THE DEVELOPMENT OF TELEPRESENCE ROBOTS FOR PEOPLE WITH DISABILITIES

BY

KATHERINE MEIFUNG TSUI
B.S. UNIVERSITY OF MASSACHUSETTS LOWELL (2004)
M.S. UNIVERSITY OF MASSACHUSETTS LOWELL (2008)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
COMPUTER SCIENCE
UNIVERSITY OF MASSACHUSETTS LOWELL

Author: .... Date: April, 2014

Dissertation Chair: ....
Dr. Holly A. Yanco

Committee Member: ....
Dr. Jill Drury

Committee Member: ....
David Kontak, OTR/L

Committee Member: ....
Dr. Brian Scassellati

The Development Of Telepresence Robots For People With Disabilities

Katherine Meifung Tsui

April 2014

Abstract

A person's quality of life is impacted when he or she is no longer able to participate in everyday activities with family and friends, which is often the case for people with special needs (e.g., seniors and people with disabilities) who are full time residents at medical and healthcare facilities. We posit that people with special needs may benefit from using telepresence robots to engage in social activities. Telepresence robots provide interactive two-way audio and video communication and can be controlled independently, allowing the person driving to use the robot to look around and explore a remote environment as he or she desires. However, 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.

Over the course of three years, we have designed and architected a social telepresence robot research platform based on a VGo Communications' VGo robot. Our work included designing a new processing and sensor system with three cameras to create a wide field of view, and laser range finder to support autonomous navigation. The images from each camera were combined into a vertical panoramic video stream, which was the foundation of our interface. Since the premise of a telepresence robot is that it is an embodiment for its user, we designed and implemented autonomous navigation behaviors that approximated a human's as much as possible, given its inability to independently translate laterally.

This research utilized an iterative, bottom-up, user-centered approach, drawing upon our assistive robotics experiences. We have conducted series of user studies to inform the design of an augmented reality style user interface. We conducted two formative evaluations (a focus group (n=5) and a follow-on "Wizard of Oz" experiment (n=12)) to investigate how members of our target population would want to direct a telepresence robot in a remote environment. Based on these studies, we developed an augmented reality user interface, which focuses primarily on the human-human interaction and communication through video, providing appropriate support for semi-autonomous navigation behaviors. We present a case study (n=4), which demonstrates

this research as a first critical step towards having our target population take the active role of the telepresence robot operator.

Acknowledgments

I would first like to acknowledge and thank my committee, Dr. Holly Yanco, Dr. Jill Drury, Dr. Brian Scassellati, and Dave Kontak. Your collective guidance and insight throughout this process have been instrumental. Thank you all for supporting my grandiose goals and helping me rein them into this resulting thesis. To Holly, thank you for nurturing me as an assistive roboticist. From the beginning of my master's through the end of my doctorate, you have been an unparalleled mentor, and I owe my research career to you.

Thanks to all of the members of the UMass Lowell Robotics Lab (present, past, and unofficial) who are part of the larger "we" appearing throughout this dissertation: Dan Brooks, Jim Dalphond, Munjal Desai, Eric McCann, Mark Micire, Mikhail Medvedev, Adam Norton, Abe Schultz, Jordan Allspaw, Brian Carlson, Vicki Crosson, Kelsey Flynn, Amelia McHugh, Sompop Suksawat, Michael Coates, and Sean McSheehy. I stand on all of your shoulders. Thank you for the many late nights coding, running experiments, analyzing data, and writing papers.

Thank you to my family and friends for their support and patience. To my parents, Terence and Theresa, and siblings, Kim and Nick, simply, I love you. To Rhoda and Amy, thank you for always having an open door. To Munjal and Dan, thank you for the miles of sanity regained from cycling and climbing. To Jim and Misha, your support in the last months pushing to complete this research will always be remembered. To Mark, you are by far my better half, who keeps me grounded and reminds me to breathe; thank you for being my tireless cheerleader.

Thank you to Thomas Ryden of VGo Communications for supporting Hugo and Margo's growing pains. Finally, although I cannot explicitly states names, I would like to give a special thank you to all the participants from Crotched Mountain Rehabilitation Center for providing inspiration and passion to continue with my research in assistive robotics. I dedicate this work to you.

This work was supported in part by the National Science Foundation (IIS-0546309, IIS-1111125).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A person's quality of life is impacted when he or she is no longer able to participate in everyday activities with family and friends, which is often the case for people with special needs (e.g., seniors and people with disabilities) who are full time residents at medical and healthcare facilities. Isolation can lead to feelings of overall sadness which can lead to additional health issues [Findlay, 2003]. Hopps et al. [2001] found that for people with disabilities, there was a negative correlation between loneliness and physical independence.

The population size of people with special needs is increasing, which is an effect caused by the "baby boomer" generation moving towards retirement age and the increased life expectancy due to medical care for people with congenital disabilities, people with developmental disabilities, and people with acquired, non-congenital disabilities due to injury, including war veterans. The number of seniors (65+) is expected to double from 40.2 million in 2010 to 88.5 million by 2050 [Vincent and Velkoff, 2010]; see Figure 1-1. According to the 2000 US Census, 1.6 million people over the age of 65 lived in a nursing home [Hetzel and Smith, 2001]. In the 2003 Olmstead report, the National Council on Disability [2003] notes that the majority of seniors do not like living in nursing homes and approximately 90% of seniors with disabilities have family members as their primary caregivers [Saynor, 2001]. Of the 35 million people over the age of 65 living in a non-institutionalized setting, 20.4% reported "difficulty going outside the home (e.g., going outside the home alone to shop

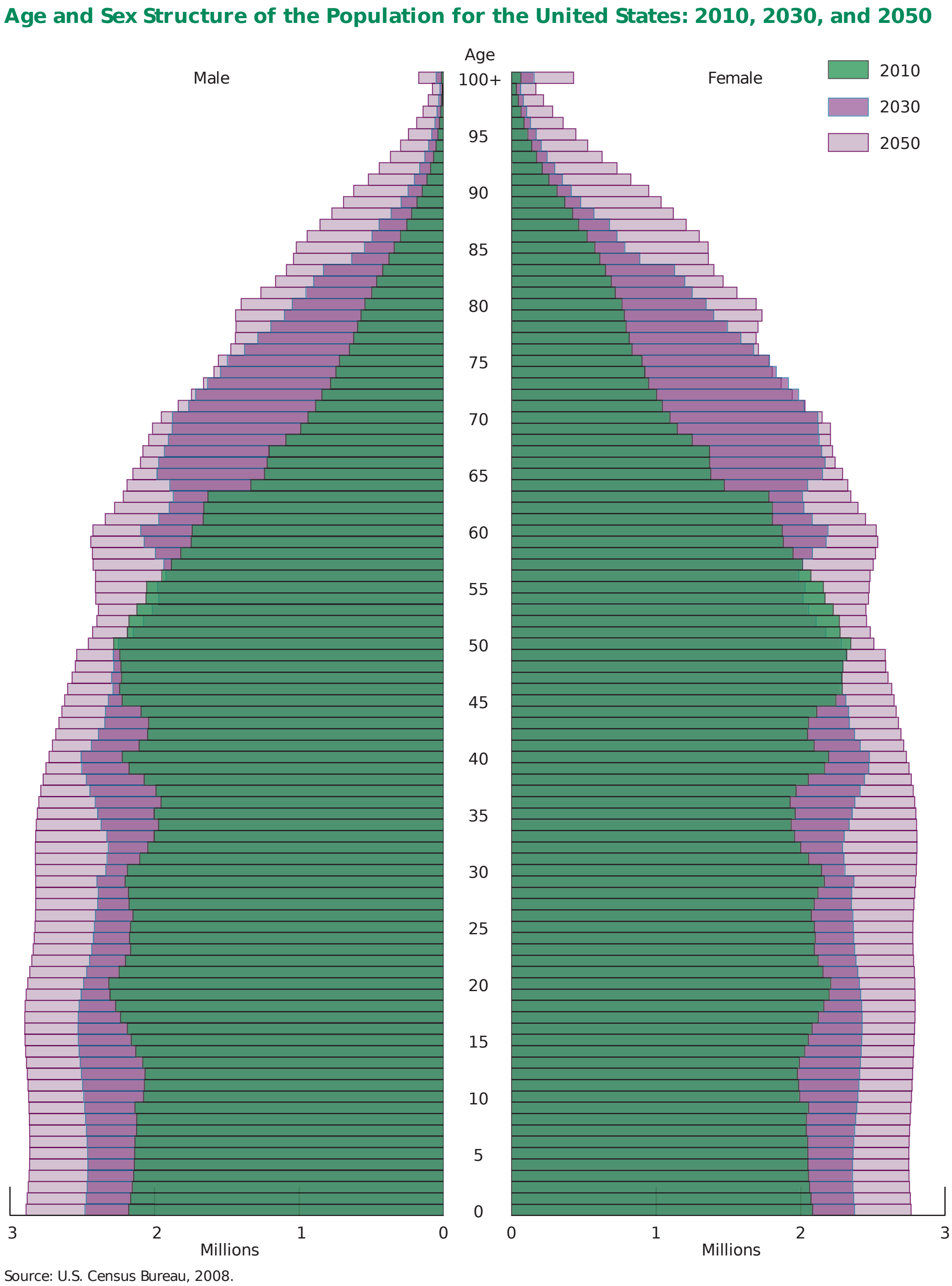


Figure 1-1: US Census population projection to 2050 [Vincent and Velkoff, 2010]

or visit a doctor's office)" [Gist and Hetzel, 2004]. Additionally, 106,000 people with developmental disabilities lived in state-operated and private institutions, and 35,000 lived in nursing home facilities [Braddock, 2002; National Council on Disability, 2003].

There is the belief that social engagement can help to mitigate depression. Researchers have investigated having people with special needs connect with each other through the Internet (e.g., [Bradley and Poppen, 2003]). Additionally, robots have been developed as social companions, including Pearl the Nursebot [Pollack et al., 2002], Paro the baby harp seal [Marti et al., 2006; Shibata et al., 2001; Taggart et al., 2005; Wada et al., 2005], PaPeRo [Osada et al., 2006], Robota [Billard et al., 2006], KASPAR [Robins et al., 2009], and Robovie [Iwamura et al., 2011; Sabelli et al., 2011]. (See Broekens et al. [2009] and Broadbent et al. [2009] for surveys.) However, Beer and Takayama [2011] note that there is a difference between companion robots and robots designed to promote social interaction between people as telepresence robots can do.

Contemporary commercial telepresence robots focus on the concepts of remote presence and telecommunication. They are designed for a broad segment of the population, including corporate executives, engineers, sales associates, office workers, doctors, caregivers, and students – as opposed to trained robot specialists, or roboticists, as with other types of telepresence robots. This new class of telepresence robots provides a human operator with social presence in a remote environment as would a telecommunication system, such as a telephone or videoconferencing system, while also providing independent mobility through teleoperation of the robot (Figure 1-2). These robots are at the intersection of physical and social presence, called *copresence* [IJsselsteijn et al., 2000; Riva et al., 2003].

We have conducted previous research to determine what types of office workers might have the most positive experiences using these telepresence robots in an office environment [Tsui et al., 2011a]. We found that people who were no longer in the same building as their teammates had the best experiences recreating the closeness with their teams using telepresence robots. We posit that similar benefits can be gained by people with special needs who wish to engage in social interaction but cannot be physically present with their family and friends.

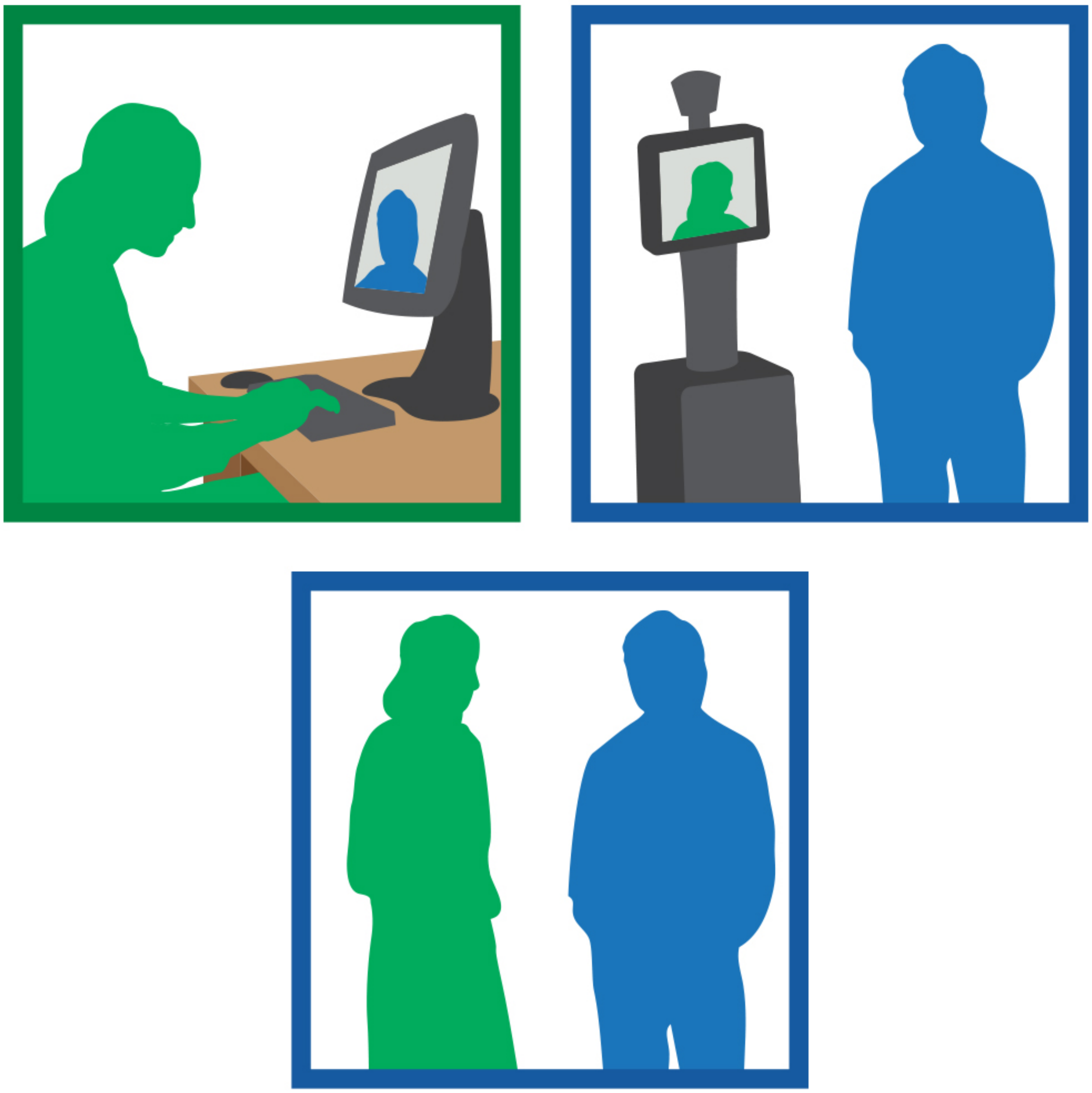


Figure 1-2: The robot's user (top left, green person) operates a telepresence robot in a remote environment (top right, left side of image). In this interpersonal communication use case, the user converses with an interactant (blue person). The degree to which the user feels telepresent with the interactant in the remote environment and vice versa (bottom) is dependent upon the quality of the user's human-computer interaction (top left) and the interactant's human-robot interaction (top right). [Tsui and Yanco, 2013]

Figure 1-3: **(Left)** A person with disabilities in passive role of the interactant (being visited by the telepresence robot). **(Right)** Our research focuses on the inverted role in which people with disabilities takes the active role of operating the robot.

1.1 Scope

The scope of this dissertation research considers the user experience from both the robot operator's perspective and also the perspective of people physically present with the robot. Our research focuses on the scenario in which people with special needs take the active role of operating telepresence robots. It should be noted there has been considerable research already done in the use case where the person with special needs is visited by a healthcare professional, family member, or friend operating a telepresence robot (i.e., passive role), discussed further in Chapter 2. The active role is depicted as the green person in Figures 1-2 (top left) and 1-3 (right), and the passive role as the blue person in Figures 1-2 (top right) and 1-3 (left).

Hassenzahl [2011] describes a "user experience" as 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. For our use case, 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 robot may be used exclusively as a conversation tool. Other people may want to check on their family and observe them, while still others may wish to attend an art exhibit opening or tour a museum. Others may simply want to be present in a space to feel more included in an activity, like attending high school via telepresence robot.

What lists the function(s) that people can do with a device [Hassenzahl, 2011]. Telepresence robots support "calls," which allow you to connect with another person.

Figure 1-4: Hugo (an augmented VGo Communication's VGo telepresence robot) is being driven remotely and being used to walk alongside a colleague, actively participating in a mobile conversation. The driver can be seen on Hugo's screen.

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 conversation. Hassenzahl [2011] 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.

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 [Tsui and Yanco, 2013]:

1. the robot itself (herein referred to as the *telepresence robot*),
2. the robot's user (herein referred to as the *user*),
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*, respectively).

How describes the design of the device and its interface [Hassenzahl, 2011]. Telepresence robot designers must consider three main interactions. First, there is the human-computer interaction between the user and the robot's interface (Figure 1-2, top left), 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., [Casper and Murphy, 2003; Fong and Thorpe, 2001; Hoover, 2011; Kac, 1998; Micire, 2008]). Second, there is the HRI in the remote environment between the interactants and the telepresence robot itself (Figure 1-2, top right); the interactants converse with the user through his or her telepresence robot embodiment. Finally, there is the interpersonal human-human interaction (Figure 1-2, bottom); if these first two interactions are successful, then robot mediation will be minimized, 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 [Draper et al., 1998; Lombard and Ditton, 1997].

Our approach in designing HRI systems is 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 approaches described in Cooper [2008] and Schulz et al. [2012]. We utilize this approach and draw upon our experiences in the domains of assistive robotics and human-computer interaction (HCI).

1.2 Problem Statement

Commercial telepresence robots are being sold as a means for ad-hoc, mobile, embodied video conferencing. These robots are typically teleoperated using a combination of mouse clicks and key presses. However, these input devices require a fine level of manual dexterity, which may not be suitable for use by people with special needs who may additionally have physical impairments. People in our target population may have significantly limited range of motion, strength, and dexterity in their upper extremities [Tsui et al., 2011d]. Additionally, there is a high cognitive workload associated with teleoperating a remote telepresence robot, and our target population may have difficulty decomposing this type of complex task [Tsui and Yanco, 2010]. To date, telepresence robots have not been designed for use by people with special needs as the robot operators.

1.3 Research Questions

This research investigated the following questions:

RQ1: What levels of abstraction and autonomy are needed for people with disabilities to effectively control a telepresence robot system in a remote environment?

It has been largely assumed that the user is always controlling the telepresence robot's movements, regardless of the robot. A robot will move forward when the up arrow key is pressed, for example, and remain moving forward until the key is released, causing it to stop. Many of the contemporary commercial telepresence robot interfaces are designed for operation from a laptop or desktop computer, and the robots are operated using a combination of key presses, mouse clicks on GUI buttons or widgets indicating proportional velocity control. The RP-7 robots also have dedicated operation consoles with joysticks [InTouch Health., 2011; InTouch Technologies, 2011]. Thus, the general perception of how to use a telepresence robot has been to provide low level forward, back, left, and right (FBLR) commands. Teleoperating a robot at this level can be a cognitively taxing task, particularly over long periods of time.

It is important to understand how members of our target population conceptualize a remote environment and what they expect a telepresence robot to be able to do in terms of navigation in the given space. In a preliminary evaluation [Tsui et al., 2011b], we found that 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. This issue is consistent with our previous work in which we found that able-bodied novice users had difficulty driving telepresence robots straight down a corridor [Desai et al., 2011; Tsui et al., 2011a]. The latency often caused the robot to turn more than the user intended and thus zig zag down the hallway. 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.

RQ2: What are the essential components of a research platform needed to inform the design of future telepresence robots for the target population?

For social telepresence robots, the user has an interface to move the robot and to communicate with people in the remote environment. To support this interaction, the robot must be able to move, have a microphone to allow the user to hear sounds from the remote site, have speakers to allow the user to be heard at the remote site, have a camera to allow the user to see the environment around the robot, and have a video screen to present the user to people in the remote environment [Tsui and Yanco, 2013]. "Interactants" and "bystanders," interacting directly and indirectly with a telepresence robot, may not be receptive to robot designs that do not share human-like characteristics [Tsui and Yanco, 2013]. User task engagement and sense of telepresence may be degraded if interactants and bystanders are unwilling to communicate with the user. Unlike virtual reality, in which all users have the same capacity to create bodies, it is imperative for interactants and bystanders to accept the telepresence robot as a representation of the user. Consequently, the robot must be able to function sufficiently well as a human proxy.

RQ3: Which design principles facilitate the development of telepresence robot interfaces for use by the target population?

As noted by Coradeschi et al. [2011], designing a user interface for people with special needs has different requirements. Large buttons and text with high contrast are necessary for low-vision users [Nielsen Norman Group Report, 2001; Tsui et al., 2009; Vanderheiden and Vanderheiden, 1992]. Simple language and familiar real world analogies may allow robot drivers to recognize how to use the interface rather than having to recall how to use it from training and/or their own experience [Nielsen, 1994a]. Within this research question, there are many challenging questions to investigate. For example, how should system status and feedback be provided without cluttering the interface? How can multiple ways of commanding a telepresence robot and navigation behaviors be represented? Can issues of latency between commanding the robot and the robot moving be overcome through interface design?

RQ4: To what degree can the target population experience remote social interaction and the remote environment itself?

Telepresence is a multifaceted continuum of user, task, system, and environmental factors [Tsui and Yanco, 2013]. 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. Empirically determined factors include visual display parameters (e.g., frame rate, latency, field of view, stereopsis, point of view, image resolution, color quality, image clarity); consistency of environmental presentation across displays; nonvisual sensing (i.e., sound, e.g., mono, directional; haptics, e.g., touch, force feedback); environmental interactivity (e.g., response rate to user input, reciprocal interaction capability between remote environment and the user, clarity of causal relationships between user actions and environmental reaction); anthropomorphism of the user representation; and so on (see [IJsselsteijn et al., 2000; Lee et al., 2010; Ma and Kaber, 2006; Slater, 2005; Slater et al., 1994; Slater and Wilbur, 1997]).

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 [Tsui and Yanco, 2013]. As previously stated, the degree to which the user feels telepresent with the interactant in the remote environment and vice versa is dependent upon the quality of the user's HRI (top left of Figure 1-2, p. 4) and the interactant's (top right). We investigated the quality of interaction through telepresence robots in pieces: the quality of a communication from a technical standpoint (audio and video), and the quality of a human-human communication through a telepresence robot.

1.4 Approach

We developed several guidelines for the design of telepresence robots based on our previous studies conducted at Google in Mountain View, CA during July and August 2010 [Desai et al., 2011; Tsui et al., 2011a, b]. 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. Direct teleoperation is impractical due to inherent network latency and the movement of people in the remote environment. To facilitate the investigation of these research questions, we designed, developed, and architected a social telepresence robot research platform, discussed in Chapter 3. We selected a VGo Communications' VGo robot as the base platform and incorporated additional processing and sensing; specifically, we added three cameras to create a wide field of view, and laser range finder to support the robot's autonomous navigation. We also address the essential components of social telepresence robot research platforms (RQ2) in Chapter 3.

Our focus on the active scenario is unique, and to give our research context, we summarize the evolution of telepresence robots and contemporary systems in Chapter 2. We describe in detail telepresence robots that have been used in the healthcare field, noting the capabilities of the robots beyond direct teleoperation and the user interface where applicable.

We conducted two formative evaluations regarding autonomous robot navigation using a participatory action design process, described in Chapter 4. First, we conducted a focus group (n=5) to investigate how members of our target audience would want to direct a telepresence robot in a remote environment using speech. We then conducted a follow-on experiment in which participants (n=12) used a telepresence robot or directed a human in a scavenger hunt task. 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, back, 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). This key insight begins to answer RQ1 regarding the level of abstraction and autonomy needed for our target population to effectively control a telepresence robot system in a remote environment.

We describe our accessible telepresence robot user interface and the HRI, HCI, and accessibility guidelines employed in Chapter 5. Our interface featured a first person, video-centric view, provided by the three cameras we added; the images from each camera were combined into a vertical panoramic video stream. We believe that the understanding of a robot's autonomous capabilities should be facilitated by the HRI interface presentation and system feedback, and we discuss which design principles can facilitate the development of telepresence robot interfaces in general for use by our target population (RQ3) in Chapter 5. Our interface was designed to support the robot's low-, mid-, and high-level autonomous movement and navigation behaviors (discussed in Chapter 6); these were represented as buttons placed at the bottom of the interface and overlaid on the video. The robot's autonomous movement updated the visual feedback displayed on the interface.

To test our end-to-end system, we designed and conducted a case study, which allowed four people of our target population to visit an art gallery. Chapter 7 is a demonstration of the end-to-end system. We investigated the 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 (RQ4) in Study 3. The quality of an interaction via telepresence robot can be measured both quantitatively and qualitatively (Appendix F), 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. All participants were able to use our system to experience the art gallery. They were able to develop an informed opinion about their favorite exhibit and provide reasoning as to why they liked it. Finally, Chapter 8 puts forth the open research questions created by, or that will extend, this research.

