. Robot rep goes to school. Wired, June 6 2006. http://archive.wired.com/techbiz/media/news/2006/06/71078. Accessed Dec. 2014.

[4] J. Beer and L. Takayama. Mobile remote presence systems for older adults: Acceptance, benefits, and concerns. In *Proc. of the 6^{th} Intl. Conf. on Human-Robot Interaction*, pages 19–26. ACM, 2011.

[5] E. Bergman and E. Johnson. Towards accessible human-computer interaction. *Advances in human-computer interaction*, 5(1), 1995.

[6] R. Bevilacqua, A. Cesta, G. Cortellessa, A. Macchione, A. Orlandini, and L. Tiberio. Telepresence robot at home: A long-term case study. In *Ambient Assisted Living: Italian Forum 2013*, pages 73–85. Springer International Publishing, 2014.

[7] G. Carruthers. Is the body schema sufficient for the sense of embodiment? an alternative to de Vignemont's model. *Philosophical Psychology*, 22(2):123–142, 2009.

[8] J. Casper and R. Murphy. Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 33(3):367–385, 2003.

[9] CastingWords. Audio transcription services: MP3s, video and more... Webpage, 2014. http://castingwords.com. Accessed Dec. 2014.

[10] R. Chellali and K. Baizid. What maps and what displays for remote situation awareness and ROV localization? In *Human Interface and the Management of Information. Interacting with Information*, pages 364–372. Springer, 2011.

[11] J. Y. Chen, E. C. Haas, and M. J. Barnes. Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(6):1231–1245, 2007.

[12] R. A. Cooper. Quality-of-life technology. *IEEE Engineering in Medicine and Biology Magazine*, 27(2):10, 2008.

[13] S. Coradeschi, A. Loutfi, A. Kristoffersson, S. Von Rump, A. Cesta, and G. Cortellessa. Towards a Methodology for Longitudinal Evaluation of Social Robotic Telepresence for Elderly. In *Proc. of Human-Robot Interaction Workshop on Social Robotic Telepresence*, 2011.

[14] M. Desai, K. Tsui, H. Yanco, and C. Uhlik. Essential features of telepresence robots. In *Proc. of the IEEE Conf. on Technologies for Practical Robot Applications (TePRA)*, 2011.

[15] J. V. Draper, D. B. Kaber, and J. M. Usher. Telepresence. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):354–375, 1998.

[16] H. Evans. Exploring robots for accessibility, in Seattle. Webpage, May 6 2014. http://futureofmuseums.blogspot.com/2014/05/exploring-robots-for-accessibility-in.html. Accessed Dec. 2014.

[17] D. Fels, L. Williams, G. Smith, J. Treviranus, and R. Eagleson. Developing a Video-mediated Communication System

for Hospitalized Children. *Telemedicine Journal*, 5(2):193–208, 1999.

[18] D. Fels, J. Waalen, S. Zhai, and P. Weiss. Telepresence Under Exceptional Circumstances: Enriching the Connection to School for Sick Children. *Proc. of IFIP INTERACT01: Human-Computer Interaction*, pages 617–624, 2001.

[19] R. Fish, R. Kraut, R. Root, and R. Rice. Evaluating video as a technology for informal communication. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 37–48. ACM, 1992.

[20] J. Fitzgerald. After Surgery, A Robot May Be at Your Side. The Boston Globe, December 12 2011. http://www.boston.com/business/technology/articles/2011/12/12/the_robot_that_makes_house_calls. Accessed Dec. 2014.

[21] N. Fliesler. A Roving Eye: Home Health Monitoring with Robotic Systems. Webpage, December 26 2011. http://www.massdevice.com/blogs/massdevice/roving-eye-home-health-monitoring-with-robotic-systems. Accessed Dec. 2014.

[22] T. Fong and C. Thorpe. Vehicle teleoperation interfaces. *Autonomous Robots*, 11(1):9–18, 2001.

[23] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, 1979.

[24] Giraff Technologies AB. giraff. Webpage, 2014. http://www.giraff.org/?lang=en. Accessed Dec. 2014.

[25] M. Gonzalez, C. Hidalgo, and A. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196): 779–782, 2008.

[26] M. Gorman. Suitable technologies introduces beam, the remote presence device. Webpage, September 26 2012. http://www.engadget.com/2012/09/26/beam-telepresence-device-suitable-technologies. Accessed Dec. 2014.