1.5 Contributions

Our research has resulted in three major contributions. First and foremost was pursuing the use case of people with disabilities taking the active role of operating a telepresence robot. Our second major contribution was the design, development, and architecture of a social telepresence robot research platform, Margo. Finally, our third major contribution was an example of an "invisible to use" [Takayama, 2011] telepresence user interface designed for users from our target population to explore a remote art gallery, which included:

- collecting a data set of first-hand accounts of users from our target population giving spatial navigation commands to a telepresence robot;
- drawing a key insight from this data set that all users gave low-, mid-, and high-level directives;
- synthesizing user interface design guidelines and principles from the domains of

HRI, HCI, and assistive technology with this insight;

- designing and implementing a telepresence robot user interface based on this synthesis;
- designing and implementing the robot's movement and autonomous navigation behaviors;
- synthesizing performance measures for quality of interaction through a telepresence robot; and
- demonstrating the end-to-end system via a case study ($n = 4$).

Chapter 2

Background

In 1980, Marvin Minksy painted a picture of people suiting up in sensor-motor jackets to work at their jobs thousands of miles away [Minsky, 1980]. He called the remote control tools *telepresences*, which emphasized the idea of remotely "being there" in such a high fidelity manner it seems as though the experience was "in person." Over thirty years later, how close are we to Minsky's vision? In 2000, the US Food and Drug Administration approved the da Vinci Surgical System by Intuitive Health for laparoscopic surgeries [Singer, 2010]. In his 2009 book "Wired for War," P. W. Singer wrote about a nineteen year old soldier living in Nevada who flew unmanned aerial vehicles to fight the war in Iraq [Singer, 2009]. In some sense, our progress is close to what Minsky projected, but these are only a small number of highly specialized telepresence systems.

Current telepresence manifests itself in a large number of places in the form of interaction through live video. Friends and family who are located across continents keep in touch with each other through their web cameras and streaming video chat applications such as iChat, Skype, and Google Talk Video, launched in 2003, 2006, and 2008 respectively [Apple, 2003; Lachapelle, 2008; Skype, 2006]. As of December 2010, there were 145 million connected Skype users, and in the fourth quarter of 2010, video calls were 42% of the Skype-to-Skype minutes [Skype S.A., 2011]. The video conference meeting is a daily activity for some workers. Telepresence through video conferencing ranges from the one-on-one video applications, to dedicated high-end telepresence rooms that show near-life size meeting participants on panoramic displays, to video kiosks designed for the person "dialing in" to embody (e.g., Microsoft's Embodied Social Proxy [Venolia et al., 2010]).

Robotics has re-entered the telepresence space but not as manipulators in sealed nuclear facilities as envisioned by Minsky [1980]. Research in the domain of telepresence robots has yielded robots such as robot submarines for subsea exploration [Hine et al., 1994], the RESQ information gathering robot used to monitor radiation levels at the Fukushima nuclear plant [Guizzo, 2011], Mars Rovers for space exploration [Kac, 1998], and Geminoid HI-1 for inter-personal communication across distances to better convey a person's remote physical presence [Sakamoto et al., 2007].

Telepresence robots can be described as embodied video conferencing on wheels. These new telepresence robots provide a physical presence and independent mobility in addition to communication, unlike other video conferencing technologies. Early telepresence robots were developed through academic research. The Personal Roving Presence (PRoP) robots were the first Internet controlled, untethered, terrestrial robotic telepresences, developed in the late 1990s [Paulos and Canny, 1998]; PRoPs enabled a single user to wander around a remote space, converse with people, hang out, examine objects, read, and gesture, which are largely the goals of contemporary telepresence robots. A number of early telepresence robots were developed as museum tour guide robots which allowed groups of remote visitors to see a given museum from the robot's perspective (e.g., Rhino [Burgard et al., 1998], Minerva [Schulz et al., 2000; Thrun et al., 2000], Xavier [Simmons et al., 2002], and TOURBOT [Trahanias et al., 2000]).

InTouch Health was the first company to commercialize their Remote Presence (RP) robots in this new communication telepresence robot market. Trials of the RP-7 robots began at rehabilitation centers and eldercare facilities in 2003 [InTouch Technologies, Inc., 2003a, b], and in hospitals in 2004 [InTouch Technologies, Inc., 2004].

After their commercial launch of the Roomba in 2002 [Fox, 2005], iRobot also approached this new communication telepresence robot space with the consumer in mind. They announced their $3,500 CoWorker robot in 2002 and their $500 ConnectR robot (a Roomba with a video camera) five years later [iRobot, 2007]. Neither product caught on, and Colin Angle noted that "off the shelf component costs still have not come down to the point that the business opportunity becomes irresistible" [Fox, 2005]. However, iRobot still believed in the concept of remote presence and presented their AVA robot at the Consumer Electronics Show in January 2011 [Hornyak, 2011a]. At the InTouch Health 7^{th} Annual Clinical Innovations Forum in July 2012, iRobot and inTouch Health revealed their RP-VITA (Remote Presence Virtual + Independent Telemedicine Assistant) robot, the next generation acute care telepresence robot [InTouch Technologies, Inc., 2012b]. iRobot and Cisco joined forces in creating Ava 500, which sold its first unit in March 2014 [Burt, 2014].

As shown in Figure 2-1, several new communication telepresence robots have emerged in the last decade though corporate efforts and partnerships between research institutions and companies (grouped by year):

- Telebotics' PEBBLES in 1997 [Telebotics, 2005],
- Fraunhofer IPA's Care-O-bot I in 1998 [Fraunhofer IPA, 2012a],
- iRobot's Co-Worker and Fraunhofer IPA's Care-O-bot II in 2002 [Fraunhofer IPA, 2012b; iRobot, 2007],
- InTouch Health's RP-7 in 2003 [InTouch Technologies, Inc., 2003a],
- RoboDynamics' TiLR in 2005 [RoboDynamics, 2011],
- Giraff Technologies' Giraff (formerly HeadThere) in 2006 [Giraff Technologies AB, 2011],
- iRobot's ConnectR in 2007 [iRobot, 2007],
- Fraunhofer IPA's Care-O-bot 3 in 2008 [Fraunhofer IPA, 2012c],
- Anybots' QA (Question and Answer), Willow Garage's Texai, Korean Institute of Science and Technology's (KAIST) Roti, and 3Detection Labs' R.BOT 100 in 2009 [Ackerman, 2009; Kwon et al., 2010; Rbot, 2012; Willow Garage, 2011b],

- Anybots' QB, VGo Communications' VGo, the KAIST's EngKey, and Yujin Robotics' Robosem in 2010 [Anybots, 2011; Ha-Won, 2010; Saenz, 2011; VGo Communications, 2011],

- RoboDynamics' Luna, iRobot's AVA, Gostai's Jazz Connect, and 9th Sense's TELO in 2011 [Ackerman and Guizzo, 2011; Gostai, 2011; Hornyak, 2011a; Manning, 2012],

- 9th Sense's HELO, InTouch Health's and iRobot's RP-VITA, Double Robotics' Double, Suitable Technologies' Beam (based on the Texai prototype, now known as Beam Pro), and CtrlWorks' Puppet in 2012 [CtrlWorks, 2013; Double Robotics, 2013; InTouch Technologies, Inc., 2012b; Manning, 2012; Suitable Technologies, Inc., 2014b],

- iRobot and Cisco's Ava 500, Orbis Robotics' Teleporter and Biolocator, and CSIRO and National Museum Australia's Chesster and Kasparov robots [iRobot, 2014; National Museum Australia, 2014; Orbis Robotics, 2014], and

- Suitable Technologies' Beam+ [Suitable Technologies, Inc., 2014a, c] in 2014.

A number of companies are targeting small-, medium-, and even large-sized businesses by selling these robots as mobile video conferencing units to support remote collaboration beyond the conference room. They envision their telepresence robots being used for a wide variety of applications including inspections at oversea manufacturing facilities and classroom education. Although this mobile video conferencing technology is currently out of the price range for many personal consumers, as the platforms range from $995 USD for a Beam+ [Suitable Technologies, Inc., 2014c] to $6,000 for a VGo robot [VGo Communications, 2011] to $5,000 monthly rental fees for an RP-7 [InTouch Health., 2011], we anticipate that in the near future the telepresence robot will become a common household electronic device, like the personal computer [Venkatesh and Brown, 2001].

In this chapter,[1] we investigate several telepresence robot systems used as healthcare

[1]Portions of this chapter were published in [Tsui and Yanco, 2013].

Figure 2-1: Since the Personal Roving Presence (PRoP) robots in the late 1990s [Paulos and Canny, 1998], several new communication telepresence robots have emerged though corporate efforts and partnerships between research institutions and companies. (Images are not to scale.)

Figure 2-2: A doctor operates the RP-7 from his console using a joystick (right) to visit his patients in their hospital room (left). Image from [InTouch Health, 2010].

support tools. Examples include doctors conducting patient rounds at medical facilities, doctors visiting patients in their home post-surgery, healthcare workers visiting family in eldercare centers, and students with disabilities attending school from home.

2.1 Patient Care

The InTouch Health Remote Presence (RP) robots were the first contemporary telepresence robots designed to let a doctor "be in two places at once" and therefore allow specialists to connect with patients beyond their own hospital [InTouch Health., 2011]. The RP-7 robot has a motorized base with holonomic drive control which allows the operator to move in any direction at any time using a joystick, as shown in Figure 2-2 [InTouch Health., 2011; InTouch Technologies, 2011]. The base contains 30 infrared distance sensors that allow the operator to see obstacles around the robot [InTouch Health., 2011]. A 15-inch LCD display is mounted as the "head" with a pan-tilt-zoom camera mounted above the screen. The RP-7 robot stands 5 feet 5 inches tall (65 in, 1.65 m) [InTouch Technologies, 2011]. In addition to the two-way live audio and video, the RP-7 supports medical sensors such as a wireless, electronic stethoscope [Lo, 2010].

Figure 2-3: InTouch Health's and iRobot's RP-VITA (Remote Presence Virtual + Independent Telemedicine Assistant. Image from [Ackerman and Guizzo, 2011].

The RP-VITA is the next generation acute care remote presence system, developed in conjunction with iRobot, shown in Figure 2-3 [InTouch Technologies, Inc., 2012b]. The RP-VITA (Remote Presence Virtual + Independent Telemedicine Assistant) robot is a similar stature to the RP-7, standing 5 feet 4 inches tall (65 in, 1.63 m) [Adams, 2012]. It features two video cameras above a large primary video conferencing screen, which can automatically pan towards the person speaking. RP-VITA's sensor suite is similar to that of iRobot's AVA (i.e., PrimeSense IR cameras, sonar, and a laser) [Ulanoff, 2012]; it is capable of autonomous navigation, which received FDA clearance in January 2014 [InTouch Health, 2013]. In addition to joystick control, an iPad can be used locally to control the RP-VITA and send it to a destination [InTouch Technologies, Inc., 2012b; Ulanoff, 2012]. The RP-VITA robot has an on-board electronic stereoscope and can connect to a number of diagnostic devices including ultrasound [InTouch Technologies, Inc., 2012b].

The RP-7 costs $200,000 to buy or $5,000 per month to rent [Hadzipetros, 2007; InTouch Health., 2011; Lo, 2010]; rental prices are similarly anticipated for the RP-VITA [Adams, 2012]. Given the expense of the RP robots and console stations, they have been primarily used in hospitals as a way to bring in a doctor's expertise when necessary. Doctors have used the RP robots to remotely supervise surgical procedures [Agarwal et al., 2007; Rothenberg et al., 2009; Smith and Skandalakis, 2005], and provide stroke expertise to community hospital facilities [InTouch Health., 2011; Lai, 2008]. Critical care hospital staff have been able to monitor patients in a neurosurgical intensive care unit (ICU) [Vespa, 2005; Vespa et al., 2007], and also provide on-call back up for surgical and burn ICUs [Chung et al., 2007]. The RP-7 system is in use at 600 hospitals, and as of February 2010, over 100,000 clinical sessions have been conducted using the Remote Presence network [Lo, 2010; Ulanoff, 2012].

The VGo Communications' VGo telepresence robots have also been used in outpatient care for doctors to check on their patients in their homes after surgery [Fitzgerald, 2011; Fliesler, 2011]. The Children's Hospital Boston (CHB) launched a five robot pilot program and sent the robots home with 40 patients [Fitzgerald, 2011]. A doctor logged into the robot to visit patients for surgical follow-up for 2 weeks. Gridley et al. [2012] report that patients who had VGo robots at home had fewer unexpected emergency room or clinic visits as well as a decrease in phone calls. Also, both the patient's parents and physician indicated higher satisfaction than those in the nonintervention group [Gridley et al., 2012]. The number of patients in the intervention group has grown to 80 as of March 2013 [Stockton, 2013].

Dr. Hiep Nguyen, Co-Director of Center for Robotic Surgery, Director of Robotic Surgery Research and Training, and an Associate in Urology at Children's Hospital Boston, foresees robotic home monitoring as a means to reduce the length of a post-surgery hospital stay. Nguyen [2012] notes that with a telepresence robot, a doctor is able to perform in-home assessments of a patient's recovery process (e.g., gait analysis of stair climbing, visual analysis of urine samples). Physicians have supervised several low-risk stent removals at patient homes through the VGo robots [Gridley et al., 2012]. The degree to which doctors feel telepresent in patient homes is unknown; however, patients and their parents feel as though the doctor is telepresent through the robot. Parents participated more and demonstrated higher levels of understanding regarding their child's postoperative care [Gridley et al., 2012]. The CHB's chief innovation officer noted the pilot program's success and that the patients "don't want to give the robot up and they really feel connected to the physician" [Parmar, 2012]. This bond may be because the VGo (48 inches) is similar in height to the patients, as posited by Nguyen [2012].

2.2 Engaging People with Special Needs

Telepresence robots have also been discussed in the context of aging in place and residential care for people with special needs, which can be demonstrated through two scenarios. The first is a passive scenario (Figure 1-3 (left), p. 5). That is, a telepresence robot can be located in the residence of the senior or person with a disability; healthcare attendants and family members can then call in and operate the telepresence robot to check on the person. This scenario has been actively researched (e.g., [Beer and Takayama, 2011; Cesta et al., 2011; Hans et al., 2002; InTouch Technologies, Inc., 2003a, b; Michaud et al., 2008]). The RP robots have also been used by healthcare staff at rehabilitation centers [InTouch Technologies, Inc., 2003a] and community eldercare facilities [InTouch Technologies, Inc., 2003b]. Telepresence robots, such as Giraff [Cesta et al., 2011], Telerobot [Michaud et al., 2008], TRIC (Telepresence Robot for Interpersonal Communication) [Tsai et al., 2007], and Care-O-bot [Hans et al., 2002], have been designed for home care assistance so that healthcare professionals, caregivers, and family members could check seniors and people with disabilities when necessary.

The second is an active scenario (Figure 1-3 (right), p. 5); That is, the person with special needs assumes the role of the operator. The telepresence robot can be located, for example, in his or her family's home, at a friend's home, at work or school, or at a museum. There are few examples of people with special needs using telepresence robots in the real world. The PEBBLES, VGo, and R.BOT 100 robots have been used by students with disabilities to attend their regularly scheduled classes.

Beer and Takayama [2011] conducted a user needs assessment of seniors (n=12; ages 63-88) with a Texai robot in both of these scenarios. The participants were visited by a person who operated the Texai (as in the passive scenario), which is

Figure 2-4: Participant interaction with a Texai. **(Left)** A participant passively interacting with the Texai. **(Right)** The participant actively operating the telepresence robot. Images from [Beer and Takayama, 2011].

shown in Figure 2-4 (left). The participants also assumed the role of operator and controlled the Texai to interact with a person, which is an example of the active scenario, shown in Figure 2-4 (right). The researchers found that in post-experiment interviews, the participants discussed significantly more concerns when visited by a person through the telepresence robot (as in the passive scenario) than the condition when the participants operated the Texai telepresence robot (as in the active scenario), which implies that seniors are willing to operate telepresence robot systems. With respect to where the participants wanted to use the telepresence robots, Beer and Takayama reported that 6 of 12 participants wanted to use the robot outside, 5 wanted to attend a concert or sporting event through the robot, and 4 wanted to use the robot to visit a museum or a theatre.

2.2.1 Telepresence Robots in the Home

When thinking of robots in the home, fictional personal assistants such as the Jetson's Rosie come to mind. Several personal robot assistants have been designed to be in the residencies of people with disabilities (e.g., Pearl the Nursebot [Pollack et al., 2002], TeCaRob [Helal and Abdulrazak, 2006, 2007]). Care-O-Bot was designed to be a personal robot assistant in the home of a person with special needs. Its functionalities included manipulating household objects, assisting with mobility through an integrated

Figure 2-5: **(Left)** Care-O-bot 2, circa 2002, listed videophone communication as a requirement [Hans et al., 2002]. **(Center and right)** Care-O-bot 3, circa 2008, uses a touchscreen on a tray to interact with the primary user (i.e., person with special needs) [Graf et al., 2009] or smart phone [Mast et al., 2012].

walker, and acting as a communication aid [Hans et al., 2002; Schaeffer and May, 1999]. The researchers specifically noted that Care-O-bot should be able to act as a video phone and be able to communicate with doctors, family, and friends; this video calling feature was significantly more desirable to caregivers than elders [Mast et al., 2012]. Fraunhofer IPA in Stuttgart, Germany has created three iterations of the Care-O-bot in 1998, 2002, and 2008 [Graf et al., 2009]. However, the work has largely focused on the design of the robot, sensors, manipulation, and navigation [Graf et al., 2009; Reiser et al., 2009]. Unlike the previous versions, Care-O-bot 3 does not include a video screen (Figure 2-5 center), although the video call functionality does exist.

When a user activates the "Make Call" service, an informal caregiver or a 24-hour professional teleassistant is then invited to take control of the robot and takes its perspective through its cameras [Mast et al., 2012]. Figures 2-6 and 2-7 are prototypes of user interfaces for an informal caregiver (UI-CG) and 24-hour professional teleassistant (UI-PRO), respectively. In the UI-CG, the caregiver can send the Care-O-bot 3 to specified location within a room by tapping on the screen, which is the primary means of navigation. The planned path to the destination is shown and can be modified during navigation by dragging it. The robot's position can be fine tuned using a circular widget from a live-video view; the robot's camera view is contextually grounded by showing a small portion of the local map with its marked field of view. Proportional control is implied from arrows increasing in size around a center circle

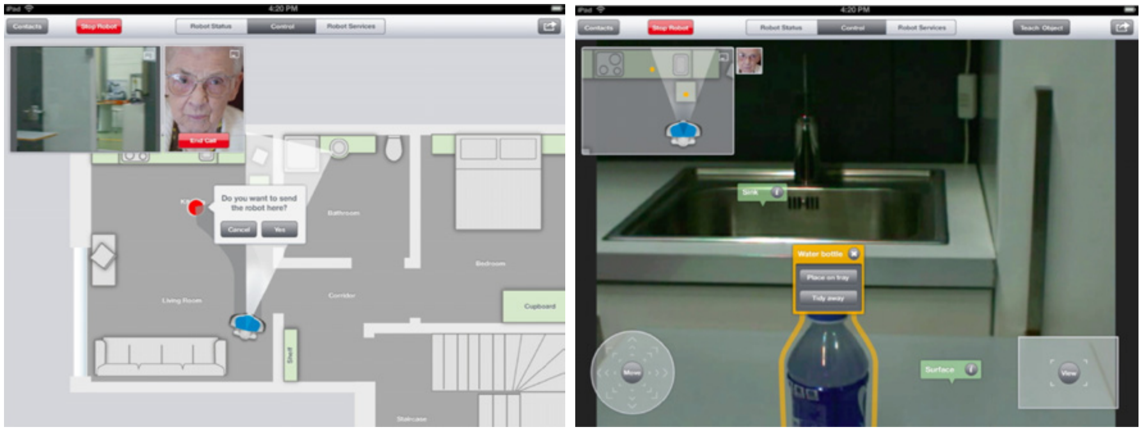


Figure 2-6: Care-O-bot 3 prototype user interface for informal caregiver for use on a tablet computer [Mast et al., 2012]. **(Left)** The caregiver can send the robot to specified location within a room by tapping on the screen. **(Right)** The robot's position can be fine tuned using a circular widget, shown in the lower left of a live-video view. Proportional control is implied from arrows increasing in size around a center circle labeled "move."

Figure 2-7: Care-O-bot 3 workstation concept for 24-hour professional teleassistant [Mast et al., 2012]. The majority of the main screen (shown in the center of the monitor) is dedicated to manipulation assistance with multi-degree of freedom devices such as the Phantom Omni (right) and SpaceNavigator (left). The prototype interface features a map of the environment overlaid with the localized robot in the bottom left of the main screen. A conventional joystick is used for controlling the robot's navigation. The workstation includes two emergency stop buttons.

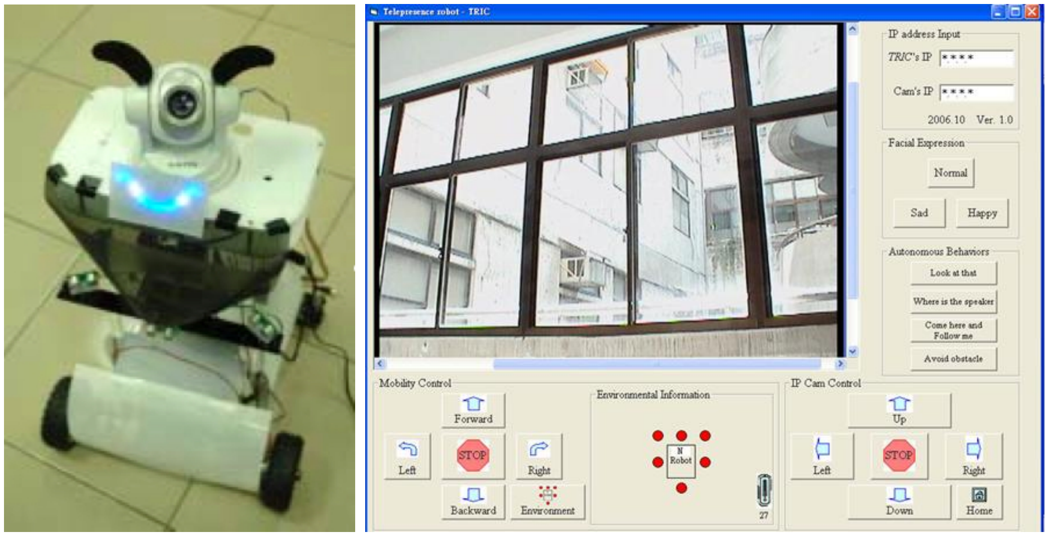


Figure 2-8: TRIC (left) and its operator interface (right) [Tsai et al., 2007].

labeled "move."

The UI-PRO interface features a map of the environment overlaid with the localized robot; a conventional joystick is used for navigating the robot. It should be noted that the researchers envision the 24-hour professional teleassistant as the final means for assisting the user when the Care-O-bot 3's autonomous capabilities were insufficient and an informal caregiver was unable to solve the problem. In a non-emergency situation, the teleassistant would likely be asked to assist with complex manipulation tasks, which is reflected in the design of prototype interface.

For some telepresence robots, the operator's face is not represented by a video stream but instead with an iconic representation of emotion (e.g., MeBot [Adalgeirsson and Breazeal, 2010], Snowie [Acosta Calderon et al., 2011]). In Taiwan, TRIC was designed as an interpersonal communication companion robot for the elderly in the home [Tsai et al., 2007]. TRIC is 29.5 inches (75 cm) tall and has a pan-tilt webcam as a "head" with two servo controlled "eyebrows" and an array of LEDs below the camera like a mouth for expression. The facial expression module has capabilities for happiness, sadness, and a neutral state; Figure 2-8 (left) shows TRIC "smiling." TRIC can be teleoperated through the interface (Figure 2-8 right), but also has autonomous capabilities for navigation assistance (obstacle avoidance, self-docking for charging). TRIC has several social autonomous behaviors including projecting a laser dot into

Figure 2-9: Telerobot (left) and its control console (right) [Michaud et al., 2010].

the environment to indicate shared attention by clicking on a location in the video feed. Additionally, TRIC can localize sound, turn towards the speaker, and move towards and track the sound. Once near the speaker, TRIC can autonomously follow the person using its sonar sensors.

Two research initiatives have focused on video communication as the primary function of personal robot assistants in the home of people with special needs. In Canada, researchers at the University of Sherbrooke developed their own telepresence robot, Telerobot. They began with a user needs assessment for robots for telerehabilitation in the home in 2003 [Michaud et al., 2010]. They use an iRobot Co-Worker robot and a robotics research platform to characterize a home including the width of hallways, door frames, transitioning over doorways. They also conducted a focus group with six participants who lived in a communal eldercare facility and eight healthcare staff.

The researchers then designed Telerobot as a circular shaped robot with omnidirectional steering. Telerobot's height can be adjusted from 26 inches (0.65 m) to 37 inches (0.95 m), as shown in Figure 2-9 (left). The base of the robot was designed to maximize video stabilization as it moves throughout a person's house across different types of flooring and across doorways. There are six infrared distance sensors on the bottom of the base to detect drop-offs. A Hokuyo URG laser, ten infrared distance sensors (five forward facing and five backward facing), and eight sonar (four forward facing and four backward facing) sensors provide information about obstacles surrounding the robot.

Figure 2-10: Telerobot version 1 interface with (A) a map of a person's home, (B) a display of range data, (C) the proportional drive control mouse-based widget, and (D) the video window of the robot's eye view [Michaud et al., 2010].

Telerobot was designed to be used by a healthcare professional, thus the control console shows the client's information on the left screen and the robot interface on the right screen in Figure 2-9 (right). The robot interface has gone through several revisions. In the first version, the video was the dominant feature in the interface (Figure 2-10). On the left side of the screen, the interface featured a known map of the person's home (top), a display of the ranging data (center), and a circular proportional control widget operated by a mouse (right). The researchers conducted a usability study with five physiotherapists and five occupational therapists controlling the robot. The results showed people with limited teleoperation experience of remote robots could navigate with Telerobot in an unknown home [Michaud et al., 2010]. The two subsequent versions of Telerobot's interface focused on the robot navigation component over the video communication.

All Version 2 interfaces were controlled using the proportional control mouse widget from Version 1. Additionally, the operator could click on the video feed or map to

Figure 2-11: **(Left)** An elderly could from Örebro, Sweden interact with a remote person through the Giraff robot. **(Right)** The Giraff interface [Coradeschi et al., 2011].

provide a destination waypoint.

In Europe, the ExCITE (Enabling Social Interaction Through Telepresence) project took a different approach by collaborating with a Swedish company, Giraff Technologies AB, which has already developed a telepresence robot [Coradeschi et al., 2011; Loutfi, 2010]. The ExCITE project is a European initiative funded by the Ambient Assisted Living project. The purpose of the ExCITE project is to increase social interaction of seniors between their family, caregivers, and other senior friends by using the Giraff robot (Figure 2-11 left). Towards this end, long-term studies will be held in Sweden, Italy, and Spain [Loutfi, 2010]; the Giraff robot will remain at twelve end-user locations for a period of six months [Coradeschi et al., 2011]. The researcher will look at the ease of installing and maintaining the Giraff system, the interface itself, any privacy concerns, and the overall acceptance and use of the Giraff system.

Coradeschi et al. [2011] describe the first participant site in Sweden. The participants were an elderly couple who live in their Örebro home (Figure 2-11 left). The husband was ambulatory, and the wife used a wheelchair in their home which had ramps and no door thresholds. The couple wished to keep in contact with family members who lived 124 miles (200 km) away. Additionally, in Sweden, it is common practice to have a "hemtjänst" (a domestic caregiver) visit a senior's home multiple times per day, and to have an alarm service which can contact health care staff in case of emergency. At this time, an interaction must be initiated by robot operator, which in this case is the family member, hemtjänst, or alarm company personnel. The participants can accept or reject an incoming call. An emergency call can be programmed as well.

The Giraff robot has a 14 inch (35.6 cm) video screen mounted in a portrait orientation on a tilt unit. A 2 mega pixel webcam with a wide angle lens is mounted above the screen. The robot is 5.9 feet tall (1.8 m) and has the option to mechanically adjust its height [Björkman and Hedman, 2010; Thiel et al., 2010]. Björkman and Hedman [2010] noted that a commercial Giraff robot relied on the operator to keep the robot safe; they augmented a Giraff with infrared sensors to detect drop offs and sonars for obstacle avoidance.

The Giraff interface is a video-centric interface similar to Telerobot's Version 1 interface. To drive the robot, the operator presses the mouse cursor on the video (Figure 2-11 right). A red curve originating from the bottom of the video is drawn. This curve represents the velocity of the robot; the magnitude of the curve indicates the speed and the angle left or right indicates how the robot should turn. The operator can also control the tilt of the robot's head; a double-click on the top of the video will tilt the camera towards the ceiling, on the bottom will tilt the camera towards the base of the robot, and in the center will level the camera. The operator can turn the robot in place by clicking on the left and right sides of the video. The interface has buttons to move the robot backwards and also to automatically turn 180 degrees to face the other direction.

2.2.2 Telepresence Robots in Classrooms

The telepresence robots described in the previous section focus on being present in the home of a person with special needs. They can also be used beyond the home; for example, the PEBBLES and VGo robots have been used by students with disabilities to attend their regularly scheduled classes. PEBBLES (Providing Education By Bringing Learning Environments to Students) was developed as a collaboration between Telebotics, the University of Toronto, and Ryerson University from 1997

Figure 2-12: **(Left)** An elementary school version of the PEBBLES II robot being used in a classroom. **(Center)** A child uses the PEBBLES II system to attend class and controls the robot in the classroom with a video game controller. **(Right)** The PEBBLES robot features a near life size display of the child's face, a webcam, and a hand to "raise" for asking questions. Images from Ryerson University [Ryerson University, 2011].

through 2006 [Fels et al., 2001, 1999; Ryerson University, 2011; Telebotics, 2005]. PEBBLES was a means for hospitalized children to continue attending their regular schools. PEBBLES has been used in Canada since 1997 and across the US since 2001 including at UCSF Children's Hospital, Yale-New Haven's Children's Hospital, and Cleveland's Rainbow Babies and Children's Hospital [Telebotics, 2005]. Each pair of robots costs $70,000, and as of June 2006, there were forty robots on loan to hospitals [Associated Press, 2006].

As shown in Figure 2-12, one robot is placed in the child's classroom and the other robot is with the child. The child controls the remote PEBBLES robot with a video game controller [Cheetham et al., 2000; Fels et al., 1999]. The interface for PEBBLES II is shown in Figure 2-13. The primary function of PEBBLES is to provide a window into the classroom, and a large video window is provided on the left side of the display for the remote video. The child can "look around" his or her class room as PEBBLES's head can move left/right and up/down. The color of the video frame changes color depending on the direction: [Fels et al., 1999] describes that when pressing the upward control, the top half of the video frame becomes yellow; when the control is released, the video frame returns to blue.

The child can also change the zoom level (in/out) and request attention and

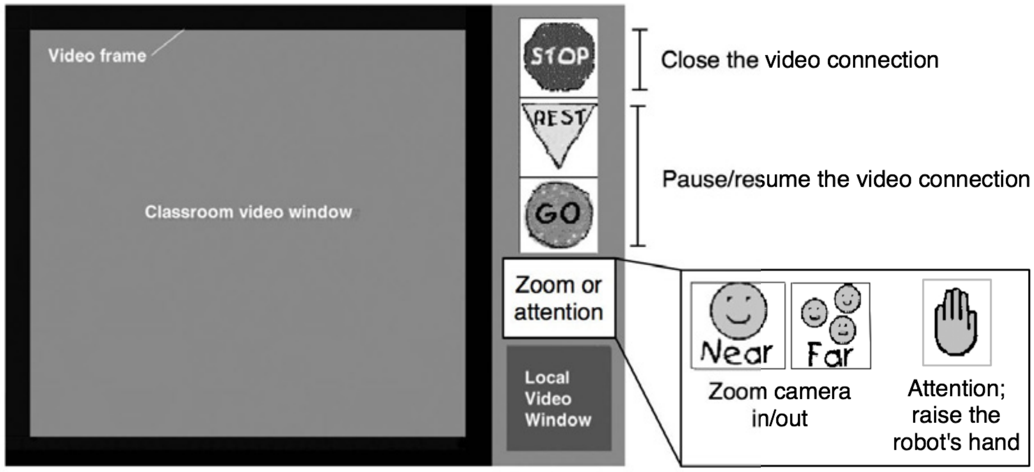


Figure 2-13: The PEBBLES II video conferencing interface that runs on the local system [Fels et al., 1999].

PEBBLES's hand can be "raised." For signaling attention and zooming the camera, the corresponding icon appears above the local video window when the controls are pressed (Figure 2-13). The child can pause and resume the video connection with his or her classroom, and at the end of the school day, close the connection between the robots. It should be noted that there are no controls on the interface to drive the robot. The PEBBLES robot is a passive mobile system and the robot operator is unable to change the robot's location independently; an attendant must push the robot from one location to another.

In 2011, the media broke the story of Lyndon Baty's using a VGo Communications' VGo robot to attend his classes; we described the VGo robot and its user interface in further detail in Chapter 3. At the time, Lyndon was a high school freshman in Knox City, TX, who has polycystic kidney disease [Robertson, 2011]. He received a kidney transplant at age 7, but when Lyndon was 14, his body began to reject it [Richards, 2011]. Lyndon stayed at home as per recommendation of his doctors so that he would not become sick. After a year of being home-schooled by a tutor, he instead attended school using his "Batybot" and can drive from classroom to classroom [Ryden, 2011]. When he has a question, Lyndon flashed the lights surrounding the Batybot's forward facing camera to get attention [Ryden, 2011]. Lyndon's mother Sheri said that "the VGo has integrated Lyndon back into the classroom where he is able to participate in classroom discussions and activities as if he were physically there. More importantly, the VGo has given back his daily socialization that illness has taken away" [Richards, 2011].

News of disabled students using robots to regain their presence in the classroom like Lyndon has become increasingly frequent. Since Lyndon first started using his robot in 2011, VGo has sold approximately 50 for use in the classroom [Massie, 2014]. Dozens of other students worldwide also go to class via their dedicated telepresence robots:

- In Moscow, Stepan Sopin used an R.BOT 100 during his recovery from Leukemia [Bennet, 2010; Hornyak, 2011b].
- Fourteen year old Lauren Robinson attended her sophomore year of high school in Fort Collins, Colorado, though her VGo telepresence robot; she had spent her freshman year homebound due to a severe dairy allergy [Hooker, 2011].
- Nick Nisius, of Janesville Community School in Iowa, began attending class in 2008 via Skype due to muscle weakness from his Duchenne muscular dystrophy [Paxton, 2011]. A teacher had hand carried Nick's laptop from classroom to classroom, but for his senior year, he used a VGo robot to join his peers between classes in the hall, at lunch, assemblies, and other events at school.
- Second-grader Aiden Bailey has had two lung transplants. Due to his suppressed immune system, he began attending school using Skype in first grade [Wiedemann, 2012]. Aiden now uses a VGo telepresence robot, purchased by the Iowa Edgewood-Colesburg school district [KWWL, 2011].
- Cris Colaluca, a student with spina bifidia in New Castle, Pennsylvania, attends his seventh grade classes at Mohawk Junior High School from his bedroom using a VGo robot [Weaver, 2012].
- Zachary Thomason of Arkansas, age 12, has extreme muscle weakness due to Myotubular Myopathy and uses a ventilator to breathe; he uses a VGo telepresence robot to go to class [Hornyak, 2012].

- Fourth-grader Paris Luckowski used a VGo telepresence robot to attend school in Newark, New Jersey, while in treatments for a brain tumor [Verizon, 2012]
- Sixteen year old Evgeny Demidov also attends his classes at Moscow School 166 through an R.BOT 100 due to his heart condition which keeps him in his home [Sloan, 2012; Wagstaff, 2012].
- Kyle Weintraub attends class in Florida while being treated for cancer in Philadelphia [Fishman, 2014].
- Thirteen year old Cookie Topp uses a VGo, purchased by her father, to attend St. Jude the Apostle School in Wauwatosa, Wisconson [Runyon, 2013].
- Lexie Kinder, age 9, operates her robot Princess VGo at her Sumpter, South Carolina school [Brown, 2013; Massie, 2014].

Future patients of the Primary Children's Medical Center in Kearns, Utah, will also be able to send a VGo telepresence robot to their regular classrooms during their stay [Winters, 2012].

Telepresence robots have also been purchased by school districts for students recovering from short-term illness, injury, or surgery [Ebben, 2014; Farrand, 2014; KWWL, 2011; VGo Communications, 2014; Washington, 2014]. The Texas Region 6 Education Center established the Morgan's Angels program and recently acquired 21 VGo robots [Region 6 Education Service Center, 2014b; Vess, 2014]. The program was named after Morgan LaRue, age 6 [Duncan, 2012]. Through the Morgan's Angels program, Daisie Hilborn, a Montgomery High School ninth grader, and Rylan Karrer, a Montgomery Intermediate fifth grader, use VGos in their schools [Lopez, 2013]. Elementary school student CJ Cook received his VGo robot from the Mogran's Angles program, which he named "Blue Deuce" [Reece, 2013].

In Texas, Baylor University, the Education Service Center Region 12, and the Texas Education Telecommunications Network (TETN) are working together to enrich education for their K-12 students [News Channel 25, 2011]. Baylor University hosts a VGo telepresence robot that students can use for virtual field trips. For example, the telepresence robot could provide remote access to the Baylor University's Armstrong Browning Library, which features illustrations and a stained-glass window of Robert Browning's poem "The Pied Piper of Hamelin." Another VGo robot "Millie" will act as a docent at the George Bush Presidential Library and Museum in the World War II exhibit [Region 6 Education Service Center, 2014a].

2.3 Summary

The majority of the work done in the field of telepresence relating to healthcare has focused on use in the hospital or in the residence of a person with special needs. As such, the design of the robots and their interfaces has been focused on the doctor, healthcare staff, or family caregiver accessing the client's file and operating the robot. Three robots have been used by people with special needs to remotely attend school. However, PEBBLES must be pushed from one room to another, thus its interface did not need to include navigation. The VGo robot used by Lyndon Baty and others can be independently driven from classroom to classroom. They use the standard VGo control software (detailed in Appendix A.1), which was designed for use in small and medium business scenarios; the interface is primarily the robot's video with a robot status bar below. Lyndon learned to navigate his new school using a paper copy of a fire drill map [Campbell, 2011], and Cris Colaluca put a hand-drawn map of his school above his desk for when he first started to use his robot [Weaver, 2012]. Occasionally, Evgeny Demidov bumps his R.BOT 100 robot on the doorway into his classroom due to the camera's limited field of view [Sloan, 2012]. Second-grader Aiden Bailey and his teacher both share control of moving his VGo robot around in his 19-person classroom [Wiedemann, 2012]; additionally, his teacher can use the VGo's infrared remote control to position Aiden's robot. In Chapter 5, we discuss how an interface can be designed for a telepresence robot so that a person with special needs can operate the remote robot.

Chapter 3

Margo: System Design and Architecture

"Interactants" and "bystanders," interacting directly and indirectly with a telepresence robot, may not be receptive to robot designs that do not share human-like characteristics. User task engagement and sense of telepresence may be degraded if interactants and bystanders are unwilling to communicate with the user. Unlike virtual reality, in which all users have the same capacity to create bodies, it is imperative for interactants and bystanders to accept the telepresence robot as a representation of the user. Consequently, the robot must be able to function sufficiently well as a human proxy. The degree to which interactants and bystanders feel the user is telepresent in the remote environment is, therefore, dependent upon the quality of their interaction with the telepresence robot embodiment.

We assert that a telepresence robot system has two components:

The robot: A physical embodiment in a real environment that is remote from the user with the capacities to independently affect the environment and interact with people in it.

The interface: Displays and/or controls with the ability to relay relevant sensor information to the user and support interaction with people in the remote environment and the environment itself.

Tables A.1 and A.2 in Appendix A summarize principles, requirements, and design guidelines from four HRI research groups investigating mobile telepresence robots for social interaction, including ones from our early work. Two key insights resulted from our early work [Desai et al., 2011; Tsui et al., 2011a, b]. First, a wide field of view is needed to operate telepresence robot, both horizontally and vertically. Second, some level of autonomous navigation is required. Direct teleoperation is impractical due to inherent network latency and the movement of people in the remote environment.

To facilitate the investigation of our research questions, we augmented a VGo Communications' VGo robot. In this chapter,[1] I discuss the augmentation requirements and constraints, the resulting design, and an overview of the corresponding software.

3.1 Base Platform

We selected the VGo Communications' VGo robot as the base for our system as it has a sophisticated audio and video communication system.[2] The VGo App is VGo Communications' video conferencing software. It supports both robot calls (i.e., from a laptop/desktop computer to a robot) and also desktop calls (i.e. between two laptop or desktop computers). The user interface is primarily a view of the robot's live camera stream with a small video of the user in the top right corner (Figure 3-1 right).

The VGo robot retails for $6,000 USD plus an annual $1,200 service contract. The VGo robot is four feet tall (48 in, 121.9 cm) and weighs approximately 18 lbs (8.2 kg) including a 6-hour lead acid battery [VGo Communications, 2012a]; a 12-hour battery is available. It uses two wheels and two rear casters to drive; each drive wheel is 7 inches (17.8 cm) in diameter and has an encoder. Its maximum speed approaches human walking speed, approximately 2.75 mph [VGo Communications, 2012a]. Additionally, the VGo's base has a front bumper and four infrared (IR) distance sensors. There is one IR distance sensor centered in front, and one on either side of the front (on the left and right); these are primarily used to warn the user about

[1]Portions of this chapter were published in [Desai et al., 2011; Tsui et al., 2013a, 2014].
[2]Synthesized from [Tsui et al., 2011a; VGo Communications, 2012a, b, c] and our own robot use.

Table 3.1: Key feature summary of VGo Communications' VGo robot.

Unit cost	$6K plus $1,200 annual service contract
Drive	2 wheels and 2 trailing casters
Wheel size	7 in (17.8 cm) diameter
Wheel base	12 in (30.5 cm)
Top speed	2.75 mph (4.4 km/h)
Height	48 in (121.9 cm) fixed
Weight	• 18 lbs (8.2 kg) with 6 hour battery • 22 lbs (10.0 kg) with a 12 hour battery
Battery type	Sealed lead acid battery, 12V
Auto charge	Auto-dock within 10 feet (3.0 m) of docking station
Microphones	4 around video screen (1 front/back pair on each side)
Speakers	2 (woofer in base, tweeter in head below the screen)
Screen size	6 in (15.2 cm) diagonal
Number of cameras	1 forward facing webcam with 2.4 mega-pixels, located above the screen
Camera pan-tilt	No independent pan. Yes, 180 degree tilt
Connection type	• WiFi (802.11b/g 2.4GHz, 802.11a 5.0GHz) • 4G LTE (requires separate contract and sim card)
Bandwidth	200kbps up to 850kbps (up and down); recommended 1.5Mbps (up and down)
Operating systems	• Windows 7/Vista/XP with SP3 • MacOS 10.6.x or higher (in beta)
Navigation control	• Mouse "Click and Go" widget • Arrows keys with customizable acceleration profile • Proportional joystick widget (in beta)

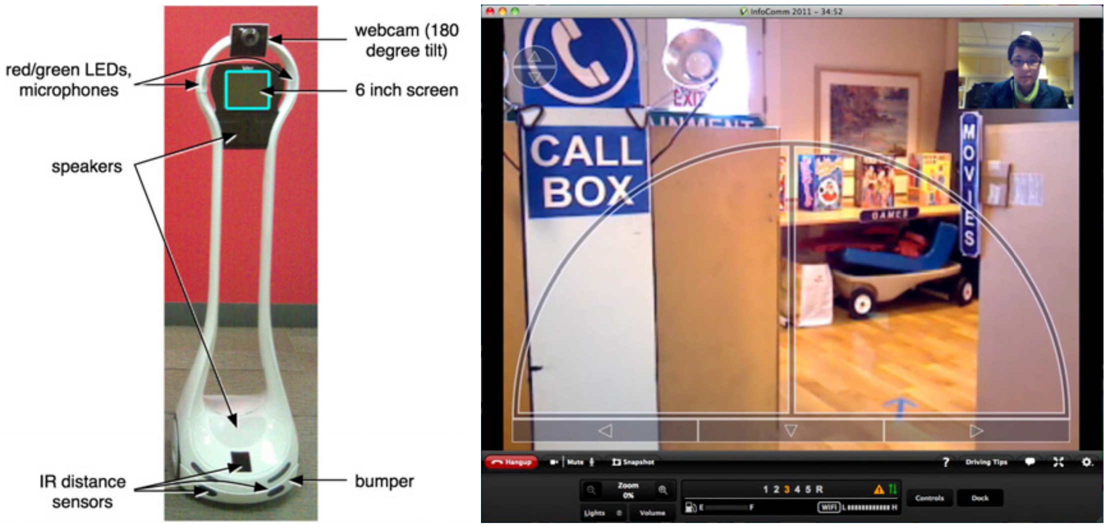


Figure 3-1: **(Left)** VGo Communications' VGo Robot. **(Right)** The VGo App's robot call user interface, seen here in an example of what the user sees controlling an unmodified VGo robot, highlighting the difficulty of single camera video based driving; screenshot from Aug. 2012. Additional details regarding the VGo App are located in Appendix A.1.

obstacle detection. The fourth IR distance sensor is located in the rear and assists with docking on the charging station. A standard VGo robot is shown in Figure 3-1 (left), and its specifications are listed in Table 3.1.[3]

The overall appearance of the VGo robot is pleasing. The landscape-oriented screen is encompassed by a ring of black plastic and thus resembles a head [DiSalvo et al., 2002]. The tweeter speaker makes the robot operator's voice appear to come from the head as a person collocated with the robot would expect. Its height is that of a small person (48 in or 121.9 cm), which Lee et al. [2009] note to be on the slightly small side of "just right." VGo's body has a slight curve and is covered completely with a white, lightweight plastic [DiSalvo et al., 2002]. Its iconic appearance resembles Eve from the Disney/Pixar film *WALL-E*, yet remains gender-neutral [Fong et al., 2003; Parlitz et al., 2008].

[3]It should be noted that some material properties and features of the VGo robot have changed since our first use in Tsui et al. [2011a]. We have augmented two robots, detailed in [Tsui et al., 2014]. The VGo base platforms for Hugo (v0) and Margo (v1) were acquired in Fall 2010 and Fall 2011, respectively. Specifically, Hugo and Margo had only one WiFi card and do not support 4G. Also, their plastic bodies were stronger than the alpha prototypes used in Tsui et al. [2011a], and Margo's drive wheels featured a softer rubber. For the purposes of this chapter, we describe our v1 robot, Margo.

Table 3.2: Hardware Design Requirements

R1	Retain all of VGo Communication's existing features
R2	Utilize existing bidirectional audio and video communication and velocity control
R3	Robot must map an unknown environment
R4	Robot must localize within the map
R5	Advanced sensing for interacting with bystanders
R6	Additional components must be commercial off-the-shelf (COTS) components, and supported by the robotics community
R7	Robot run time must be at least 2 hours
R8	Robot must maintain friendly appearance
R9	Robot must maintain its stock stability
R10	Robot must have a dedicated camera for driving
R11	Robot must have forward and downward facing camera views

3.2 Requirements

We wished for our augmentations to function seamlessly with the stock platform. First and foremost, we needed to retain the use of all or most of VGo Communication's existing features (Requirement 1, R1). In particular, we needed to utilize the robot's bidirectional audio and video communication system (hardware and software), as well as the built-in velocity control (R2).

In order to investigate autonomous and semi-autonomous navigation behaviors, we needed the robot to be able to map an unknown environment (R3) and perform basic localization (R4). We also needed advanced sensing for interacting with bystanders who may be asked to provide navigation assistance (R5).

Additional components had to be commercial off-the-shelf (COTS) components and well supported by the robotics community (R6). The augmented robot's run time must be at least 2 hours (R7). For all augmentations, we needed to retain the robot's friendly appearance (R8) and maintain the stability of the robot given its center of gravity (R9) [Tsui and Yanco, 2013].

3.2.1 Additional Sensing

The standard VGo robot has one camera above its screen, which can be tilted down for navigation assistance. However, given its limited field of view, we found that it is not possible to simultaneously use the camera for conversation and navigation [Desai et al., 2011; Tsui et al., 2011a]. Thus, we needed the robot to have a dedicated camera for driving (R10).

Research has shown that a view of the robot's base can provide scale and a static and physical real-world frame of reference to the user [Keyes et al., 2006; Mine et al., 1997]; a user can consciously map the robot embodiment to include an object or image [Carruthers, 2009]. Therefore, the user's view should include both forward looking and downward looking perspectives in one continuous view (R11).

The addition of a laser range finder and IMU was required to achieve a particular level of performance for navigation based activities such as environment mapping, obstacle avoidance, and map based localization techniques. Proximity or low precision distance sensing behind the robot was also required (R12). Finally, the robot needed sufficient computational power to processes these additional sensors (R13).

3.3 Design Constraints

With all requirements in place, there were several design constrains for component placement. The first constraint (C1) was that all additional components must fit within the footprint of the standard VGo robot footprint. This meant that the majority of the components must fit within the volume of the vertical stocks (approximately 4in wide $\times$ 23.6in tall $\times$ 3.4in deep; 104mm wide $\times$ 600mm tall $\times$ 87mm deep), atop its head, or on its base. Table 3.3 lists all of the constraints for additional components.

3.3.1 Power

Although the individual power sources for the robot could be used to simultaneously power both the stock platform and augmentations, VGo's managed power system

Table 3.3: Robot Design Constraints

C1	Robot must retain its stock footprint
C2	Robot must retain its stock power and charging systems
C3	Industrial design must be maintained
C4	The laser range finder must be positioned on the front of the base and have the widest possible field of view
C5	The IMU must be positioned at the center of rotation and parallel to the ground
C6	The downward facing camera must be mounted as high as possible
C7	All additional computing power must be incorporated into the robot

inside the robot was only suitable to provide power for the stock robot and could not support the augmentations. Adding separate additional power sources increases the weight, reduces the form factor, and increases the complexity of using the system. Therefore, the vast majority of our augmentations needed to be powered alongside the stock robot using the on-board large capacity 12V 15Ah lead acid battery, stock wall charger, or robot dock charger (C2). With two parallel power systems, power related activities such as switching between power sources and powering up or down needed to be integrated into to the stock platform without disrupting the standard VGo system.

3.3.2 Space and Component Positioning

The additional components needed to fit within the VGo's footprint (C2) in a manner that maintained its streamlined industrial design (C3).