[27] C. M. Gridley, A. B. Retik, B. Cilento, and H. T. Nguyen. In-home robots can effectively engage children and their parents in post-operative care, and allow for cost-efficient remote physician monitoring. In *American Academy of Pediatrics National Conference and Exhibition*, October 20 2012. Session abstract. https://aap.confex.com/aap/2012/webprogram/Paper17310.html. Accessed Dec. 2014.

[28] E. Guizzo. When My Avatar Went to Work. IEEE Spectrum, September 22 2010. http://spectrum.ieee.org/robotics/industrial-robots/when-my-avatar-went-to-work. Accessed Dec. 2014.

[29] H. Haas. The influence of a single echo on the audibility of speech. *J. Audio Eng. Soc.*, 20(2):146–159, 1972.

[30] M. Hassenzahl. User Experience and Experience Design. Webpage, 2014. http://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html; chapter 3. Accessed Dec. 2014.

[31] G. J. Hole, P. A. George, K. Eaves, A. Rasek, et al. Effects of geometric distortions on face-recognition performance. *Perception-London*, 31(10):1221–1240, 2002.

[32] R. Hoover. The Force is Strong with NASA's Smartphone-Powered Satellite, July 8 2011. http://www.nasa.gov/mission_pages/station/main/spheres_smartphone.html. Accessed Dec. 2014.

[33] IBM Centres for Solution Innovation. Telbotics – PEBBLES, 2008. http://www.customerfacingsolutions.com/pdfs/work/Telebotics%20PEBBLES.pdf. Accessed Dec. 2104.

[34] W. A. IJsselsteijn, H. de Ridder, J. Freeman, and S. E. Avons. Presence: concept, determinants, and measurement. In *Electronic Imaging*, pages 520–529. International Society for Optics and Photonics, 2000.

[35] E. Kac. Live from Mars. *Leonardo*, 31(1):1–2, 1998. ISSN 0024-094X.

[36] S. Kiesler, A. Powers, S. Fussell, and C. Torrey. Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, 26(2):169–181, 2008.

[37] D. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, J. Bricout, Z. Wang, and A. Behal. How autonomy impacts performance and satisfaction: Results from a study with spinal cord injured subjects using an assistive robot. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 42(1):2–14, 2012.

[38] K. Koffka. *Principles of Gestalt psychology*. Routledge, 4 edition, 2013.

[39] Z. G. Kotala. Robotic Arm's Big Flaw: Patients Say It's 'Too Easy'. Webpage, September 23 2010. http://today.ucf.edu/robotic-arms-big-flaw-patients-say-its-too-easy. Accessed Dec. 2014.

[40] A. Kristoffersson, S. Coradeschi, K. S. Eklundh, and A. Loutfi. Towards measuring quality of interaction in mobile robotic telepresence using sociometric badges. *Paladyn, Journal of Behavioral Robotics*, 4(1):34–48, 2013.

[41] A. Kristoffersson, S. Coradeschi, and A. Loutfi. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, 2013.

[42] S. Kurniawan and P. Zaphiris. Research-derived web design guidelines for older people. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pages 129–135. ACM, 2005.

[43] L. Lewington. Robots explore Tate Britain's artwork after dark. Webpage, August 12 2014. http://www.bbc.com/news/technology-28742582. Accessed Dec. 2014.

[44] A. Lindgren-Streicher and C. Reich. Visitor experience monitoring project: Fiscal year 2010 report. 2011.

[45] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.

[46] K. Massie. Headline classroom robots make school possible for home-bound students. Webpage, April 5 2014. http://www.simpsonstreetfreepress.org/science-and-technology/robot-substitution. Accessed Dec. 2014.

[47] M. Micire. Evolution and field performance of a rescue robot. *Journal of Field Robotics*, 25(1-2):17–30, 2008.

[48] M. J. Micire. *Multi-Touch Interaction for Robot Command and Control*. PhD thesis, University of Massachusetts Lowell, December 2010.

[49] G. Miller and J. Licklider. The intelligibility of interrupted speech. *J. of the Acoustical Soc. of Amer.*, 1950.

[50] M. R. Mine, F. P. Brooks Jr, and C. H. Sequin. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 19–26. ACM Press/Addison-Wesley Publishing Co., 1997.

[51] J. Nielsen. *Usability Engineering*. San Francisco, CA: Morgan Kaufmann, 1993. ISBN 0-12-518406-9.