However, certain sensory components require specific position that also had to be accommodated. The laser range finder need to be positioned near the ground, near the front of the robot, with as wide a field of view as possible (C4). The IMU needed to be mounted in the center of the robot's axis of rotation, and parallel with the ground (C5).

Cameras needed to be positioned as high as possible on the robot (C6). Downward facing cameras must be mounted at the highest position possible, thereby allowing the widest field of view of the area surrounding robot's base for navigation purposes.

Forward facing cameras must be mounted at an appropriate height in order to capture useful and interesting data from interactions. As the height of the VGo robot is only 48 in (121.9 cm), all the forward facing cameras needed to be mounted at the highest position possible on the VGo robot – atop the robot's stock camera.

3.3.3 Processing

Due to the vast amount of information produced by modern sensors, a robot's mobile nature, and limitations in wireless technology, all sensor processing needed to be performed on-board the robot (C7). The embedded computer robot was neither physically or computationally suitable for supporting more than one camera. Additional computation power had to be incorporated into the robot design to provide the hardware interfaces for the sensors and run essential logic needed for communicating with the augmented system.

3.4 Sensing and Processing Augmentations

3.4.1 Navigational Sensors

A Hokuyo UGH-08 laser ($3,500) and a MicroStrain 3DM-GX3-45 inertial measurement unit (IMU, $3,800) were used for navigation purposes [Park and Kuipers, 2011]. An array of six Sharp 2Y0A02 IR distance sensors ($12 each) provided cursory information about the region behind the robot.

3.4.2 Cameras

Three additional cameras were added to the robot. We chose to incorporate an Asus Xtion Pro Live ($170), which provides a dedicated forward facing, color video camera and 3D information from an IR painter/camera pair. This camera can be used for capturing the interactions of people physically present with the robot. The Asus camera and had several benefits over a Microsoft Kinect. It required only 1 USB 2.0 port, thereby reducing the power consumption. The Xtion had higher image resolution and a larger field of view than the Kinect [Gonzalez-Jorge et al., 2013]. Its form factor was much smaller than the Kinect; in fact, the internal circuit board and heat sink was only slightly larger than the Kinect's camera board. The other two cameras were Logitech C910 webcam ($70), one of which provided a downward looking view of the area around the robot's base [Kirbis, 2013], and the second overlapping the other two cameras (Figure 3-3).

3.4.3 Processing and Interfaces

Our design included a fitPC-2 [CompuLab, 2012], running Ubuntu 12.04 LTS and ROS fuerte for interfacing with the augmented system and stock robot. The fitPC-2 ($400) has an Intel Atom Z550 2GHz processor and 2GB RAM. It was chosen for its small size (4 in; 10.2 cm $w \times$ 4.5; 11.4 cm in $h \times$ 1.05 in; 2.7cm d), wide power range (8-15V), and efficient power consumption (0.5W standby, 8W full CPU). It has an onboard Mini PCI-E 802.11b/g WiFi, a DVI port, an ethernet jack, audio input and output jacks, and 6 USB 2.0 ports. The fitPC-2 connects to the VGo robot using two USB-RS422 adapters. A 2×20 character Phidget display and adapter show status information (e.g., power levels, wifi strength). A heat sink and cooling fan dissipated residual heat. We selected the Evercool EC8010LL05E ($10) fan for its airflow (14.32 CFM) given its minimal noise level (<22 dBA) power consumption (5V at 0.25A) [Evercool USA, 2010]; the majority of the heat sink was covered by this fan (3.15 in; 8 cm $w \times$ 3.15 in; 8 cm h), with a minimal increase to the profile (0.39 in; 1 cm d).

There is one 4-port USB 2.0 hub, and a Phidget 1019 InterfaceKit (IFK) with a 6-port built-in USB 1.1 hub. The Phidget 1019 IFK can read in 8 analog and 8 digital signals and also drive 8 digital outputs. The Sharp IR distance sensors interface with the Phidget 1019 IFK using the corresponding Phidget 1101 analog adapters.

To accommodate the large amount of onboard processing, we incorporated Lenovo x230 laptop ($1,500) with an i7-2620M processor and 8GB RAM running, which also ran Ubuntu 12.04 with ROS fuerte. The laptop is attached to back of the robot using a U-shaped mount through the robot's rear battery compartment and velcro straps (Figure 3-6). The two computers were connected together using gigabit ethernet.

Two of the cameras are connected directly to the laptop to prevent the USB bus from being overdrawn - according to the USB 2.0 standard, only 80% of the maximum bandwidth can be allocated since traffic may be bursty [Ideas On Board, 2012]; allocation for a video device is based on image resolution and does not account for frame rate. Since the fitPC-2 unfortunately has only one USB 2.0 root hub, some of the sensors must be connected to the laptop. The Lenovo x230 has three USB root hubs: one USB 2.0 connected through its docking station, one USB 2.0 and one USB 3.0 in the laptop itself.

3.5 Platform Augmentations

The majority of these components were mounted on three custom 6.35mm inch acrylic panels, located in the vertical space between Margo's stalks; Figure 3-2 shows an expanded and side view of the panels. To prevent hardware from falling into the robot's internals potentially causing damage to the system, we developed a technique in which hex nuts were held captive between two 0.0625 inch (1.59mm) acrylic layers, bonded on either side of the 0.25 inch (6.35mm) panel using a solvent cement (Weld-On 4). We perforated this layer to reduce the weight of the middle and back acrylic panels. We bonded acrylic O-ring risers to also prevent damage to exposed circuit boards due to over-tightening when (re)attaching the components to the panels.

The middle panel was first mounted between Margo's stalks, and the front and back panels then mounted to it. Therefore, the middle plane had to be as stable as possible. We placed three crossbars between the VGo's two stalks with one just below the head, one toward the bottom of the stalks just above the VGo's latching mechanism, and one between them. The crossbars were T-shaped; the top and bottom crossbars were inserted from the same stalk, and the middle from the other stalk to minimize horizontal movement of the middle plane when attached. The largest and heaviest component was the fitPC-2 and drove the layout of components. We mounted the fitPC-2 toward the bottom of the middle plane between the middle and bottom crossbars to maximize its stability and that of the robot overall. To prevent horizontal

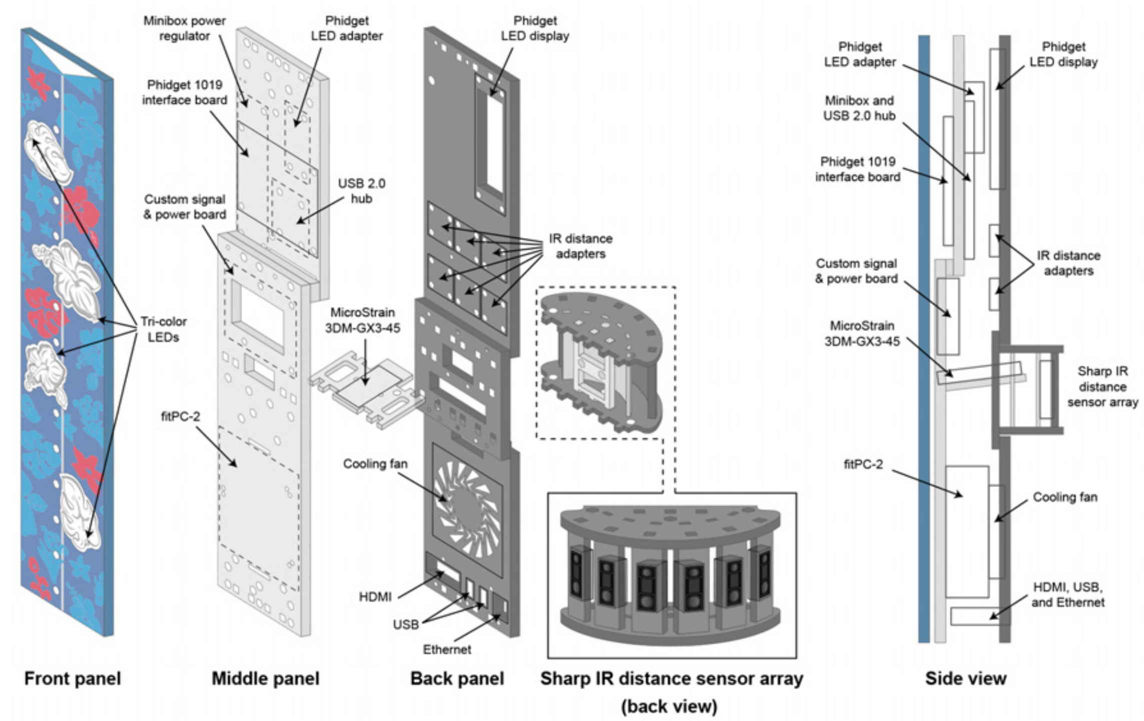


Figure 3-2: Expanded and side view of the augmentation panels

and vertical movement, we emulated a VESA mount using three M2 screw pairs; a velcro strap was used to secure the fitPC-2 to the middle plane.

3.5.1 The Hat

The three cameras are housed in a "hat" which sits above the VGo's servo tilt camera. To minimize the size of the hat, we removed the cameras' plastic exteriors. Functional prototypes were rectangular in shape. However, the rectangular shape detracted from the VGo robot's curved body design, and we increased the height of the hat to incorporate an arch to soften the harsh line (Figure 3-3). The final version of the hat adds approximately 6 inches (15.2 cm) to the robot's height, increasing its total height to a socially comfortable 54 inches (137.2 cm) [Lee et al., 2009].

3.5.2 Power System

The power system had three components. First, we used a Minibox DCDC-USB buck-boost power supply ($60) which regulated the power directly from the 12V 15Ah

Figure 3-3: Margo's hat design features three cameras rigidly mounted together. One Logitech C910 is mounted below the Asus Xtion, and the other above.

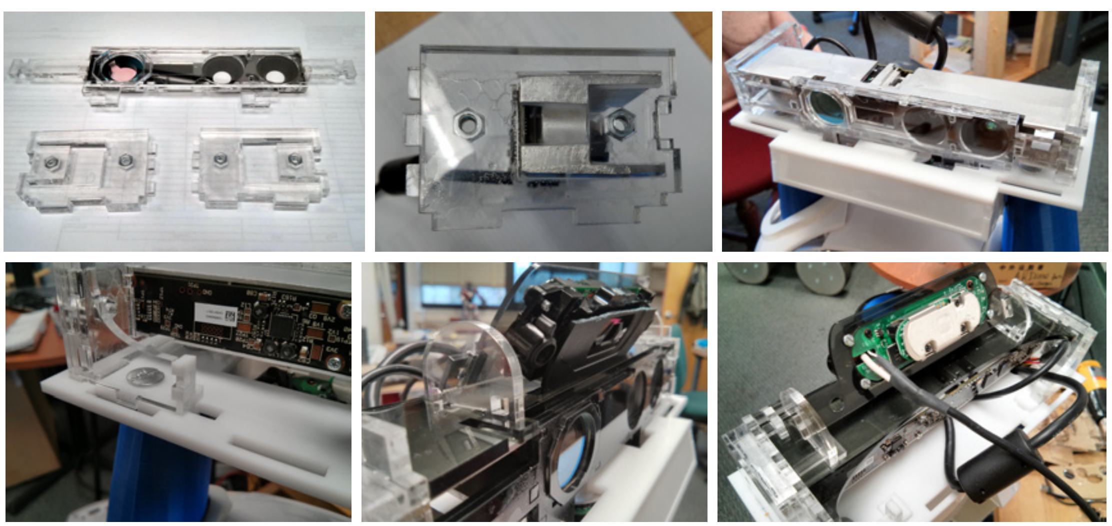


Figure 3-4: Assembly process of the Asus Xtion and upper Logitech C910 webcam for the hat design.

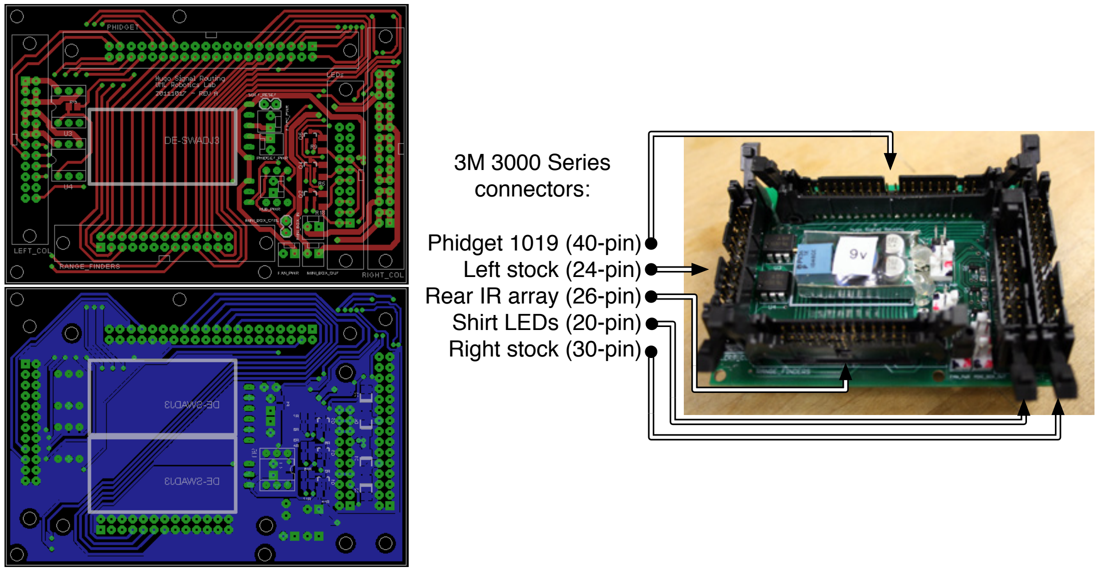


Figure 3-5: A custom signal and power routing board was designed to cleanly access the robot's stock power and control systems, as stated in Table 3.2.

battery for the laser. We designed two single-layer, double-sided custom printed circuit boards.[4] The signal and power routing board (3.5 in; 8.89 cm $w \times$ 2 in; 5.08 cm h, Figure 3-5) bridged the wiring between the separate components in the robot's left stalk, right stalk, three acrylic panels located between the stalks, the base laser, and the hat containing the Asus camera and Logitech webcams. To prevent mistakes in (re)assembling the robot, we used flat ribbon cable (3M 3302 series) with keyed, locking connectors (3M 3000 series) and strain relief; additionally, each connector on the signal and power routing board had a unique number of pins. Three adjustable step down switching regulators (model DE-SWADJ 3, $25 each) provided power to the fitPC-2 (9V), the Phidget 1019 IFK (9V), and the cooling fan and the USB 2.0 hub (5V); each of these 2-pin power connectors were also keyed and locking.

A second custom power-OR board located in the base powered our augmentations from the battery, dock charger, or wall charger, whichever was the highest voltage power source. This power-OR board also featured a relay, which we triggered when the VGo's power button was pressed so that our augmentations would only receive power when the VGo was powered on.

[4]Design by Chris Granz formerly of UMass Lowell.

3.5.3 Emergency Stop

We determined there was a need for an emergency shut off for our robot to ensure safety in operation among humans, and therefore we augmented Margo with a software emergency stop (e-stop) on the top, back of the robot (Figure 3-6). When the black lanyard is pulled, the red e-stop containing a magnet releases contact with the switch (Phidget 3560, BR-1014), which is connected to a digital input on the Phidget 1019 IO board. `/IFK1019` node was monitoring the board. It started publishing messages to `/estop_physical` topic indicating that the emergency stop was pulled as soon as the switch was triggered. We call it a "software" e-stop because it was implemented inside `velocity_manager` node. If the software e-stop is engaged, `velocity_manager` overrides all velocity commands sent to the motor with zeros.

The `velocity_manager` node also filtered the velocity commands before they were sent to motors. We recognized that the jittery movements of the robot were not pleasing features and we determined the need to control robot's acceleration. `velocity_manager` node had three profiles: one for autonomous acceleration, one for gamepad controlled acceleration, and one for deceleration. All velocity commands were processed by the `velocity_manager` before being used to control the motors. This ensured there were no sudden changes in speed even if the commanded velocity was not continuous.

3.5.4 Caster Replacement

Due to the significant number of augmentations made to the robot, the weight began to deform the outer plastic casing on the right caster, causing it to bind when transitioning between driving straight and turning in place. We replaced the existing, soft, rubber-wheeled casters with steel ball-bearing casters (Figure 3-7). This augmentation highlights a tradeoff - the stock casters attributed to the overall audio quality (i.e., very quiet) but Margo was unable to autonomously navigate. The latter are louder [Desai et al., 2011; Tsui and Yanco, 2013] but allow the robot to move; fortunately, VGo's video conferencing software can choose to use only the forward

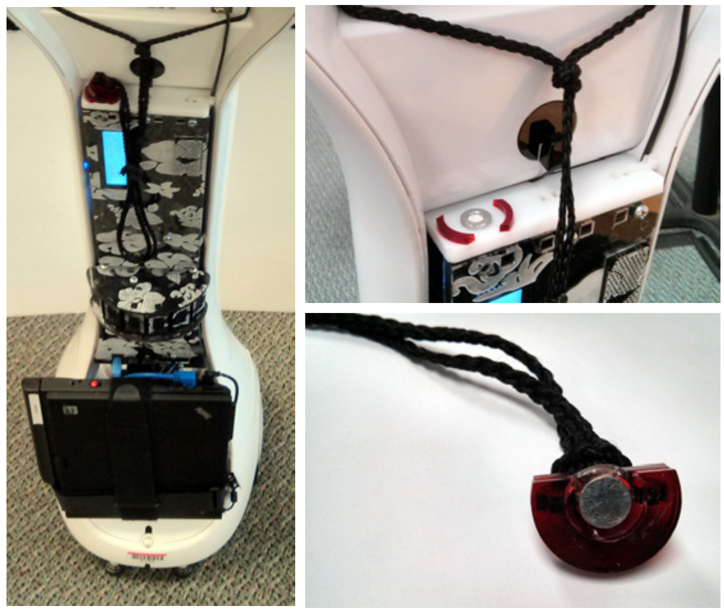


Figure 3-6: Margo emergency stop and laptop augmentations. The e-stop engages when the lanyard is pulled, similar to a snowmobile or jetski throttle kill ripcord.

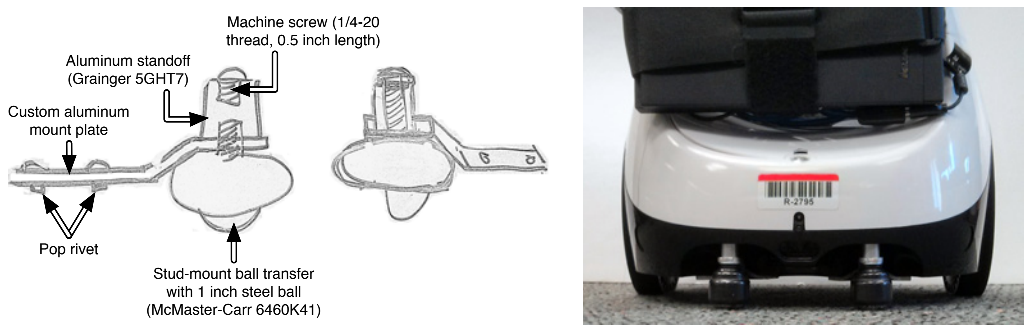


Figure 3-7: Margo rear caster replacement featuring stud-mount ball transfers.

facing microphones.

3.5.5 Appearance

To people physically present with the robot, it appears to only have one function – mobile video conferencing. It was important to mount the additional sensing and computation in a hidden manner to maintain the VGo robot's aesthetic design. Robots with mechanical appearances are viewed as aggressive and angry [Woods et al., 2004], and an overly sophisticated appearance may lead to underwhelming interactions [Blow et al., 2006]. Since human-like robots are preferred for social roles, we "dressed" the

Figure 3-8: Margo (leftmost robot) exhibiting the front view of our resulting design with red LEDs lit in the shirt. View of the back panel on a sibling robot, Largo (rightmost robot).

front area below the robot's screen with a Hawaiian shirt which implies that the robot has a fun, friendly personality [Goetz et al., 2003] and appeals to a large audience [Parlitz et al., 2008]. The shirt is bright blue [Woods et al., 2004] and decorated with tropical flowers. Four flowers are clear with inset RGB LEDs which will be used to indicate state. For example, green may indicate that the robot is currently in use, which matches with VGo's "in call" status. The gender-neutral Hawaiian shirt motif is suitable for both male and female operators [Lee et al., 2009; Parlitz et al., 2008]. Figure 3-8 shows Margo (leftmost robot) with red shirt LEDs light.

The back of the robot is designed for an administrator's use only. The back panel and the IR sensor array are cut from translucent, dark gray Plexiglas and the Hawaiian shirt pattern is repeated in white (Figure 3-8 rightmost robot). Our goal was to deemphasize the partially exposed components, namely the fitPC-2's cooling fan, the IR sensor array, and the administrator's access to the internal system. We incorporated the Phidget 2×20 character display into the back plane at the top of the shirt. Below the IR array, we exposed the fitPC-2's USB, ethernet, and HDMI ports and a recessed soft reset button for debugging purposes.

We recess mounted the laser in the forward center of the robot's base over the base speaker (Figure 3-9). Additionally, the laser's formerly orange cap has been covered with white silicon Sugru to minimize it visually. The UGH-08 laser has a 240° field of view; however, the VGo's vertical stocks block the laser's view such that it only returns meaningful values for the front 180°. We designed an array of six Sharp IR distance sensors to provide cursory information about the space behind the robot (Figure 3-10). The IR array bustle extends from the back panel but remains within the form factor of the VGo's base (Figure 3-8 rightmost robot). The Sharp IR distance sensors have a range of 8 to 60 inches (20 to 150 cm).

3.6 Software

Both the fitPC and Lenovo laptop run Ubuntu 12.04 LTS and ROS fuerte [Willow Garage, 2012]. The use of ROS is prevalent among academic and industry researchers

Figure 3-9: **(Left)** The Hokuyo UGH-08 laser is mounted in the forward center of the VGo robot's base over the base speaker. **(Right)** The laser's formerly orange cap has been covered with white silicon Sugru to visually minimize it.

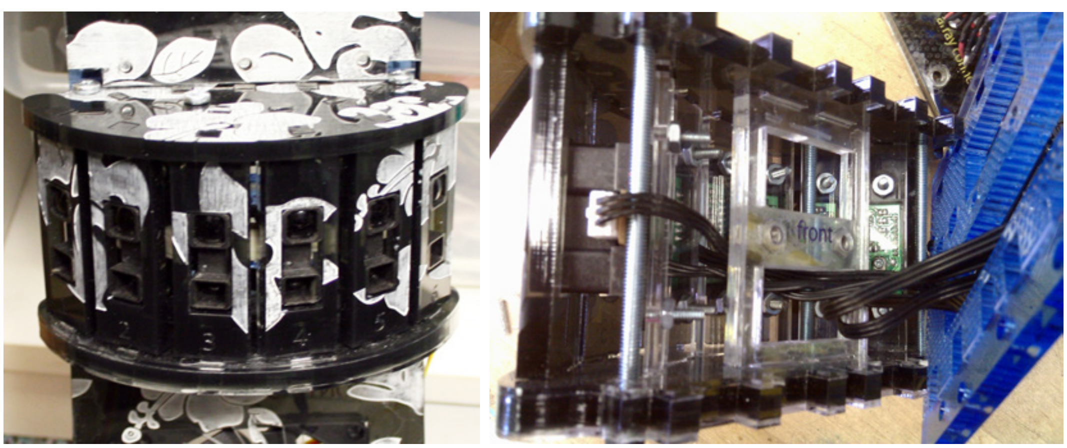


Figure 3-10: Rear IR array bustle with six Sharp 2Y0A02 sensors. **(Left)** External view. **(Right)** Internal view.

and, furthermore, is well suited for the complexity of our system. We are able to leverage the distributed architecture and modular communication ROS provides. We have built upon existing ROS stacks and packages contributed by the community, thereby allowing the rapid development of robot specific code for the VGo robot. Figure 3-11 provides an overview of Margo's software. The core system shown in Figure 3-12 is described below. We discuss the hat camera and movement subsystems in detail in Chapters 5 and 6, respectively.

The fitPC runs the ROS master and oversees low-level function such as motor controls and digital I/O. The software was distributed between the two machines so that all ROS nodes using image processing launched on the laptop, and the video stream from the down facing webcam showing the robot's base passed over the private interface. Additionally, the laptop was setup to host a ROSBridge webUI for user interface development [Mace, 2013].

3.6.1 Communication with the VGo Base

The VGo robot uses serial communication to allow the "head" and "base" to exchange information. We modified this connection by inserting a fitPC-2 between them. The fitPC-2 establishes serial communication with the base through a custom ROS node, called `vgo_serial_comm`. There are two modes in which the `vgo_serial_comm` node operates. First, the `vgo_serial_comm` node also has the ability to directly connect the VGo's head and base together by reading incoming packets and forwarding them to the appropriate destination. We refer to this as the serial passthrough mode, which can be set using a rosservice call to `/vgo_base/EnableSerialPassthrough`. The power to the base and head are reset via the Phidget 1019 IFK when the serial passthrough is enabled.

In the second mode, the fitPC-2 is able to directly communicate with the VGo base, which can be set by disabling the serial passthrough using a rosservice call to `/vgo_base/DisableSerialPassthrough`. The `vgo_serial_comm` node uses the VGo library[5] to enable serial communication, initialize the type of base sensor values it

[5]This serial communication library is property of VGo Communications and used with permission.