[52] J. Nielsen. Enhancing the explanatory power of usability heuristics. In *SIGCHI Conf. on Human Factors in Computing Systems*, pages 152–158, 1994.

[53] J. Nielsen. 10 usability heuristics for user interface design. Webpage, January 1 1995. http://www.nngroup.com/articles/ten-usability-heuristics. Accessed Dec. 2014.

[54] D. Norman. How might people interact with agents. *Communications of the ACM*, 37(7):68–71, 1994.

[55] D. A. Norman. *The design of everyday things*. Basic books, 2002.

[56] K. Pernice and J. Nielsen. Usability guidelines for accessible web design. Technical report, Nielsen Norman Group, 48105 Warm Springs Blvd., Fremont, CA 94539-7498 USA, 2001. http://www.nngroup.com/reports/usability-guidelines-accessible-web-design; originally titled *Beyond ALT Text: Making the Web Easy to Use for Users with Disabilities*. Accessed Dec. 2014.

[57] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong. On the Levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking (TON)*, 19(3):630–643, 2011.

[58] L. Richards. I, Student. Homebound Teen is Attending Classes via Wheel-bound Robot. Webpage, January 19 2011. http://www.timesrecordnews.com/news/2011/jan/19/i-student. Accessed Dec. 2014.

[59] L. D. Riek. Realizing Hinokio: candidate requirements for physical avatar systems. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 303–308. ACM, 2007.

[60] G. Riva, F. Davide, and W. A. IJsselsteijn. *Being there: Concepts, effects and measurements of user presence in synthetic environments*. Ios Press, 2003.

[61] J. Rosenberg. Quality matters. United Communications magazine, August 2010. http://www.tmcnet.com/ucmag/columns/articles/99344-quality-matters.htm. Accessed Dec. 2014.

[62] R. Rubin, A. Rubin, E. Graham, E. Perse, and D. Seibold. *Communication Research Measures II: A Sourcebook*. Routledge, Taylor & Francis, 2009.

[63] Ryerson University. PEBBLES. Webpage, 2011. http://www.ryerson.ca/pebbles/index.html. Accessed Dec. 2014.

[64] R. Schulz, S. R. Beach, J. T. Matthews, K. L. Courtney, and A. De Vito Dabbs. Designing and evaluating quality of life technologies: An interdisciplinary approach. *Proceedings of the IEEE*, 100(8):2397–2409, 2012.

[65] B. Shneiderman. Direct manipulation: A step beyond programming languages. *Sparks of Innovation in Human-Computer Interaction*, page 17, 1993.

[66] D. Sirkin, G. Venolia, J. Tang, G. Robertson, T. Kim, K. Inkpen, M. Sedlins, B. Lee, and M. Sinclair. Motion and attention in a kinetic videoconferencing proxy. In *Human-Computer Interaction–INTERACT 2011*, pages 162–180. Springer, 2011.

[67] Suitable Technologies. How beam works. Webpage, 2014. https://www.suitabletech.com/beam/#howBeamWorks. Accessed Dec. 2014.

[68] Suitable Technologies, Inc. Shelbot (aka Shelbot), January 9 2013. http://blog.suitabletech.com/2013/01/09/shelbot-aka-shel-bot. Accessed Dec. 2014.

[69] L. Takayama. Toward making robots invisible-in-use. *New Frontiers in Human-Robot Interaction*, 2, 2011.

[70] R. Toris and B. Alexander. The Standard ROS JavaScript Library, June 11 2014. http://wiki.ros.org/roslibjs. Accessed Dec. 2014.

[71] K. Tsui, M. Desai, H. Yanco, and C. Uhlik. Exploring use cases for telepresence robots. In *Proc. of Intl. Conf. on HRI*. ACM/IEEE, 2011.

[72] K. Tsui, A. Norton, D. Brooks, H. Yanco, and D. Kontak. Designing telepresence robot systems for use by people with special needs. In *Proceedings of the International Symposium on Quality of Life Technologies 2011: Intelligent Systems for Better Living, held in conjunction with RESNA 2011 as part of FICCDAT*, 2011.

[73] K. M. Tsui. Design and evaluation of a visual control interface of a wheelchair robotic arm for users with cognitive impairments. Master's thesis, University of Massachusetts Lowell, May 2008.

[74] K. M. Tsui. *The Development of Telepresence Robots for People with Disabilities*. PhD thesis, University of Massachusetts Lowell, April 2014.

[75] K. M. Tsui and H. A. Yanco. Design challenges and guidelines for social interaction using mobile telepresence robots. *Reviews of Human Factors and Ergonomics*, 9(1):227–301, 2013.

[76] K. M. Tsui, K. Abu-Zahra, R. Casipe, J. M'Sadoques, and J. L. Drury. A Process for Developing Specialized Heuristics: Case Study in Assistive Robotics. Technical report, University of Massachusetts Lowell, 2009. Available at http://teaching.cs.uml.edu/techrpts.

[77] K. M. Tsui, K. Abu-Zahra, R. Casipe, J. M'Sadoques, and J. L. Drury. Developing Heuristics for Assistive Robotics. In *Proc. of Intl. Conf. on HRI*. ACM/IEEE, 2010. Late breaking paper.

[78] K. M. Tsui, M. Desai, and H. Yanco. Towards Measuring the Quality of Interaction: Communication through Telepresence Robots. In *Proc. of the Performance Metrics for Intelligent Systems Workshop (PerMIS)*, 2012.

[79] K. M. Tsui, K. Flynn, A. McHugh, H. A. Yanco, and D. Kontak. Designing speech-based interfaces for telepresence robots for people with disabilities. In *IEEE International Conference on Rehabilitation Robotics (ICORR)*, 2013.

[80] K. M. Tsui, E. McCann, A. McHugh, M. Medvedev, H. A. Yanco, D. Kontak, and J. L. Drury. Towards designing telepresence robot navigation for people with disabilities. *International Journal of Intelligent Computing and Cybernetics*, 7 (3):307–344, 2014.

[81] K. M. Tsui, A. Norton, D. J. Brooks, E. McCann, M. S. Medvedev, J. Allspaw, S. Suksawat, J. M. Dalphond, M. Lunderville, and H. A. Yanco. Iterative design of a semi-autonomous social telepresence robot research platform: a chronology. *Intelligent Service Robotics*, 7(2):103–119, 2014.

[82] K. Urrutia. Creating a CSS3 pulsating circle. Webpage, January 24 2012. http://kevinurrutia.tumblr.com/post/16411271583/creating-a-css3-pulsating-circle. Accessed Dec. 2014.

[83] G. Vanderheiden and K. Vanderheiden. Guidelines for the design of consumer products to increase their accessibility to persons with disabilities or who are aging. Webpage, 1992. http://trace.wisc.edu/docs/consumer_product_guidelines/toc.htm; working draft v1.7. Accessed Dec. 2014.

[84] P. Vespa. Multimodality Monitoring and Telemonitoring in Neurocritical Care: From Microdialysis to Robotic telepresence. *Current Opinion in Critical Care*, 11(2):133, 2005. ISSN 1070-5295.

[85] P. Vespa, C. Miller, X. Hu, V. Nenov, F. Buxey, and N. Martin. Intensive Care Unit Robotic Telepresence Facilitates Rapid Physician Response to Unstable Patients and Decreased Cost in Neurointensive Care. *Surgical Neurology*, 67(4):331–337, 2007. ISSN 0090-3019.

[86] VGo Communications, Inc. VGo robotic telepresence for healthcare, education and business, 2014. http://www.vgocom.com. Accessed Dec. 2014.

[87] W3C. Web Content Accessibility Guidelines (WCAG) 2.0. Webpage, December 11 2008. http://www.w3.org/TR/2008/REC-WCAG20-20081211. Accessed Dec. 2014.

[88] W3C. User agent accessibility guidelines (UAAG) 2.0. Webpage, November 7 2013. http://www.w3.org/TR/2013/WD-UAAG20-20131107; draft. Accessed Dec. 2014.

[89] J. Wineman, J. Peponis, and R. Conroy Dalton. Exploring, engaging, understanding in museums. In *Proceedings of the Space Syntax and Spatial Cognition Workshop: Spatial Cognition '06*. Springer, 2006. Monograph Series of the Transregional Collaborative Research Center. Universität Bremen, Bremen.

[90] H. A. Yanco, H. J. Kim, F. G. Martin, and L. Silka. Artbotics: Combining art and robotics to broaden participation in computing. In *AAAI Spring Symposium on Robots and Robot Venues: Resources for AI Education*, 2007.

[91] S. Yarosh and P. Markopoulos. Design of an Instrument for the Evaluation of Communication Technologies with Children. In *Proc. of the 9th Intl. Conf. on Interaction Design and Children*, pages 266–269. ACM, 2010.