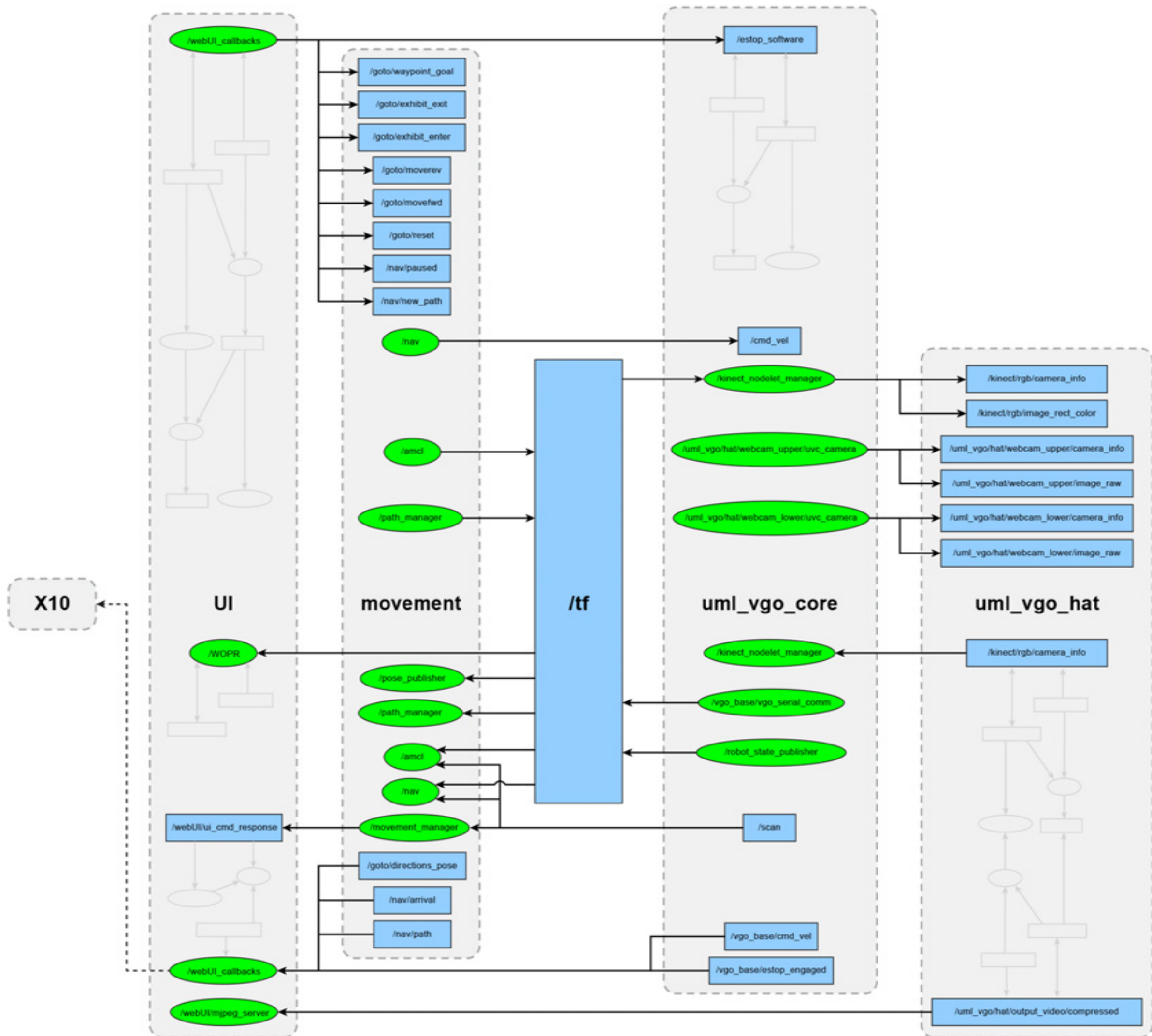


Figure 3-11: High-level system diagram showing ROS nodes in green and topics in light blue. Core subsystem diagram is shown in Figure 3-12 on p. 57. Camera subsystem diagram is shown in Figure 5-7 on p. 110. Movement subsystem diagram is shown in Figure 6-1 on p. 128.

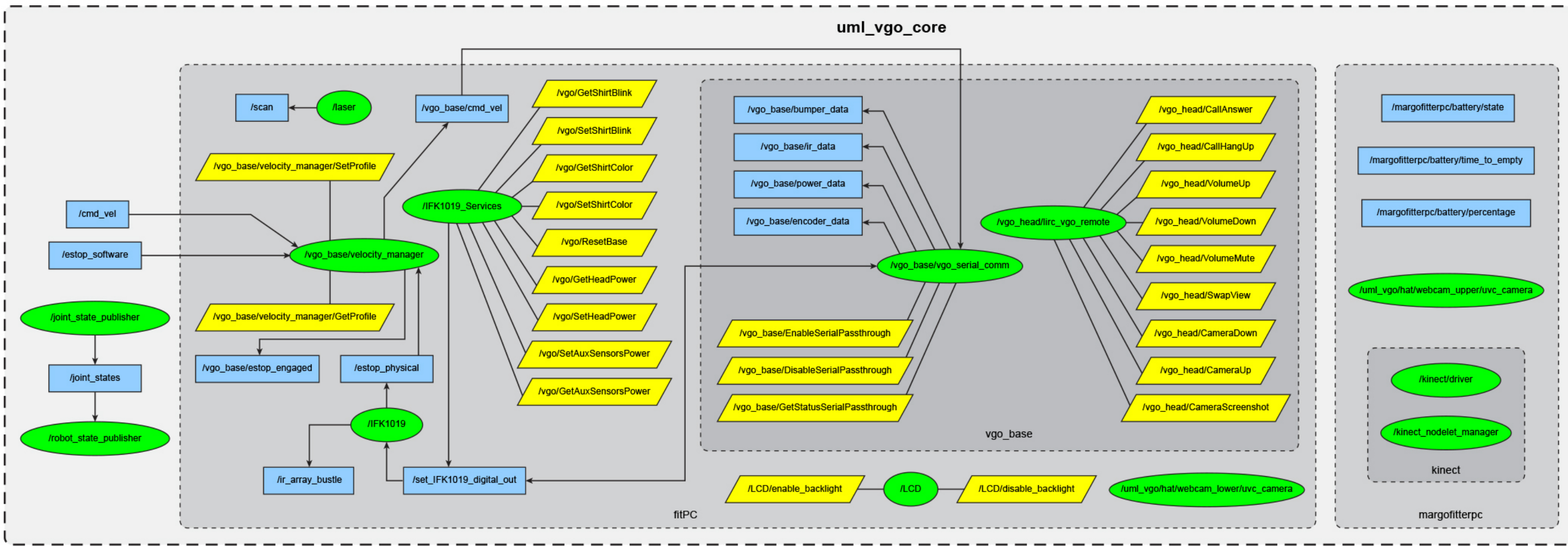


Figure 3-12: Core system component diagram showing ROS nodes in green, topics in light blue, and services in yellow. The full system diagram is shown in Figure 3-11 on p. 56.

should report, and set the frequency of reporting. In this mode, `vgo_serial_comm` publishes four topics:

- `/vgo_base/power_data`,
- `/vgo_base/bumper_data`,
- `/vgo_base/encoder_data`, and
- `/vgo_base/ir_data`.

Messages on these rostopics are published every 100 ms, and each topic has its own custom message.

The `/vgo_base/power_data` topic primarily provides the base's status with respect to power, such as if the robot is being charged, if the battery is indicating low power, if the base is being reset or shutdown. `vgo_serial_comm` monitors the `/vgo_base/power_data` status information to enforce a graceful shutdown for our fitPC-2 and augmentations when the robot is asked to reset or shutdown; otherwise, the fitPC-2 will crash when power is cut from the system. Supplementary information about the battery fuel gauge and current battery charge is also included in the `/vgo_base/power_data` message. `batt_fuel_gauge` is the percentage of battery life and is represented as an unsigned 16-bit integer with an expected range of 0 to 100. `batt_curr_mV` is the current voltage of the battery and represented as a 16-bit float with an expected range of 10,000 to 13,000 mV (10 to 13 V).

The VGo base has four Sharp IR sensors, and the data from these sensors are published to the `/vgo_base/ir_data` rostopic. One is located on the back of the robot; this distance sensor is used primarily for automatically docking the VGo robot. Another IR distance sensor is in the front of the robot facing upward, which is used to tell when the robot is approaching a tall obstacle such as a table that could not be sensed well by the laser. There are two IR distance sensors on the front left and right of the robot. Additionally, the VGo robot has two switch sensors in its front bumper, and the boolean states for the `left_bumper` and `right_bumper` are published on the `/vgo_base/bumper_data` rostopic.

The VGo robot features high resolution motor encoders. Values of the left and right motors are given as signed 16-bit integers and published on the /vgo_base/encoder_data rostopic. We have empirically determined that 5.18 encoder clicks is equal to a 1 degree turn of wheel. Odometry is published on the /tf transform rostopic.

To control the base's motors, vgo_serial_comm subscribes to the /cmd_vel rostopic and checks for new messages every 100 ms. /cmd_vel uses a ROS standard message Twist; the x value of the linear component and the z value of the angular component are used to set the translation and rotation of the VGo robot's motors.

3.6.2 Sensors

As previously described, we have enhanced our robot's sensing ability. A Hokuyo UGH-08 laser and a MicroStrain 3DM-GX3-45 inertial measurement unit (IMU) are used for navigation purposes [Park and Kuipers, 2011]. We utilize hokuyo_node package by Gerkey et al. [2011] to access our UGH-08 laser's values, and microstrain_3dmgx2_imu package by Leibs and Gassend [2011] for our IMU. These two nodes are instantiated as laser and microstrain_3dmgx2_imu_node, respectively, in Figure 3-11.

An array of six Sharp IR distance sensors provide cursory information about the space behind the robot. The Sharp IR sensors interface with the Phidget 1019 IFK using the corresponding Phidget 1101 analog adapters. Our IFK1019 node handles the connection to the Phidget 1019 board (also known as PhidgetInterfaceKit 8/8/8). It publishes the raw data from each of the six Sharp IR sensors to the /ir_array_bustle rostopic every 1000 ms.

An Asus Xtion and two Logitech C910 webcam are housed in a "hat" which sits above the VGo's servo tilt camera. The Xtion is used for capturing the interactions of people physically present with the robot and providing a fixed, forward facing view from the robot. We utilize the openni_launch unary stack by Rabaud [2012] for accessing the Xtion, and the gscam package by Jay and Crick [2012] to access the webcam's video. These nodes are instantiated as hat_kinect_camera and hat_down_webcam in Figure 3-11.

3.6.3 Emulating the VGo IR Remote Control

The VGo remote control activates user interface functions on the robot's head: answering and hanging up a call, tilting the VGo's camera up and down, turning the robot's volume up and down, muting the volume, and taking a picture. A USB IguanaWorks IR transceiver is used to emulate the VGo's IR remote control using the Linux Infrared Remote Control (LIRC) package [Bartelmus, 2011]. From our `lirc_vgo_remote` node, we provide rosservices for each of remote control functions: `/vgo/CallAnswer`, `/vgo/CallHangUp`, `/vgo/CameraUp`, `vgo/CameraDown`, `/vgo/CameraScreenshot`, `/vgo/VolumeDown`, `/vgo/VolumeMute`, and `/vgo/VolumeUp`. We emulate a button press by executing `irsend` calls at the command line and specifying the corresponding hex code; see Figure A-5 in Appendix A.2.

3.6.4 Localization

We utilize the `amcl` package by Gerkey [2011] for generating maps from the laser and the robot's encoder values. `amcl` subscribes to the laser's `/scan` rostopic and the `/tf` topic to which `vgo_serial_comm` publishes its odometry transforms. We also use `amcl` to localize on a given map with the robot's location in 2D space and orientation, respectively subscribed to as `/map` and `/initialpose` rostopics. When localizing, `amcl` publishes its 2D location and orientation on the `/amcl_pose` rostopic and its transform on `/tf`.

3.6.5 Manual Control

We utilize the `joy_node` package by Quigley et al. [2012] to interface with a USB gamepad; the node is instantiated as `gamepad` in Figure 3-11. `gamepad` publishes to the `joy` rostopic. The `sensor_msgs/Joy` message contains and arrays of the joystick's buttons and axes. Our custom `vgo_teleopJoy` rosnode subscribes to the `joy` rostopic. The linear and angular position of an analog joystick on the gamepad provide the basis for a proportional velocity `Twist` message. We are tuning the acceleration and deceleration to prevent rocking and sudden jolts with abrupt starts and stops; Margo's center of gravity has changed due to the augmentations above and below the head and increased weight. We scale both the translation and rotation values prior to publishing to /cmd_vel.

3.6.6 Auxiliary Infrastructure

We previously noted that our IFK1019 node publishes the values from the IR array to the /ir_array_bustle rostopic. It also publishes the boolean value of the reset button to the /soft_reset_button topic, and subscribes to the /set_IFK1019_digital_out rostopic.

The IFK1019_Serivces node is responsible for a number of rosservices including setting the robot's shirt color (i.e., /vgo/SetShirtColor), managing power to the robot base (i.e., /vgo/ResetBase) and head (i.e., /vgo/SetHeadPower), and managing the Minibox power to the laser and Kinect (i.e., /vgo/SetAuxSensorsPower). These services instantiate custom /vgo_msgs/IFK messages which are set with a Phidget 1019 IFK port and boolean state. The messages are then published to the /set_IFK1019_digital_out rostopic.

A Phidget 2×20 character display and adapter show status information (e.g., power levels, wifi strength). The LCD nodes subscribes to the /LCD/set_text and /LCD/set_brightness rostopics.

3.6.7 Logging

We have created a node, data_logger, which wraps the rosbag command line functionality. When a rosservice call to /data_logger/start is made, the process forks and, if successful, the child process calls a system exec to rosbag record with a list of rostopics. The rostopics to be recorded are set in the launch file's parameters and include /rosout and /rosout_agg, which provide access to ROS node console messages. These two topics are subscribed by and published to the rosout logging node, which is launched on startup as a part of roscore.

We primarily use rostopics as a means to access sensor data and rosservices as function calls. As rosbag is limited to recording only rostopics, we created a class, VGoLogger, to act as a translator using a custom vgo_msgs/VGoLog message. The message inserts the time the message was sent (time) and identifies the executing node (node). The function that sends the message (func) and the message itself (message) are manually specified. The /vgo/log rosstopic is set to the INFO verbosity level (i.e., log_level=2). For example, the lirc_vgo_remote node utilizes the VGoLogger class. Within each service's function, a temporary instantiation of VGoLogger is set the name of the rosservice corresponding to emulated remote control button press, and the message "sending command irsend." When a lirc_vgo_remote service is called, a vgo_msgs/VGoLog message is published to the /vgo/log rosstopic with the sent time, lirc_vgo_remote as the node, the name of the called service's function, and the "sending command irsend" message.

3.6.8 Simulation

In order to facilitate the development process, it is often useful to have a simulation environment. We used Stage, from the Player/Stage project [Gerkey et al., 2003], to replace the real robot sensors and actuators with an abstraction. This allowed us to rapidly test higher level behaviors without a need to constantly tend for the robot. The Stage simulation is faster to setup compared to simulation using Gazebo [Koenig and Howard, 2004], but it provides a simplistic world without accurate physics.

The simulation contained an approximation of Margo. The simulated robot had a front facing laser range finder and non-holonomic base. When the system was run in simulation, as opposed to running with a physical robot, only a few components were modified. The /vgo_base/vgo_serial_comm was replaced with a dummy version that did not need VGo hardware present. VGo hardware was emulated with the Stage simulator, which used /vgo_base/cmd_vel to command the movement, and provided simulated odometry and laser reading (/scan topic).

3.7 Summary

Since the premise of a telepresence robot is that it is an embodiment for its user, the robot must be able to function sufficiently well as a human proxy. To support remote social interaction, a telepresence robot must at a minimum (1) be able to move, (2) have a microphone to allow the user to hear sounds from the remote site, (3) have speakers to allow the user to be heard at the remote site, (4) have a camera to allow the user to see the environment around the robot, and (5) have a video screen to present the user to people in the remote environment [Tsui and Yanco, 2013]. Additionally, a wide field of view is needed to operate telepresence robot, both horizontally and vertically [Desai et al., 2011; Tsui et al., 2011a; Tsui and Yanco, 2013]. Second, some level of autonomous navigation is required. Direct teleoperation is impractical due to inherent network latency and the movement of people in the remote environment. Finally, several requirements developed for our VGo augmentation (Table 3.2) can be generalized to all future social telepresence robots, namely:

- Robot must map an unknown environment (R3).
- Robot must localize within the map (R4).
- Robot must have advanced sensing for interacting with people in the remote environment (R5).
- Robot must maintain friendly appearance (R8).
- Robot must maintain its stability (R9).
- Robot must have a dedicated camera for driving (R10).
- Robot must have forward and downward facing camera views (R11).

These are the essential components of a research platform needed to inform the design of future telepresence robots for our target population (RQ2).

Chapter 4

Informing Interface and Navigation Behavior Design

Designing human-robot interaction (HRI) systems for people with cognitive and/or physical impairments is difficult. There are several approaches, originating from the research areas of human-computer interaction, human factors engineering, and assistive and rehabilitation technology. Philosophies for designing computer systems for use by a maximal portion of the population largely fall into one of two categories: *Universal Design* (UD), and *Universal Accessibility* (UA) [Stephanidis et al., 1998]. We note that UD and UA as defined by Stephanidis et al. [1998] are logically mutually exclusive. Universal Design (also known as "design for all," "barrier free design," and "inclusive design") refers to designing a system such that anyone can use it without the need for additional individual access methods (e.g., [Abascal, 2002; Abascal and Azevedo, 2007; Abascal and Nicolle, 2005; Steinfeld and Smith, 2012]). Universal Accessibility (also known as "ability based design" and "user sensitive inclusive design") refers to designing a system such that the majority of people can use it without a separate accessibility aid, while providing compatibility with alternative means of accessing the system for individuals who may require them [Newell and Gregor, 2000; Stephanidis et al., 1998; Wobbrock et al., 2011]. Stephanidis et al. [1998] note that UA is typically considered for special populations including seniors and people with disabilities (e.g., [Obrenovic et al., 2007]).

Participatory Design (PD) is a user-centered design process in which stakeholders (i.e., the target end-users) are involved throughout the entire design process [Muller, 2003; Muller and Kuhn, 1993]. Issues regarding accessibility of an interface resulting from an HCI design process can be considered either after the primary design process is complete (i.e., summative evaluation), or ideally they can be considered proactively (i.e., formative evaluation) [Stephanidis, 2001]. Participatory Action Design (PAD) is an application of PD with a goal of providing assistive benefits to end-users to improve their quality of life [Ding et al., 2007a]. In addition to the end-users themselves, other stakeholder groups such as domain experts and caregivers are often included in PAD processes [Allen et al., 2008; Choi, 2011]. PAD has been successfully used for designing assistive HRI systems, including the Quality of Life Technology Center's PerMMA (Personal Mobility and Manipulation Appliance) robot wheelchair with two manipulator arms [Allen et al., 2008; Cooper et al., 2012].

Participation of a domain expert in the PD process is a common practice when designing and developing assistive technologies (see [Allen et al., 2008; Choi, 2011; Sulaiman et al., 2010]). Allen et al. [2008] describe a continuum of expert participation in the design process. On one end of the continuum, the expert acts as a researcher and works most closely with the HCI researchers and can supplement the relevant domain knowledge to the HCI researchers that they may not themselves have. On the other end, the expert acts as a target user, using their domain knowledge to provide the perspective of a target user based on experience gained during previous work with the target population. At the midpoint of the continuum is the liaison role, in which an expert works equally with both HCI researchers and target users; an expert in the liaison role can employ his or her domain knowledge to facilitate communication with target users that might not otherwise be able to communicate without possible frustration of the target users or HCI researchers.

Our approach in designing HRI systems is 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 approaches described in Cooper [2008] and Schulz et al. [2012].

In this chapter,[1] we describe our formative evaluation strategies in which we investigate how people with cognitive and motor impairments would want to direct a telepresence robot in a remote environment. We first conducted a focus group (n=5) with the target audience and experts in the design process who facilitated the discussion of a number of scenarios. Then we conducted a follow-on experiment to investigate the differences and similarities between participants from the target audience giving spoken spatial commands to a person versus a remote robot, which was perceived to be autonomous through "Wizard of Oz" control [Kelley, 1984]. It is important to understand how people conceptualize a remote environment and what they expect a telepresence robot to be able to do in terms of navigation in the given space. We anticipate that people's expectations will change depending on the size and arrangement of an environment. We believe that larger navigation movements, such as moving from room to room, should be autonomous, while FBLR directives are more appropriate for precision control.

4.1 Related Work in Spatial Navigation Commands

A number of corpora have been constructed investigating spatial commands. One approach has been to investigate spoken interactions between human dyads in a spatial navigation task. The HCRC Map Task (n=64) corpus was gathered from unrestricted, verbal dialogues between pairs of people [Anderson et al., 1991]. The Instruction Giver was provided a map with labelled landmarks, a start location, an end location, and a goal path; the Instruction Follower had a corresponding map without the path. The CReST (n=14) corpus' task utilized a cooperative, remote, search task of six offices [Eberhard et al., 2010]; the Instruction Follower wore a helmet with a video camera that the Instruction Giver was able to see in real time, in addition to a map of the six offices. The SCARE (n=30) and the GIVE (n=72) corpora are similar, but were conducted in virtual reality [Gargett et al., 2010; Stoia et al., 2008].

Another approach is to investigate how people give spatial commands to robots. In the TeamTalk corpus, Instruction Givers (n=35) directed one robot "Mok" to a

[1]Portions of this chapter were published in [Tsui et al., 2013b].

goal position near another robot "Aki" in an empty environment [Marge and Rudnicky, 2011]. The TeamTalk corpus featured only monologue-style instructions as Instruction Givers were asked to formulate their instructions before audio recording them. The IBL corpus featured both monologue and dialogue-style instructions in which Instruction Givers directed a small robot to points in a miniature city [Bugmann et al., 2004, 2001]. In the asynchronous monologue condition, Instruction Givers (n=16) were told that their instructions would later be used by Instruction Followers to remotely drive the robot using only its camera view; participants in this condition were allowed to use unconstrained speech and were asked to reuse routes they had previously defined. In the synchronous dialogue condition, Instruction Givers (n=8) were told that the Instruction Followers were taking notes in the adjacent room. The Instruction Follower was an experimenter (i.e., a "wizard" [Kelley, 1984]); he interrupted if the instruction was deemed "too long" and provided responses as though he understood all of the directions. During both experiments, the robots did not move.

Riek [2012] notes that "Wizard of Oz" studies [Kelley, 1984], in which a person "acts" as the robot, are commonly used by the human-robot interaction (HRI) community in experiments involving natural language comprehension. Typically, the "wizard" is a person familiar with the robot's abilities and limitations. However, studies sometimes utilize naive robot operators to provide feedback and perform as the robot. For example, Koulouri and Lauria [2009] collected a data set based on a modified version of the Map Task. The Instruction Follower was changed to a robot partner, whose actions and responses were controlled by the second participant; the type of responses given by the robot operator was varied. Additionally, the Instruction Follower's map showed only the robot and the area surrounding its current position, and the Instruction Giver saw the full scale static map without the robot's position and orientation. Chernova et al. [2010] gathered data from 558 two-player online games in which one person took the role of the human astronaut and the other a robot on Mars; the interaction between the players was also typed text chat.

None of the aforementioned corpora and data sets involved people with disabilities as participants, and it is unknown how our target audience would want to direct robots in a remote environment. Also, there is an underlying assumption in the human-human corpora that people will talk to robots in the same manner that people talk with other people (e.g., Map Task [Anderson et al., 1991], CReST [Eberhard et al., 2010], SCARE [Stoia et al., 2008], GIVE [Gargett et al., 2010]). Finally, we note that responses and feedback from the robot are largely overlooked with the exception of Koulouri and Lauria [2009]. The role of the Instruction Follower was usually occupied by a human partner (e.g., [Anderson et al., 1991; Eberhard et al., 2010; Gargett et al., 2010; Stoia et al., 2008]) or a human acting in the role of the robot in an unconstrained manner (e.g., [Bugmann et al., 2004, 2001; Chernova et al., 2010]).

4.2 Study 1: Telepresence End-user Focus Group

In the first study, our goal was to understand how people with cognitive and motor impairments want to give instructions to a telepresence robot in a remote environment both explicitly and implicitly. Through a series of scenarios, we investigated how a robot should provide feedback to the operator and how it should ask for help.

4.2.1 Participants

We recruited five 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 the focus group. Three participants used tracheostomy tubes for assisted respiration (P1, P2, and P4). Three participants used a power wheelchair: one independently operated his chair with a mouth stick (P1), and two participants required the assistance of a caregiver (P2 and P4). Two participants used a manual wheelchair: one required the assistance of a caregiver (P3), and one was independently able to propel his wheelchair (P5).

4.2.2 Focus Group Design

We first described the robot system as a physical entity in a remote space which provides an embodiment for a person's virtual self; the robot and the robot operator are separated. We further explained that telepresence robots provide interactive two-way audio and video communication. Additionally, we told the participants that a telepresence robot's movements can be independently controlled by an operator, which means that the person driving can explore and look around as he or she desires. We demonstrated the two-way audio and video communication of a VGo Communications' VGo robot [2011] (see Figure 4-1, p. 77) and drove it around the conference room.