A HCI Heuristics and Accessibility Guidelines

We surveyed the design guidelines and heuristics that have been developed for both general-purpose user interfaces and interfaces for assistive technologies. These guidelines are provided below in Tables 9 through 13; note that guidelines are a subset, and concepts may be repeated with slightly different wording or emphasis from their authors. Nielsen's usability heuristics [51, 53] are listed in Table 9. Web Content Accessibility guidelines [87] by W3C are listed in Table 10. Kurniawan and Zaphiris's guidelines for older adults [42] are listed in Table 11. Finally, Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [83] are listed in Tables 12 and 13.

Table 9. Relevant Nielsen Usability Heuristics [51, 53]

ID	Heuristic	Description
N-1	**Visibility of system status**	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
N-2	**Match between system and the real world**	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
N-3	**User control and freedom**	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
N-4	**Consistency and standards**	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
N-5	**Error prevention**	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
N-6	**Recognition rather than recall**	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
N-7	**Flexibility and efficiency of use**	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
N-8	**Aesthetic and minimalist design**	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

Table 10. Relevant W3C Web Content Accessibility Guidelines [87]

ID	Guideline	Description
W-1	**Perceivable**	Information and user interface components must be presentable to users in ways they can perceive.
W-1a	**Text Alternatives**	Provide text alternatives (e.g., large print, speech, symbols, or simpler language)
W-1b	**Distinguishable**	Make it easier for users to see and hear content including separating foreground from background (i.e., color, audio, contrast, scale)
W-2	**Operable**	User interface components and navigation must be operable.
W-2a	**Enough Time**	Provide users enough time to read and use content.
W-3	**Understandable**	Information and the operation of user interface must be understandable.
W-3a	**Readable**	Make text content readable and understandable.
W-4	**Robust**	Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies.

Table 11. Relevant Kurniawan and Zaphiris's Web Design Guidelines for Older Adults [42]

ID	Description
KZ-1	Provide larger targets.
KZ-2	There should be clear confirmation of target capture.
KZ-3	Graphics should be relevant and not for decoration.
KZ-4	Icons should be simple and meaningful.
KZ-5	Avoid pull down menus.
KZ-6	Avoid scroll bars.
KZ-7	Language should be simple and clear.
KZ-8	Avoid irrelevant information on the screen.
KZ-9	Information should be concentrated mainly in the centre.
KZ-10	Screen layout, navigation and terminology used should be simple, clear and consistent.
KZ-11	Provide ample time to read information.
KZ-12	Reduce the demand on working memory by supporting recognition rather than recall (N-6) and provide fewer choices to the user.
KZ-13	Background screens should not be pure white.
KZ-14	High contrast between the foreground and background should exist.
KZ-15	Content should not all be in color alone.
KZ-16	Support user control and freedom. (N-3)

Table 12. Relevant Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [83] (Part 1 of 2)

ID	Guideline	Description
V-1	**Seeing visual output clearly**	– Make letters and symbols on output as large as possible/practical. – Use upper and lowercase type to maximize readability. – Use high contrast between text or graphics and background. – Keep letters and symbols on visual output as simple as possible. – Replace or supplement color coding with different shape or relative position coding.
V-2	**Understanding output**	– Use simple screen layouts. – Hide (or layer) seldom used commands or information. – Keep language as simple as possible. – Use attention-attracting and grouping techniques (e.g., putting a box around things or color blocking.) – Present information in as many (redundant) forms as possible/practical or provide as many display options as possible.
V-3	**Physically operating controls**	– Space controls out to provide a guard space between controls. – Provide for operation with left or right hand. – Avoid controls that require twisting or complex motions (e.g., push and turn.) – Space, position and size controls to allow manipulation by individuals with poor motor control or arthritis.

Table 13. Relevant Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [83] (Part 2 of 2)

ID	Guideline	Description
V-4	**Understanding how to operate controls**	– Minimize dual purpose controls. – Use selection techniques where the person need only make a single, simple, non-time-dependent movement to select. – Reduce or eliminate lag/response times. – Minimize ambiguity. – Provide a busy indicator or, preferably, a progress indicator when a product is busy – Use simple concise language. – Use redundant labeling (e.g., color code plus label). – Lay out controls to follow function – Standardize - Use same shape/color/icon/label for same function or action.