We described the robot in example scenarios involving an environment well known to the participants, then discussed how the interaction should change if the robot begins traveling in a new place. We developed a series of scenarios to investigate these interactions (Table 4.1 and Appendix B); a subset of these scenarios was selected for discussion in the telepresence end-user focus group. Toward the end of the focus group session, we inquired if there were things the robot could do that would increase or decrease the participants' trust in the system. We asked if they had any prior experience with speech recognition technology, if they had a preference for the robot's voice sounding male or female, and if they had any pets.

The focus group was 120 minutes in length. Participants were offered $75 each as compensation for their time and travel. The focus group was audio and video recorded, and the recordings were transcribed by CastingWords [2012].

4.2.3 Insights

We noted several themes related to navigation: accommodation of individuals' access/input methods, command content, command structure, and the user interface elements necessary to support navigation functionality.

Table 4.1: Study 1: Sample scenarios and open-response questions developed to help focus group participants imagine themselves using a telepresence robot.

Scenario 1	Things I want to see today. Unordered list: Community garden space. Group cooking class at 1:00pm. Visit with your friend. Gymnasium (open from 3:00pm to 6:00pm for general use). Watch a movie. Cafeteria (food options, coffee). • How would you tell or show your robot what you wanted to do? • How should the robot tell or show you that it understands? • When would you tell or show the robot each event? • Where would you expect the robot to go first? Next? (and so on)
Scenario 2	Imagine your friend gives you a message that he or she has a visitor coming at 1:30 pm and would like you to join them, using the robot, for a walk on a wheelchair accessible nature trail. You will meet your friend and his or her visitor at the reception desk. • How would you tell or show your robot about this new event? • How should the robot tell or show you that it understands this? • Should the robot tell or show you about the rest of your day? If so, how? • Should the robot remind you that the gymnasium opens at 3:00pm? If so, when would you expect a reminder?

Input Devices and Access Methods

The focus group participants emphasized that control of a telepresence robot must be customizable and accommodate both an operator's primary and secondary communication access method or input device. Participants mentioned the following access methods, as can be seen by example quotes:

- Eye gaze: "If you were to stare forward and then look left, it would turn left."
- Joystick: "If you have a disability where you can't talk, we also want the option of being able to run it with a joystick."
- Mouth stick: "He turns that chair extremely accurately with his mouth."
- Computer mouse: "If you can use your mouse, you can drive a machine."
- Voice: "You've got to have voice command."

Focus group members were concerned about having the flexibility of alternative access methods because their ability to use a particular method may be situationally dependent. For example, if they are in a noisy environment or "if you had a cold," it may be difficult to control a robotic wheelchair via voice command. In fact, concerns regarding voice commands and the logistics of noise cancellation for background sounds were raised throughout the duration of the focus group.

Voice Command Structure

The structure of voice commands can be characterized by their level of simplicity and form of speech. Commands can be structured to be simple or complex, where a simple command is defined as containing a single instruction and a complex command is formed by conjunctions of simple instructions (that is, it becomes a compound instruction). Participants noted that the robot should be controlled with simple and distinct commands. P4 stated, "We're going to have commands in relationship to where it's going to go, how it's going to go, and which way it's going to go... If you have too many words, that can drag. And you'd say ... probably five or six words.

[Commands] can't be mistaken for something else that you're saying." P3 reported that he would give the robot instructions "one at a time" as opposed to stacking or queuing commands.

Toward the end of the focus group, P1 gave example voice commands including "park it," "follow the nurse," "go to my room," "bring me there," and "show the nurse to my room." This last command is an example of a complex instruction, because it can be decomposed into sequential simple commands "find the nurse" and "bring the nurse to my room." First, "find the nurse" implies potential navigation to multiple locations, not just the nurses' station. Second, "bring the nurse to my room" implies success of the first command, understanding where the robot is currently, planning a path between the robot's current location and a destination point, and navigating the planned path while ensuring the nurse accompanies the robot.

Example commands used by focus group participants took the form of three different types of speech: imperative, declarative, and interrogative. The majority of the voice command examples were simple, imperative statements with an implicit subject of "you, the robot," one verb with an optional adverb, and an optional predicate noun or prepositional phrase.

Voice Command Content

Commands can encompass high-level concepts, such as "go to <*destination*>", or low-level speech, such as "move forward" (or back, left or right). In general, the high-level concepts refer to global environmental knowledge – information that presupposes an understanding of the entire space in which the robot navigates – while the low-level speech constructs refer to local environmental knowledge – information that pertains only to the area immediately surrounding the robot. Commands based on either of these levels can be simple or complex, depending upon whether they are expressed in the form of single or compound instructions.

Early parts of the focus group session centered on simple, low-level commands. For example, P1 reported he would tell the robot to "move forward, move forward, and then... it stops." However, during the course of the focus group, P4 proposed that each building should have a sensor for known destinations. The participants then envisioned themselves speaking to the robot at a higher level (e.g., go to location, follow someone). P4 provided some voice command examples that leveraged known locations (e.g., “go to the gym,” “go to the cafeteria,” “go to the principal’s office”) and also local destinations within a given space (e.g., “go to the cashiers,” “go to the counter”); P2 stated her support for P4’s discussion of high-level driving commands, and P1 also gave examples of “go to <*destination*>” driving commands during the focus group.

We posed a scenario to P2 in which the goal location did not yet exist in the robot’s map. P2 stated she would “find out where we’re going [by] ask[ing] a nurse;” then she would “tell the robot.” P3 reported that he would “just talk [to] it” when instructing the robot to “go find her [the nurse].” We asked what P2 would do if the nurse did not know, and she noted that she would ask “until I found someone who did know.” P1 also reported he would ask for directions or for the person to “bring me there” or the robot to “follow the nurse.”

User Interface Elements

The concept of maps is important to navigation and recurred during the focus group discussions. P1 requested a map; “If it’s [the robot] got a map of the building, first, I would say ‘go to’ whatever location, say gymnasium, and it would go there.” It is unclear, however, if P1’s request was for the robot to have a map representation or if a map should be provided for the robot operator to reference in a user interface.

We inquired about the participants’ prior use of voice recognition software and posed the concept of the telepresence robot having this functionality. P4 stated, “My thought here is for people with handicaps and people on the other end talking to somebody on the screen... was a bouncing ball at the bottom or something, [with] the words of what you’re saying" in addition to the video conferencing component. P4 continued, “They can see you, but they can also read on the screen what you’re saying.” We asked how the robot should provide feedback to the operator. We posed the scenario that “you tell the robot something, you need it to do something, you need it to go somewhere. How do you know the robot knows exactly what you told

[it]?" P4 again noted that the verbal commands should be displayed on the operator's screen below the video, and the operator can repeat or give new commands to correct it. Also, they noted that verbal commands should be displayed on the operator's screen, which has the implication of speech recognition. It was unclear, however, if the robot should infer commands when "listening in" on operators' conversations.

4.2.4 Discussion

The insights from this focus group provided examples of how robot operators might want to interact with a telepresence robot. We observed that the participants envisioned instructing the robot in a seemingly different manner than if they were giving instructions to an adult, child, or pet. The example commands were largely simple imperative statements (e.g., "move forward/back/left/right," "go to <*destination*>").

It should be noted that only two of the focus group participants regularly used video conferencing, thus the concept of a telepresence robot was quite abstract to many of the participants. It was unclear if the participants recognized that the telepresence robot was a representation of themselves or if they thought of the robot as a personal service assistant. Therefore, we designed a follow-on experiment in which people with cognitive and motor impairments were placed in the situation of instructing an agent in a scavenger hunt task.

4.3 Study 2: Scavenger Hunt Experiment

For this second study, we designed an experiment in which people gave verbal instructions to guide a remote shopping assistant within a space similar to a retail store. Our goal was to see how the language used changed if the participants thought that 1) they were commanding an autonomous robot named "Margo" in the environment, or 2) a person named "Kelsey" was moving a camera in the environment for them.

Table 4.2: Study 2: Participant Descriptions

P#	Age	Gender	Medical Condition[2]	Cognitive Impairment	Literacy Ability	Speech Ability	Vision Ability
P1	36	M	TBI	moderate	intact	intact	functional when corrected with glasses
P2	63	M	ALS	none	intact	intact; uses trachostomy tube	functional when corrected with glasses
P3	67	M	TBI	moderate	intact	intact	functional
P4	45	M	TBI	moderate	intact	intact	functional when corrected with glasses
P5	32	M	MD	none	intact	intact; limited breath support	functional
P6	24	F	TBI	moderate	intact	intact	functional when corrected with glasses
P7	20	F	SB	mild	moderate	intact	functional
P8	22	M	SCI	none	intact	intact; used trachostomy tube	functional
P9	32	M	TBI	mild	intact	intact	visual field loss on right side
P10	52	M	TBI	moderate	intact	intact	functional vision when corrected with glasses; significant hemispheric neglect
P11	64	M	CVA	moderate	intact	intact; clarity slightly affected	functional when corrected with glasses
P12	27	M	TBI	moderate	intact	intact	functional

[2] Amyotrophic lateral sclerosis (ALS), cerebrovascular accident (CVA, or stroke), muscular dystrophy (MD), spina bifida (SB), spinal cord injury (SCI), traumatic brain injury (TBI)

4.3.1 Recruitment and Participants

We recruited 12 participants for our between-subjects study. Participants were members of the Crotched Mountain Rehabilitation Center community, including inpatient clients from the Brain Injury Center and participants in the residential program. Each spoke English and had a condition that significantly limited their ability to travel and maintain contact with important individuals. Their medical conditions included amyotrophic lateral sclerosis (ALS), cerebrovascular accident (CVA, or stroke), muscular dystrophy (MD), spina bifida (SB), spinal cord injury (SCI), traumatic brain injury (TBI). We excluded people with blindness, low arousal levels, or other conditions preventing benefits from use of the robot. People with severe cognitive challenges who were unlikely to understand that the robot was a representation of themselves, as opposed to a TV show or video game, were also excluded.

Ten men and two women participated in the experiment; the average age was 40.3 years (SD=17.3). A summarized description of their abilities is given in Table 4.2. All participants had functional vision ability; one participant had visual field loss on his right side (P9) and another had significant hemispheric neglect (P10). Eleven participants had intact literacy ability; P7 had moderate literacy. All participants had intact speech ability. P2 and P8 used tracheostomy tubes for assisted respiration. P5 drew oxygen from a tube and had limited breath support; consequently, he had a quiet voice. P11 had a cerebral stroke, which slightly affected the clarity of his speech.

P2, P5, and P8 had intact cognition. P7 and P9 had mild cognitive challenges (able to function in most environments independently and execute activities of daily living with assistance from memory aids). The remaining seven (P1, P3, P4, P6, P10, P11, and P12) had moderate cognitive challenges (may have significant memory loss but able to perform most activities of daily living with minimal support, except cooking and bathing for safety reasons).

Four participants had prior experience with voice recognition. P2 and P8[3] had used Dragon NaturallySpeaking, and P12 used Dragon Dictate. P5 noted that he

[3] It should be noted that P2 and P8 also participated in the Study 1 focus group as P4 and P1, respectively.

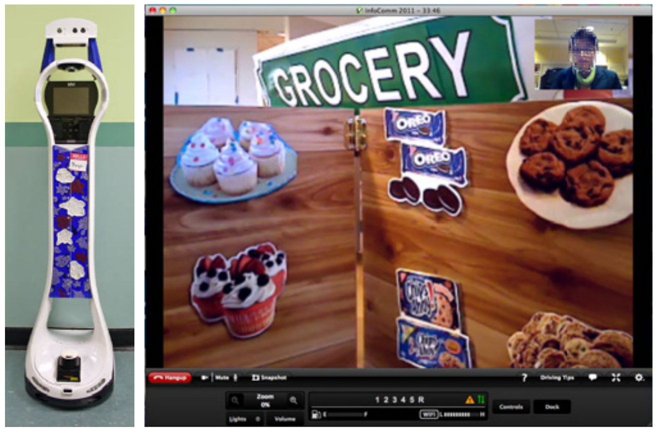


Figure 4-1: **(Left)** Margo, a modified VGo telepresence robot. **(Right)** Participant's view of the remote shopper's webcam using the desktop VGo App.

experience with a voice activated computer but did not list specific software.

4.3.2 Experimental Design

Participants provided their informed consent or assent, as appropriate. Then, the experimenter read a script describing the remote shopping experience. The premise was that participants were to host a themed party. They would shop for costumes, food, drinks, party games, and movies to show at their parties (Task 1, 15 minutes) with the help of a remote shopping assistant located at "KAS Party Central." The remote shopping assistant would show the store to the participants using a webcam. At the end of the scenario, they would finalize their choices at a checkout station with a party planner (Task 2, 5 minutes).

The live video from the webcam on our VGo robot, Margo, was maximized on a 22 inch (55.9 cm) Dell monitor with 1920×1280 pixels (Figure 4-1). The webcam was described as having 2-way audio, so that they could hear what was going on at the store, and the remote shopper could hear anything they said to it. Participants were told to talk to tell the shopper where to go and what they were looking for. Participants were also told that there was a few second delay between when they spoke and when the remote shopper heard them.

The experimenter provided the participants with a store directory (indicated by blue hash marks in Figure 4-2 (left)) and a written description of the two tasks, including the

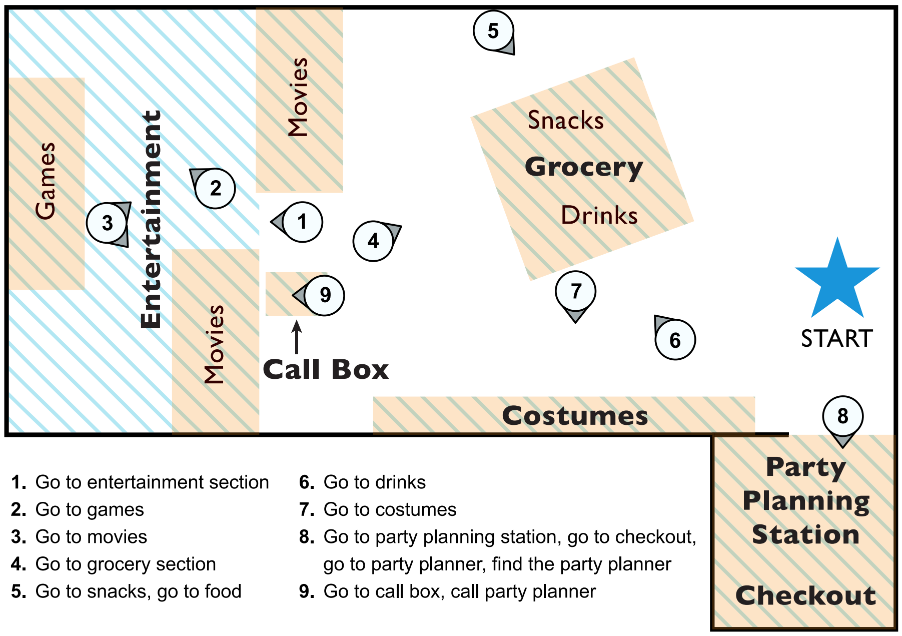


Figure 4-2: "KAS Party Central" store layout. The numbered circles with triangles indicate location and orientation for "go to <*named destination*>" instructions for the wizard. Participants were given a version of the store directory shown as blue hashes with the Entertainment, Grocery, Costume, Party Planning Station, and Checkout sections shown with bold text. For detailed item placement see Figure C-1 in the appendix.

shopping list (i.e., 2 types of snacks, 1 drink, 1 movie as recommended by a party planner, 1 party game, 1 costume). After completing both tasks, participants were asked to draw their path on a map and engaged in a post-experiment interview. At the end of the session, the telepresence robot and wizard were introduced to the participant, and the researchers answered any remaining questions about the study. Estimated time for this study was 60 minutes, and participants were compensated with a $20 gift certificate.

It should be noted that a training period was not provided as the purpose of this study was to understand how people provide verbal navigation instruction. Our experiment featured a single wizard whose role was to interpret the participants' verbalized navigation instructions. The wizard then controlled the telepresence robot in accordance with the given instruction. For both the robot-agent and human-agent conditions, the instructions were interpreted in the same manner, and the wizard moved the VGo telepresence robot using a USB gamepad.

4.3.3 Wizarding

Based on the Study 1 focus group, we anticipated that the participants would use a wide range of language including low-level FBLR directives (e.g., turn right, go right), relative descriptions using local area information (e.g., take the next left), and global destinations (e.g., go to a named location). In addition to the robot moving as directed by the participant, our wizard provided a limited amount of scripted verbal feedback. Koulouri and Lauria [2009] found that when a small set of limited feedback (Condition 3) is provided to a robot operator, the operators reverted to giving low-level FBLR commands, ignored the feedback, and focused on the robot's movement. In Condition 3, their wizard gave seven fixed responses: "hello," "goodbye," "yes," "no," "sorry, I don't understand," "sorry, I can't do that," and "what?" In Condition 2, their wizard was additionally allowed to "request clarification" from and "provide information" to the robot operator in an open text format. Our verbal protocol expanded on this level of feedback, which we detail as we define how the wizard controlled the robot.

In general, the wizard rephrased the participant's command with acknowledgement. Figure 4-2 shows the robot's target location and orientation for "go to <*named destination*>" instructions. For example, our wizard responded "going to <*named destination*>" when initiating movement, and "Ok, I'm here" when the robot was positioned and oriented. The choices for each of the shopping list items were not known to the participants beforehand; the shopping assistant started each run as if the list were unknown to the robot or person as well. Thus, "go to <*named object*>" and "find the <*named object*>" instructions were not valid (e.g., "go to the robot movies," "find the cupcakes"), as prior knowledge about the environment was not a dependent variable in this experiment.The wizard responded "I don't understand," "I'm sorry," and/or "Please repeat" to invalid or ambiguous instructions.

FBLR command actions were set according to what people would expect. For example, "turn right" resulted in the robot turning at most 90 degrees, as opposed to turning to the right continuously until commanded to stop. If the wizard was instructed to drive for a potentially long distance (e.g., "drive straight," "drive forward," "drive down

Table 4.3: Study 2: Remote shopper's verbal protocol

Response to participants' commands	***n***	$\bar{x}$ **seconds**	***SD***
[Hello	Hi]	4	3.0	1.4
Going to ...	68	7.1	4.7
I am here.	42	–	–
Picture taken.	81	7.2	4.5
How may I help you?	5	10.4	12.3
What should I do?	4	29.3	17.6
I don't understand.	3	12.3	6.0
[Sorry	Please repeat]	7	7.3	2.7
[Yes	Ok]	12	4.6	2.8

the hall"), the wizard followed the instruction until a wall, shopping display, or obstacle was encountered. The wizard did not provide verbal feedback for FBLR directives.

Our verbal protocol included a number of additional feedback messages. When a participant indicated an item selection, the wizard responded by saying "picture taken." The wizard prompted the participant at the beginning of Task 1 ("how may I help you?") if he or she did not initiate instruction, and also during the task if the participant was silent for 30 seconds after the last robot movement or command ("what should I do?"). If a participant chained multiple commands, the wizard acknowledged the sequence, rephrased and acknowledged the first command (e.g., "going to ..."), acknowledged the completion of the first command, rephrased and acknowledged the second command (e.g., "now going to ..."), and so on. The wizard incorporated awkward silences between words in the robot agent condition; otherwise, the wizard used her regular speaking voice. Table 4.3 summarizes the remote shopper's verbal protocol and the number of responses spoken to the participants; the latency of the remote shopper's "I am here" responses was dependent upon the distance between the current location and the requested destination.

4.3.4 KAS Party Central

We divided a space (approximately 20×40 ft (6.1×12.2 m)) into four sections. There were two large sections which each had two subsections; the snacks and drinks were located in the "grocery" section, and the movie posters and party games in the "entertainment" section (Figure 4-2). The party planning station was co-located with the checkout in an adjacent room. The costumes were grouped together in one section. Large, green signs denoted the costume, grocery, entertainment, and party planning sections, and smaller, blue signs denoted the subsections within grocery and entertainment (Figure 4-3). A call box was set in the middle of the store, out of the line-of-site of the party planning station.

We populated the store with party refreshments and theme items. There were multiple images associated with each choice, as the participants may not have been familiar with any one in particular. The images show both brand name and generic items.[4] It was not important for the participant to select one specific image within a choice. Figure 4-3 (top) shows examples of the two types of the four drink choices (i.e., milk (option 1), fruit punch (2), orange soda (3), and water (4)) on one side of the grocery section. The snack choices were pretzels (option 1), apples (2), cupcakes (3), and cookies (4). We created four party themes (i.e., circus (theme 1), Christmas (2), Halloween (3), robot (4)), and each theme was assigned three times. There was one choice for each theme's costume and party game, and three movie choices per theme. Costumes were comprised of a t-shirt and a mask, wig, or hat (Figure 4-3 top, shown center). Movie choices were shown as large poster and grouped according to theme (Figure 4-3 bottom, shown center and right).

The placement of specific themed items, snack options, and drink options is shown in Figure C-1 in the Appendix. If the robot was commanded to go to the costume section or games subsection, all themed items were visible in the robot's camera view. For all other sections and subsections, we positioned the robot such that it was unable to view all of the items at once. The four movie themes were split into two separate subsections with the circus and Christmas themes grouped together, and

[4] All trademarks are the property of their respective companies.

Figure 4-3: "KAS Party Central." **(Top)** View of the party planning station and checkout, costumes, and drinks. **(Bottom)** View of the games table and movies.

the Halloween and robots themes together (Figure 4-3 bottom). The remote shopper would initially face the posters that did not contain participants' themed movies; note the two orientation indicators for location 3 in Figure 4-2. If the remote shopper was asked to go to the entertainment section, the robot's camera view would show the large green section sign with a portion of the games subsection visible on the right side (location 1 in Figure 4-2).

The grocery section walls formed an "X." If the remote shopper were asked to go to the drinks subsection (location 6 in Figure 4-2), two of the four choices would be shown: water (option 4) on the left, and milk (option 1) on the right. The walls showing options 2 and 3 (fruit punch and orange soda, respectively) were perpendicular. Similarly, two of the four choices would be shown if the remote shopper was asked to go to the snacks subsection (location 5). Finally, if the remote shopper were asked to move to the grocery section, only one snack choice and drink choice would be visible; cupcakes appeared on the left and and orange soda on the right, respectively.

4.3.5 Data Collection and Analysis

We recorded video and audio for each session, and the recordings were transcribed using CastingWords [2012]. We developed and refined a categorical coding scheme through open and axial coding [Glaser and Strauss, 1967] based on the participants' utterances (Tables 4.4 and 4.5). Utterances are separated by a verbal response from the wizard, the start of the command action by the wizard, or elapsed silence by the participant of at least 10 seconds. Cohen's kappa for inter-rater reliability was computed as κ=0.86 (excluding chance).

Eleven of the twelve participants completed the primary shopping task within the 15 minutes allotted; we have removed P10 from the statistical analysis due to non-completion of the task. We coded 312 total utterances from these 11 participants. As utterances could contain more than one sentence or phrase, we considered the whole utterance and noted all appropriate categories. Participants in the human agent condition (n_H=6) spoke a total of 178 utterances ($\bar{x}_H$=29.7, SD=12.9), and participants in the robot agent condition (n_R=5) spoke a total of 134 utterances ($\bar{x}_R$=26.8, SD=15.1). Unless otherwise noted, we computed two-tailed Student's unpaired t-tests with a confidence interval of 95% (α=0.05) on the categorial frequency count between the human agent and robot agent conditions.

4.3.6 Results and Discussion

Overall, speech used to direct the human remote shopper and the robot remote shopper had few statistically significant differences (see Figure 4-4 and Table 4.6). We believe that this result is due to the experimental design. In both conditions, participants were given a very simple and limited description of the remote shopper's capabilities. We did not provide details in the scenario description as to how the remote shopper would move in the environment, if the remote shopper implicitly knew to avoid obstacles, and if the remote shopper knew what items were located in the store and where they were located. Additionally, we did not ask the participants any pre-experiment questions to prevent biasing their language style. We solicited the participants' experience with voice recognition at the end.

Figure 4-4 (left) shows the averages and standard deviations of the utterances that required no, local, and global environmental knowledge for the remot e shopper to complete the requested command; no statistical differences were found between the

Table 4.4: Category coding definitions (Part 1 of 2); κ=0.86 (excluding chance)

Category	Description	Examples
ENVIRONMENTAL KNOWLEDGE		
None	The requested command can be done regardless of the robot's location (disregarding obstacle avoidance)	Stop, forward, back, left, right, tilt, zoom
Local	The requested command requires information from local sensors (i.e., camera view)	Take a picture
Global	The requested command requires knowledge beyond local sensors	Go to <*named destination*> (not in camera's current field of view)
SENTENCE COMPLEXITY		
Simple	One independent clause:[5] simple or compound subject and simple or compound verb	Go to the snacks.
Compound	Two independent clauses joined using: for, and, nor, but, or, yet, so	Go to the snacks and turn right.
Complex	One independent clause joined with one or more dependent clause(s) using a subordinating conjunction (after, although, as, because, before, if since, though, unless, until, when, whenever, where, whereas, wherever, while) or relative pronoun (that, what, who, which)	Go to the snacks, which are on the other side of groceries from drinks.
SENTENCE TYPE		
Declarative	Sentence that makes a statement	Ex. 1: I need a snack. Ex. 2: The first item on the list is a snack. Ex. 3: I choose pretzels.
Imperative	Expresses a command, request, or selection. Subject may be implicit ("you") or explicit (proper name of remote shopper). Verb may be implied when the predicate is only an adverb phrase or direct object.	Ex. 1: [You,] go to the snacks. Ex. 2: [You, turn] left. Ex. 3: [You, go to] snacks, please.
Interrogative	Sentence that asks a question	Could you go to the snacks?
Interjection	A single word or non-sentence phrase which is not grammatically related to the rest of the sentence	Ok, all right, yes, well, hi, thanks

[5] A clause has a subject and a predicate; a predicate minimally contains a verb. An independent clause is a sentence.

Table 4.5: Category coding definitions Ppart 2 of 2); κ=0.86 (excluding chance)

Category	Description	Examples
FEEDBACK TO REMOTE SHOPPER		
Praise	Utterance includes positive feedback given to remote shopper when beginning, while executing, or completing an instruction.	Good, excellent
Confirmation/ acknowledge-ment	Utterance includes neutral feedback given to remote shopper when beginning, while executing, or completing an instruction.	Ok, yes
SOCIAL ETIQUETTE		
Greeting	Utterance includes acknowledgment of remote shopper.	Hello, hi
Expressing polite request	Utterance includes "please."	Please
Expressing polite gratitude	Utterance includes "thank you."	Thank you, thanks
Addressing by name	Utterance includes the remote shopper's name.	Margo (robot), Kelsey (human)
OTHER		
Not to remote shopper	The utterance is not directed at or spoken to the remote shopper.	
No code	There is no appropriate category for this utterance.	

Table 4.6: Resulting frequency counts (n) and averages ($\bar{x}$) from categorical coding of participants' transcripts, excluding P10. Statistically significant p-values are bolded.

Category	Condition	n	$\bar{x}$	SD	p
ENVIRONMENTAL KNOWLEDGE					
None	Human agent	51	8.5	15.5	0.75
	Robot agent	57	11.4	13.4	
Local	Human agent	46	7.7	3.2	0.55
	Robot agent	46	9.2	4.6	
Global	Human agent	45	7.5	3.7	0.85
	Robot agent	39	7.8	0.4	
SENTENCE COMPLEXITY					
Simple	Human agent	134	22.3	11.3	0.67
	Robot agent	132	26.4	17.8	
Compound	Human agent	6	1.0	2.0	0.70
	Robot agent	3	0.6	1.3	
Complex	Human agent	18	3.0	1.9	0.56
	Robot agent	9	1.8	4.0	
SENTENCE TYPE					
Declarative	Human agent	67	11.2	8.2	**0.05***
	Robot agent	13	2.6	1.9	
Imperative	Human agent	79	13.2	18.0	0.71
	Robot agent	89	17.8	21.8	
Interrogative	Human agent	22	3.7	5.2	0.24
	Robot agent	49	9.8	9.3	
Interjection	Human agent	79	13.2	6.9	**0.02***
	Robot agent	18	3.6	3.2	
FEEDBACK TO REMOTE SHOPPER					
Praise	Human agent	5	0.8	2.0	0.83
	Robot agent	3	0.6	1.3	
Confirmation/acknowledgement	Human agent	58	9.7	4.8	**0.01****
	Robot agent	9	1.8	2.9	
SOCIAL ETIQUETTE					
Greeting	Human agent	3	0.5	0.5	0.77
	Robot agent	3	0.6	0.5	
Expressing polite request (please)	Human agent	8	1.3	2.2	0.21
	Robot agent	17	3.6	3.0	
Expressing polite gratitude (thank you)	Human agent	4	0.7	1.2	0.92
	Robot agent	3	0.6	0.9	
Addressing by name	Human agent	6	1.0	0.9	0.30
	Robot agent	50	10.0	16.9	

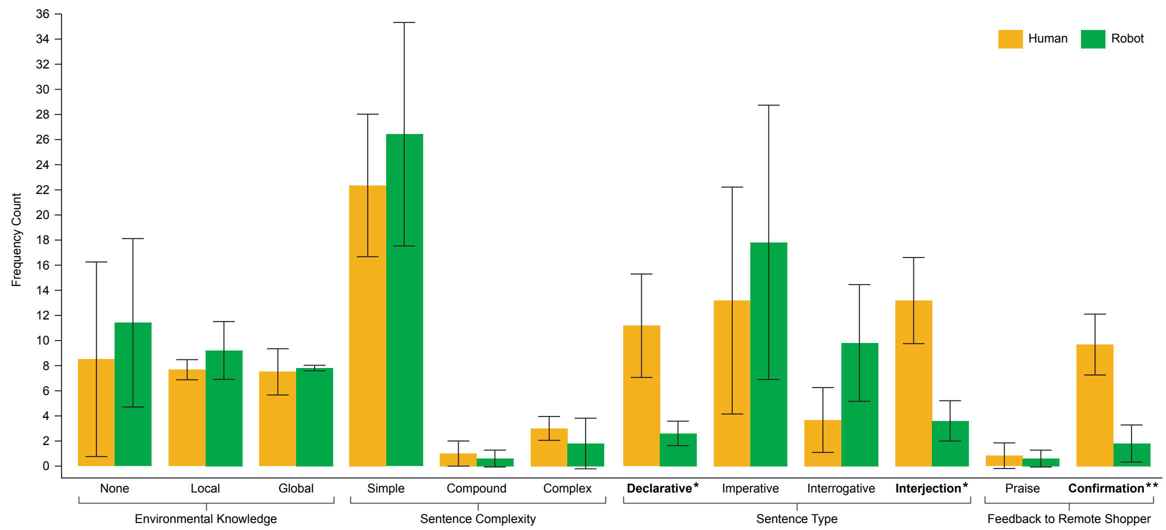


Figure 4-4: Average and standard deviations of coded categories. Significant differences were found between the human and robot conditions for the declarative and interjection categories in sentence type ($p \leq 0.05$), and the confirmation category in feedback to remote shopper ($p < 0.01$); shown as bolded category titles.

human agent and robot agent conditions overall. Figure 4-4 (center) shows the averages and standard deviations for the sentence type and sentence complexity categories. No statistical differences were f ound between the human agent and robot agent conditions for the sentence complexity categories, and the imperative and interrogative s entence types.

The differences in the participants' verbal instructions can largely be attributed to personal style. Some participants in both conditions primarily gave imperative commands that did not require any environmental knowledge to fulfill th e request (Figure 4-5 left: left, right, turn, stop). P2 in the robot agent condition gave three times as many commands that did not require environmental knowledge (n=33) than global on es (n=11), and P5 in the human agent condition gave more than four times as many (n=40 and 9, respectively); see Figures C-2 and C-4, respectively. All participants utilized commands that required global environmental knowledge (Figure 4-5 right: go to). P3 (human agent condition) primarily used declarative language, while P6 and P8 (robot agent condition) used interrogative language (see Figures C-5 and C-7, respecti vely). Two participants, P2 and P5 (one in either condition), praised the remote shopper's completed actions. The majority of the

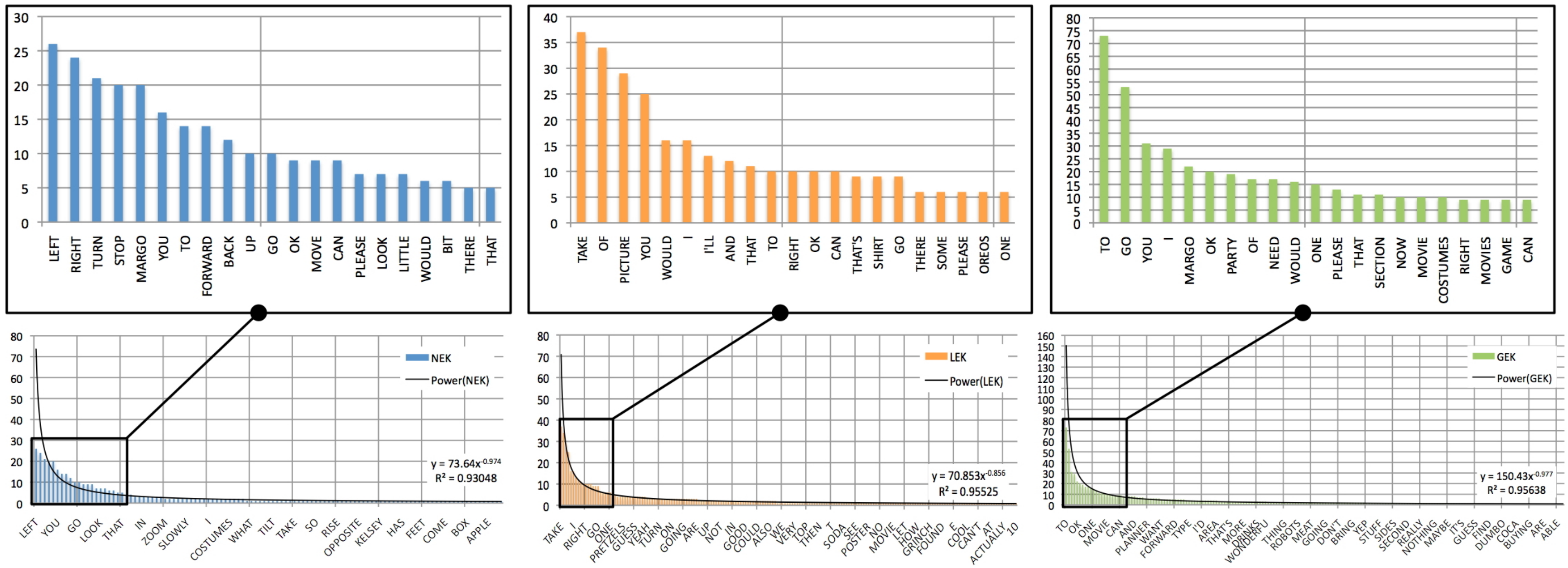


Figure 4-5: Unique word histograms of top 21 utterances coded by levels of environmental knowledge. **(Left)** None: Directives require no environmental knowledge, and the request can be performed regardless of the robot's location; n=384, 67% representation. **(Center)** Local: Directives require information from the robot's sensors at an instance in time; n=573, 51%. **(Right)** Global: Directives require knowledge beyond the robot's local sensors; n=802, 51%.

utterances were simple sentences (n=266; n_H=134, $\bar{x}_H$=22.33, SD=11.33; n_R=132, $\bar{x}_R$=26.40, SD=17.78); two participants, P6 and P9 (one in either condition), provided the majority of the compound and complex language(n_{P6}=9, n_{P9}=11); see Figure C-5.

In the post-experiment interview, participants stated that they used a number of strategies to shop for their party items. P2, P3, P6, and P7 noted going in order of the checklist. P5 stated that he wanted to be efficient by looking around the store. P9 noted that he directed his remote shopper using the store directory map and went to the next closest item. We asked inquired about the participants' strategies for providing verbal instructions to the remote shopper and how it may have changed during the experiment. P2 and P12 gave vague responses; P2 noted that he "spoke concisely and clearly" and P12 said that his strategy was to "be polite, like I would to talk to somebody." P3, P5, and P8 were more specific. P3 stated he "told her [Kelsey] where to go" and P8 also noted that he "just asked her [Margo] to go to certain areas." P5 noted that he "initially talked to her [Kelsey] like a human" but changed his verbal instruction style to give FBLR directions, which he said felt that he had more control. Only P5 stated that his strategy for talking to the remote shopper changed.

Declaring Item Selection

Participants in the human agent condition spoke significantly more declarative utterances (n_H=67, $\bar{x}_H$=11.2, SD=8.2) than participants in the robot agent condition (n_R=13, $\bar{x}_R$=2.6, SD=1.9); p=0.05 with $t(9)$=2.25. We found that 55 of the 67 declarative utterances in the human agent condition (82.1%) and 8 of the 13 in the robot agent condition (61.6%) were first person declarative sentences (e.g., "I want," "I need," "I'll take," "I'll choose"). Participants in the human agent condition spoke significantly more first person declarative utterances ($\bar{x}_H$=9.7, SD=7.2) than the participants in the robot agent condition ($\bar{x}_R$=1.6, SD=1.8); p<0.05 with $t(9)$=2.27. The functions of these first person declarative utterances related to item selection and specifying locations, which require local and global knowledge, respectively.

We further investigated the function of the commands that required local environmental knowledge, which included item selection by taking a picture. Seventy-seven of the 92 utterances that required local environment knowledge (83.7%) involved item selection. Participants in the human agent condition had a similar number requiring local knowledge (n_H=40, $\bar{x}_H$=6.7, SD=2.7) as the participants in the robot agent condition (n_R=37, $\bar{x}_R$=7.4, SD=2.7); p=0.67.

We then looked at the sentence type corresponding to item selection and found that participants spoke a total of 24 declarative utterances (n_H=13, n_R=1), 38 imperative (n_H=17, n_R=21), and 23 interrogative (n_H=6, n_R=17). Again, participants in the human agent condition spoke significantly more declarative utterances ($\bar{x}_H$=3.8, SD=3.1) than those in the robot condition ($\bar{x}_R$=0.2, SD=0.4); p=0.03 with $t(9)$=2.52. Participants in the robot agent condition spoke a greater number of imperative utterances ($\bar{x}_R$=4.2, SD=4.1) and interrogative utterances ($\bar{x}_R$=3.4, SD=3.2) than those in the human condition ($\bar{x}_{H_{imp}}$=2.8, SD=2.7; $\bar{x}_{H_{int}}$=1.0, SD=2.4, respectively), though not significantly so (p=0.54 and p=0.21, respectively).

Finally, we note that for all utterances involving item selections, all but four identified the specific item. Participants also provided descriptive information including the color of the item (e.g., P7: "I'll do the black shirt with the skeleton"), its location in the camera's field of view (e.g., P6: "And also the milk that's on the right side of the bottom?"), and its location with respect to other items (e.g., P9: "Up above the black shirt, there was a mask. Can you take a picture of that?").

Interjections Indicating Confirmation

In addition to participants in the human agent condition using more declarative statements, they also used more interjections ($\bar{x}_H$=13.2, SD=6.9) than those in the robot agent condition ($\bar{x}_R$=3.6, SD=3.2); p=0.02 with $t(9)$=2.90. Fifty-eight of the 79 (73.4%) interjections in the human agent condition were categorized as confirmation or acknowledgment feedback to the remote shopper, in contrast to 9 of the 18 (50%) interjections spoken to the robot shopper. This difference was also significant (p=0.01 with $t(9)$=3.25; $\bar{x}_H$=9.7, SD=4.8; $\bar{x}_R$=1.8, SD=2.9). We believe this difference is due to the perception of giving instructions to a person versus robot. In both conditions, the wizard performed commands and provided feedback in the same manner. The webcam was used solely to provide the participant a view of the remote environment and not used to coordinate with the remote shopper in the human agent condition. Verbal acknowledgements are a compensatory strategy used in human-human remote collaboration (e.g., [Boyle et al., 1994; Eberhard et al., 2010; Gergle et al., 2004; Rosenberg, 2002; Trafton et al., 2005]).

Margo, stop!

It is imperative that the robot is able to stop on command. There were 23 utterances that contained the keyword "stop" (n_H=4, n_R=19). In 21 of these 23 utterances (91.3%), the participant was directing the remote shopper to cease the current movement. There was no statistical difference between the human agent condition ($\bar{x}_H$=0.3, SD=0.8) and the robot agent condition ($\bar{x}_R$=3.8, SD=6.9); p=0.36. Additionally, there were five utterances containing an implied stop command (n_H=4, n_R=1). Colloquialisms including "that's good" (n=1) and "[*blank*, hold it, stay] right there" (n=4) should also be given the same importance as "stop."

Social Etiquette

Six of the eleven participants greeted their remote shopper by name (n_H=3, n_R=3). Two participants introduced themselves by name as well: one in either condition (P9, P12). Four participants thanked their remote shoppers (n_H=4, n_R=3). P1 and P9 (human agent condition) said "thank you" once and three times, respectively; P2 and P12 thanked their robot shopper once and twice, respectively. Six participants said "please" a total of 26 times (human agent: n_{P3}=3, n_{P5}=5; robot agent: n_{P4}=2, n_{P6}=3, n_{P8}=5, n_{P12}=8). There was no significant difference between the human agent ($\bar{x}_H$=1.3, SD=2.2) and robot agent ($\bar{x}_R$=3.6, SD=3.0) conditions; p=0.20.

There were 47 instances in which a participant addressed the robot by its name, in addition to the three greetings. Addressing the robot by name was one strategy for giving a new command. Three participants said "Margo" in this manner (n_{P2}=40, n_{P6}=2, n_{P8}=2). It was also used for checking if the robot was still awaiting commands by P6 and P8 (n=3); participants gave a subsequent request following the robot's acknowledgement. In the human agent condition, the remote shopper's name "Kelsey" was spoken in 6 instances: three times as a greeting and three times at the start of a declarative sentence (P1, P3). There was no significant difference between the human agent and robot agent conditions (p=0.30). In a one-to-one scenario, addressing the remote shopper by name may have been considered superfluous; vocative expression is an explicit manner of direct address which people use in group situations [Akkeer, 2009].

4.4 Summary

Prior to this research, it was unknown how our target audience would want to direct robots in a remote environment. A number of corpora have been constructed investigating spatial commands; however, none involved people with disabilities as participants.

By investigating a speech interface, our intention was to mitigate associations that participants from our target population had with joystick controls, which are common input devices for operating power wheelchairs. We conducted a formative assessment of user expectation utilizing a participatory approach, performed an experiment with twelve participants from our target audience, and collected a corpus of their first-hand spatial commands. We can now begin to understand how people from our target audience would direct telepresence robots.

Our corpus provides a first hand account regarding the level of navigation functionality these robots are expected to have (answering RQ1). By investigating a speech-based interface and providing sufficient feedback, we drew one key insight: users would command the robots using multiple levels of abstraction. Users wanted to give directives at the low-level (i.e., forward, back, 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). Our robot's corresponding autonomous navigation behaviors for each level are described in Chapter 6.

Chapter 5

Accessible Human-Robot Interaction for Telepresence Robots

Telepresence robots can provide a means for remote social interaction with people or to visit places. However, it is not always obvious to a user as to *how* to operate a robot system beyond low-level FBLR directives (i.e., forward, back, left, right, stop). 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 [Nielsen, 1994b]. Human mobility follows a power-law distribution (e.g., by vehicle [Gonzalez et al., 2008], walking [Rhee et al., 2011], activities of daily living at home [Aipperspach et al., 2006]); we posit that users' intentions and, consequently, directives will also follow this pattern, which we observed in our formative assessments (Chapter 4). We developed a graphical user interface that leveraged this key insight.

Nielsen's 10 usability heuristics for designing user interfaces are the gold standard in HCI [Nielsen, 1994b]. However, Bergman and Johnson [2008] 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* [1993], "disabilities are only mentioned in a few brief sentences in the entire book." 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.” Fortunately, there have been several groups focused on disability issues and accessibility in HCI. The W3C group [W3C, 1994, 2008] and the Nielsen Norman Group [Nielsen Norman Group Report, 2001] have focused on web accessibility, and Bergman and Johnson [2008] 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 [1992] provide an introduction to disabilities. Our research draws upon our previous work [Tsui, 2008; Tsui et al., 2009, 2010], Nielsen’s usability heuristics [1994b], W3C’s guidelines [W3C, 1994, 2008], Kurniawan and Zaphiris’s web accessibility guidelines for older adults [Kurniawan and Zaphiris, 2005], and Vanderheiden and Vanderheiden’s guidelines [1992]. In this chapter, we highlight a number of these guiding heuristics and their manifestation in our telepresence robot user interface.

5.1 Description of Art Gallery Built for Case Studies

To give users an interesting environment in which to explore through a telepresence robot, we built an Artbotics gallery. 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 [Yanco et al., 2007b]. Each exhibit uses sensors such as IR distance sensors or buttons to allow a person to interact with the art. When the sensors are triggered, the robots 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 Figure 5-1. 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 8 $\times$ 7 foot plywood wall,

Figure 5-1: The gallery contained five Artbotics exhibits: (left to right) Sunflower (yellow background), Vincent (red background), Face (purple), Monkey (green background), and Music (blue background).

which was painted to match the theme of the corresponding exhibit. The exhibits were configured in a "U" shape with one facing back and two on either side perpendicular to the back (see Figure 5-1). The outer dimensions of the space were 24 × 16 feet, leaving a 23 × 14 foot interior area for the robot to move around. Three IR distance sensors were placed, centered, under a kick plate in front of each exhibit (see Figure 5-2). 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 approached from each angle and both a close and far distance.

5.2 Interface Overview

Our interface enabled the users to navigate through an art gallery. 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 semi-autonomous robot navigation behaviors were integrated into the interface design. (Details regarding the robot navigation follow in Chapter 6.) The interface was touch based, and users would touch and release elements to move the robot in the remote environment.

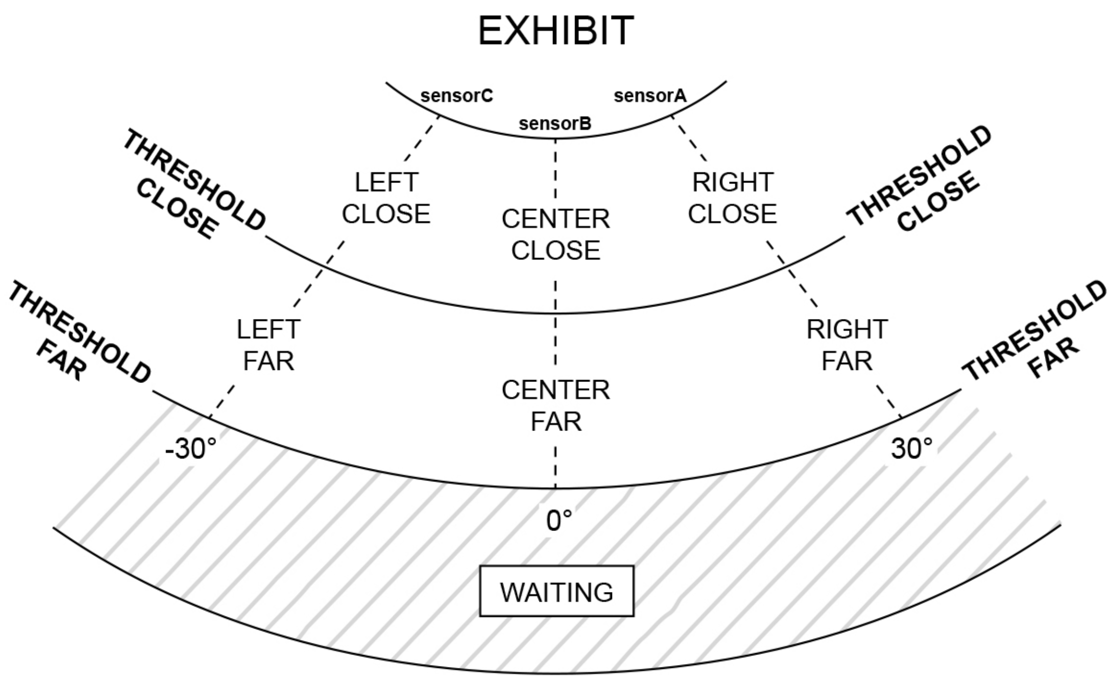


Figure 5-2: Three sensors were placed in front of and below each exhibit. The center Faced straight out while the left and right were aimed 30 degrees in either direction with respect to the center.

5.2.1 Components

The primary function of the interface was to provide an experience in which the user would be telepresent in the space, therefore much of the screen (63%) was comprised of the camera view (details follow in Section 5.3.1 on p. 108). The bottom of the camera view featured an iconified representation of the robot's base, referred to herein as the "robot base icon." All of the interactive components of the interface were position around the robot base icon (Figure 5-3, p. 97). Grouped navigation commands at bottom half of screen to prevented user from reaching and over extending upper limb (sustained activity) [Fitts, 1954].

As determined by our previous studies discussed in Chapter 4, 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.

To provide global navigation, five buttons were positioned around the robot base

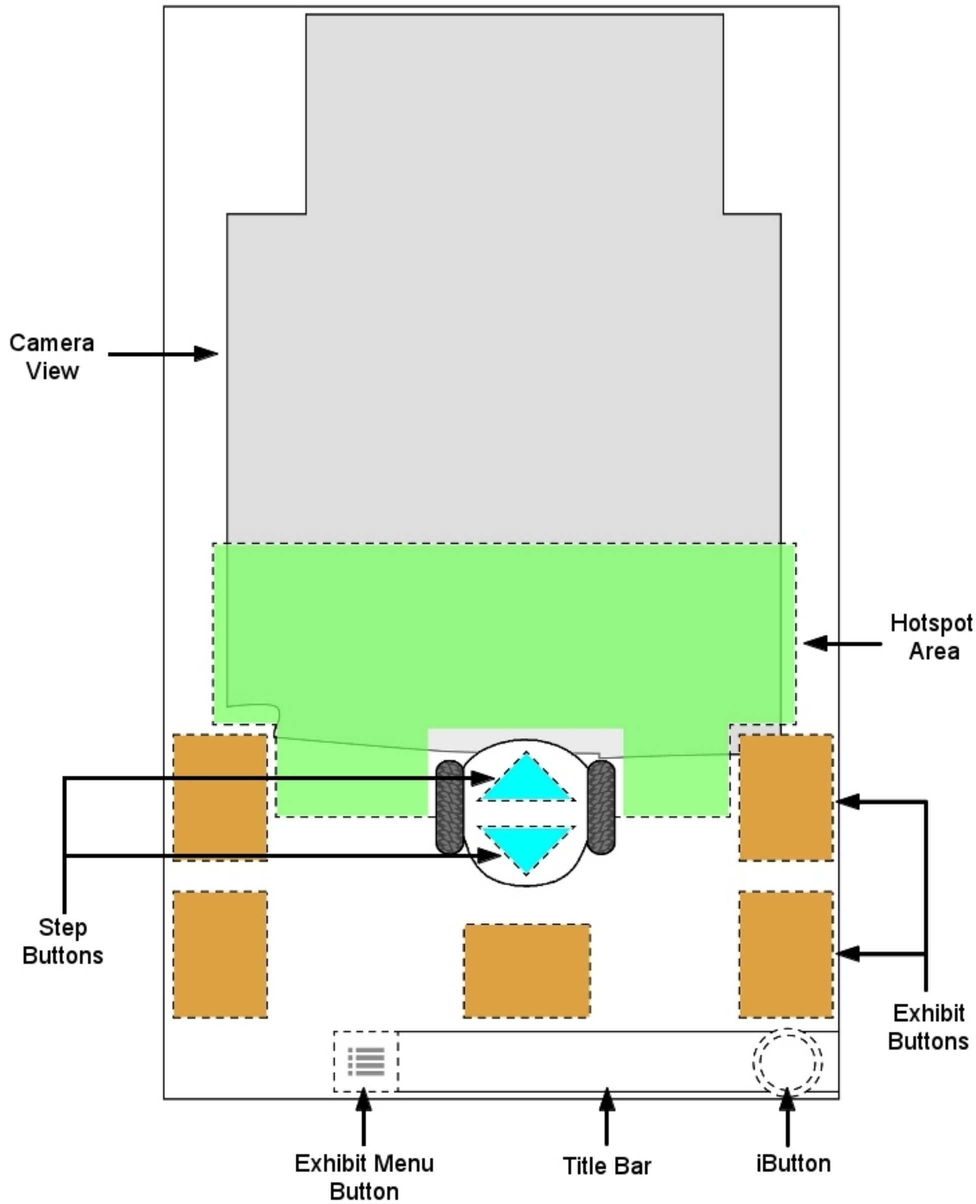


Figure 5-3: Wireframe representation of user interface. Dotted lines denote areas of the interface which are interactive. The large grey box at the top of the screen denotes the area taken by the video container. The area highlighted in green represents the area where the exhibit hotspots will appear when on screen. The area highlighted in orange indicate the position of the exhibit buttons are positioned when on screen. The cyan arrows indicate the positions of the forward and reverse step buttons.

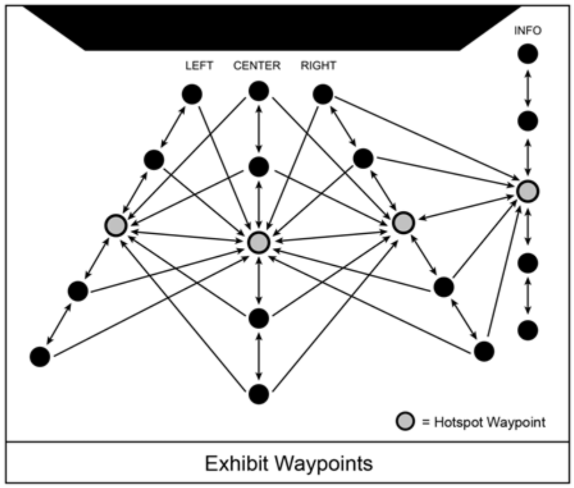


Figure 5-4: Connected graph of the robot's movement forward and backward from hotspots corresponding to step button selections.

icon. 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. Figure 5-3 shows the locations of the "exhibit buttons" denoted as orange rectangles. 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.

Local navigation once at an exhibit was represented as pulsing rings around the robot base icon. These pulsing rings, herein referred to as "hotspots," represented the real-world locations around each exhibit. (Chapter 6 details how these positions are calculated.) The area in which the hotspots were dynamically positioned on the screen is denoted by the green highlighted area in Figure 5-3.

Finally, low-level control of the robot was provided by two cyan arrow buttons, herein referred to as "step buttons" (Figure 5-3, p. 97). 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 reverse a maximum of two times each from the hotspot, as shown in Figure 5-4.

In addition to the navigation and low-level movement controls, there was one other component of the interface which is broken into three parts. The "title bar," positioned at the bottom of the screen below the exhibit buttons, provided additional an user control to the robot (Figure 5-3, p. 97). At the leftmost part of the title bar, there was the menu button, which, when pressed, opened the exhibit menu as a pop-up modal (Figure 5-14, p. 121). The center component gave the user more information about the current status of the robot in plain text (4.75 em). To the right of the title bar was the "iButton." The iButton both displayed an iconified version of the system state, and allowed the user to change the robot's state described below.

5.2.2 Implementation

Our interface was implemented as a web-based application. 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 [Desai et al., 2011]. Our interface was designed for a 22" touchscreen monitor in portrait orientation with a resolution of 1050 × 1680 pixels.

The interface was written using HTML5 and javascript. Communication with the robot was provided through roslibjs [Toris, 2013]. This allowed the interface to communicate with the robot bi-directionally, both receiving system updates and sending commands to be executed.

All of the buttons in the interface were given a visual affordance 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.

The visual affordances [Gibson, 1979; Norman, 2002] were different for each type of button. The exhibit buttons were rectangular in shape (130 × 180 px, 36 × 51 mm). They were designed to have 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. Each exhibit button was colored with its exhibit's corresponding background color, and an pictoral icon of the exhibit was placed in the center (Table 5.1). 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.

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 [Urrutia, 2012]. The hotspots were outlined with a 10 pixel white border, and the color of the center matched the exhibit color (Table 5.1). 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 on a hotspot, the pulsing ceased and the opaque hotspot displayed at its maximum size until it was released.

The step buttons were triangular buttons. They were bright cyan colored in their default state, and dark cyan with a grey gradient when pressed. The step buttons also utilized a similar shadowing technique as the exhibit buttons.

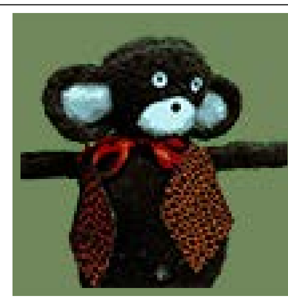
For all buttons, if a release event occurred within the bounds of the same button, we considered it a selection and the command was sent to the robot. Otherwise, the touch or click was ignored. In both cases, the buttons would revert to their original state upon release.

5.2.3 User Access Method

All buttons were placed in the bottom half of the touchscreen to facilitate our target population's access given the large size of the monitor. The 22" 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 (Figure 5-5 right). For users with limited motor skills in their hands, we designed an optional clear acrylic keyguard with cutouts for the elements of our interface (Figure 5-5).[1]

[1] 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

Table 5.1: 100 × 100 px icons of each exhibit. HTML colors sampled from painted exhibits' background.

Exhibit name	**HTML background color**	**Icon (100×100 px)**	**Resulting button fill**
Sunflower	**#BBBC56**		
Vincent	**#7C2625**		
Face	**#4C2F5D**		
Monkey	**#70875B**		
Music	**#71809D**		

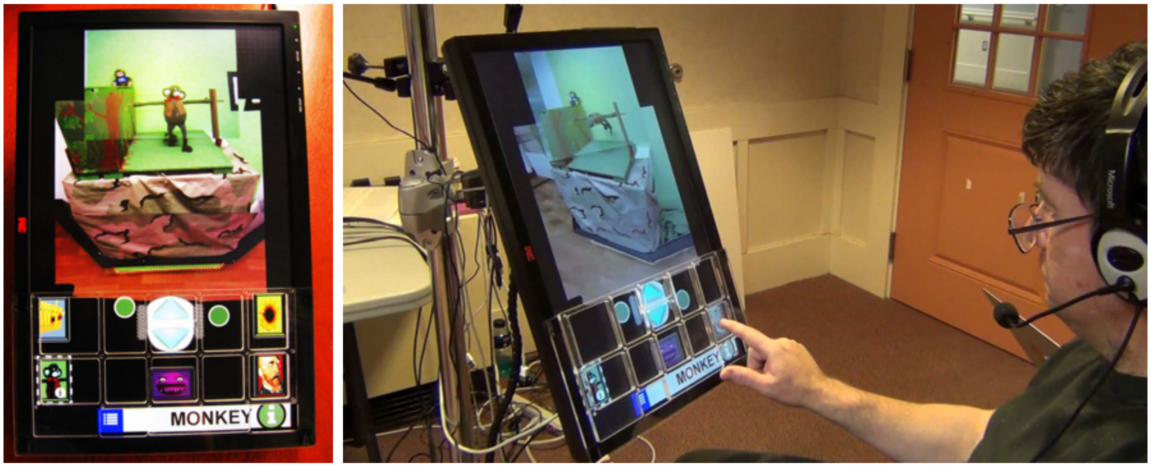


Figure 5-5: **(Left)** A custom clear acrylic keyguard is shown over the lower third of the interface with cutouts for the exhibit buttons, step buttons, area around the robot base icon, menu button, and iButton. **(Right)** The 22" monitor (shown with keyguard) was mounted on an Ergotron cart with an adjustable swing arm, which allowed it to be placed in a manner that was comfortable for the user to see and manipulate.

Alternatively, the interface could support any access method that emulated a mouse cursor (e.g., computer mouse, RollerBall2 Joystick).

5.3 Heuristics and Guidelines At Work

We drew from both HCI interface design guidelines (i.e., Nielsen's usability heuristics in Table 5.2), guidelines for accessible web design (i.e., W3C in Table 5.3, Kurniawan and Zaphiris's guidelines for older adults in Table 5.4) and accessible consumer products (i.e., Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging in Tables 5.5 and 5.6), and our previous research (Appendix D). The guidelines in Tables 5.2 through 5.6 are a subset, and concepts may be repeated with slightly different wording or emphasis from their authors.

Our telepresence user interface is rife with subtle design choices, which balanced

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. 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.

Table 5.2: Relevant Nielsen Usability Heuristics [1994b; 1995]

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 5.3: Relevant W3C Web Content Accessibility Guidelines [2008]

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 5.4: Relevant Kurniawan and Zaphiris's Web Design Guidelines for Older Adults [2005]

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 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.

Table 5.5: Relevant Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [1992] (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 5.6: Relevant Vanderheiden and Vanderheiden's Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging [1992] (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.

accessibility and usability based our experience [Tsui, 2008; Tsui et al., 2009, 2010]. To create an effectively "invisible-to-use" interface [Takayama, 2011], our design approach was to use a minimalist aesthetic (N-8) and 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, 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." This design choice supports accessibility and maintains the user's focus and sense of being telepresent in the remote environment.

5.3.1 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 foundation 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 [Tsui and Yanco, 2013]. Chen et al. [2007] found that a 45 degree FOV was regarded as too small for remotely driving a vehicle.

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 (Figure 5-6) (KZ-3). The images from each of the Logitech c910 webcams were 864 $\times$ 480 px, and the Asus Xtion was 640 $\times$ 480 px. The individual video streams were rectified to remove lens distortion [Hole et al., 2002; Riek, 2007] prior to tiling the images together. Figure 5-7 (p. 110) shows the hat stitcher subsystem; 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-9).

Figure 5-8 (p. 111) shows our interface with its vertical panoramic camera view and the VGo App, 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

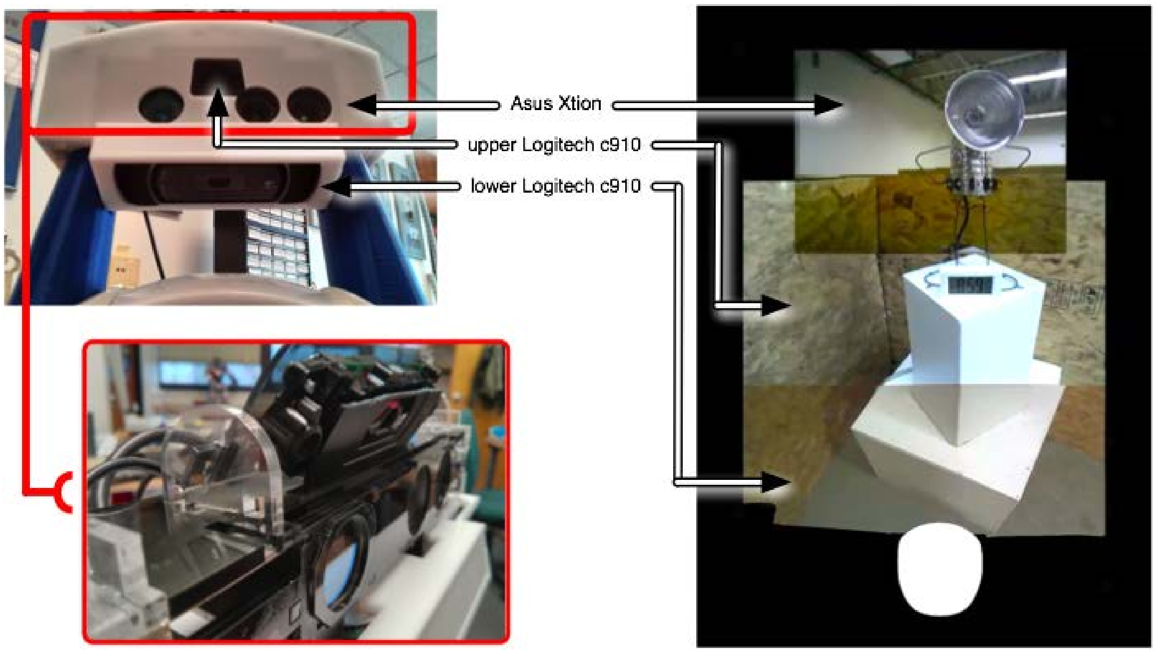


Figure 5-6: Margo's three cameras in the hat are stitched together in a single vertical panoramic video stream. The camera and corresponding part of the view are shown.

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 [Chellali and Baizid, 2011]. However, users may still achieve a physical real-world frame of reference [Mine et al., 1997] in the remote environment through a sense of the location of parts of the robot or other objects via video images [Carruthers, 2009].

The camera view showed a portion of the robot in its view, which provided the user scale in the remote environment [Keyes et al., 2006]. However, the Hawaiian shirt was visible in the camera view in addition to the robot's base, which yielded an inflated sense of the robot's size. We corrected this issue by placing a black mask over the robot's stalks, shirt, and base in the video (Figure 5-9).

To reinstate the proper sense of scale, we designed a to-scale icon of the robot's

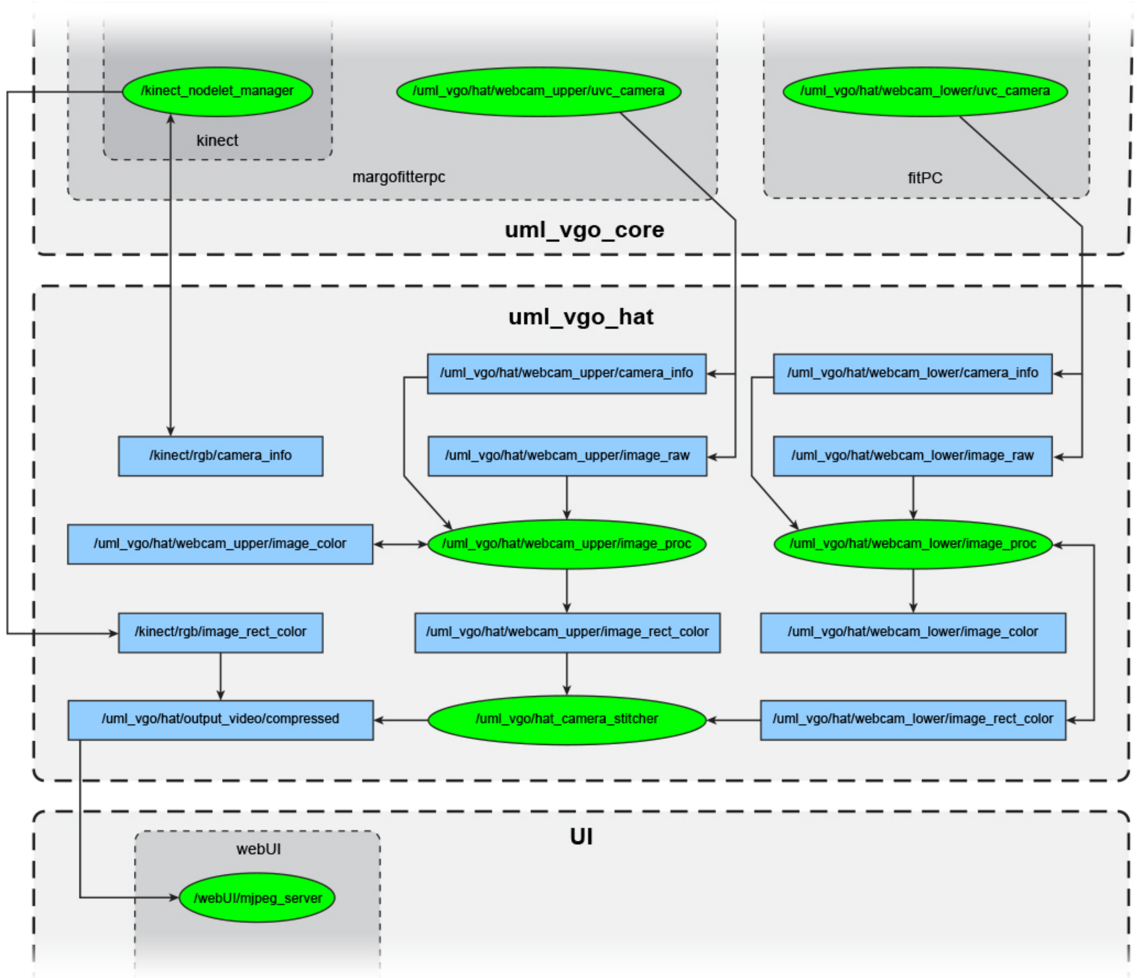


Figure 5-7: Diagram of hat stitcher subsystem showing ROS nodes in green and topics in light blue. The full system diagram is shown in Figure 3-11 on p. 56.

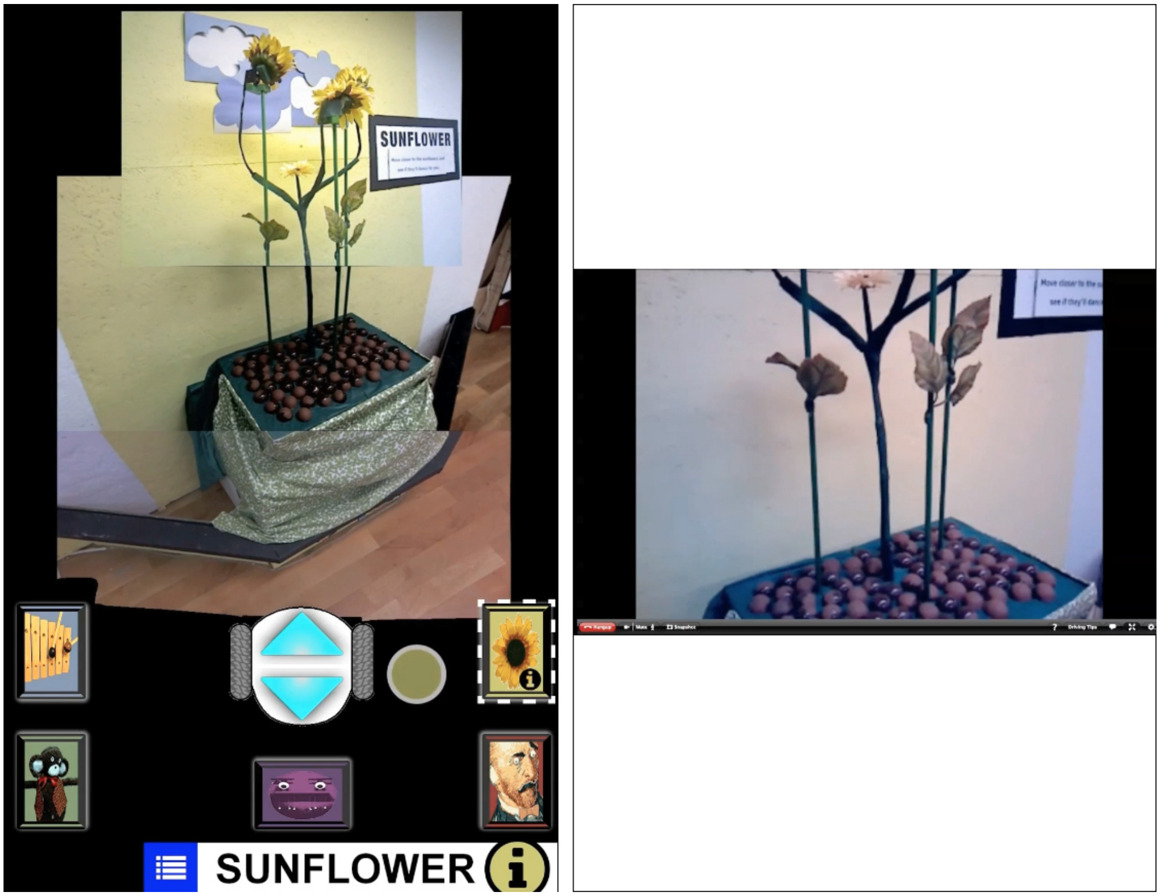


Figure 5-8: The robot is located at the left hotspot of the Sunflower exhibit. Our interface is shown on the left, and the VGo App on the right.

Figure 5-9: Demonstration of pre- and post-image mask of Margo's shirt front and stalks in its camera view (left and right, respectively).

Figure 5-10: Design of the robot base icon. **(Left to right)** The base was photographed from the height corresponding to the downward facing webcam. Its shape was filled with solid white, then outlined in black. Finally, dark grey tires with light grey tread were added on either side.

base. Figure 5-10 details the process of developing the robot base icon. From the height corresponding to the downward facing webcam, we took a picture of only the robot's base. The contour of the robot's base provided an outline for the icon. The final design of the robot base icon included two wheels, one on either side (Figure 5-10 right). This depiction is an exaggeration of the robot's drive wheels, which were recessed under the base and therefore not visible. The rear casters were also recessed under the base but were omitted from the robot base icon. While the width of the wheels may be exaggerated, it does, however, provide the user a grounded frame of reference (KZ-4).

5.3.2 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 state of the system. In addition to seeing the robot's video update, we have 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, as shown in Figure 5-11 (p. 119). 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. (The implementation of the animation is described in the following chapter.) The Giraff, Texai, and Beam user interfaces employ a similar feedback technique by showing a vector originating from the bottom center of the video on the interface [Coradeschi et al., 2011; Giraff Technologies AB, 2011; Gorman, 2012; Guizzo, 2010; Suitable Technologies, 2012].

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. 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). If the robot handler felt as though the robot was acting in an unintentional manner, he or she could activate the robot's emergency stop. The iButton also reflected this state by displaying an orange exclamation point in a bright red circle. 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. 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). Figure 5-12 shows these four states on p. 120.

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 Figure 5-11. When the robot reached the selected exhibit, the title text showed "<**exhibit_name**>" at full height, and the interface played a short audioclip of an arrival sound (W-1b).

5.3.3 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 for control. The user had limited low-level control and could move the robot two positions forwards or backwards 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). 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).

5.3.4 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. 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). We utilized the Gestalt principle of similarity [Koffka, 2013]. All navigation buttons including the menu buttons, exhibit buttons, and hotspots were color matched to the exhibit that they represented (V-1, V-4). Every exhibit had four hotspots, positioned similarly around the exhibit and the wall information sign (N4). These hotspots correspond to actual positions in the environment from which the robot can view an exhibit.

The exhibit buttons were 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).

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. 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 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 and always available. 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).

5.3.5 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 Figure 5-13 right (p. 120), only two buttons were displayed when the robot was moving (i.e., menu button and pause), which both caused the robot to pause as previously described. 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 [2010] found that widgets on a touchscreen should have a minimum size of 30mm. All buttons in our interface were a minimum of 100 × 100 pixels in size, which meets this standard. The majority of our buttons exceed Micire's findings, with the largest button having a touchable area of 244 × 122 pixels (79.5 × 66 mm). Due to their triangular shape, the step buttons featured additional padding to allow clicking around the triangle (transparent).

Gestalt principles of grouping [Koffka, 2013] 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 ordering (Figures 5-13 (left), p. 120, and 5-14, p. 121). The hotspots were positioned around the robot base icon emanating from it. The forward and reverse step buttons were overlaid on the robot base icon (V-4), which also adhered to Shneiderman's theory of direct manipulation [Shneiderman, 1993]. The shape of the buttons corresponded to functional groupings as well (V-4); exhibit buttons were rectangular, hotspots circular, and step buttons triangular.

Coloring and contrast were used to make the buttons highly visible [Nielsen Norman Group Report, 2001; Vanderheiden and Vanderheiden, 1992] and emphasize groupings [Koffka, 2013]. 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. Except for the Sunflower exhibit, the information symbols were white which highly contrasted with the saturated red, purple, green, and blue background colors. The background color for the Sunflower exhibit was a light yellow; black information symbols were used in this case. Although translucent, the hotspots' white borders created contrast against both the exhibit color, and the interface's black background (KZ-13) and the gallery's beech wood colored flooring if overlaid on the camera view.

Several interface elements employed a static color profile, namely the interface's black background and the robot base icon. The title bar was a solid white color, which contrasted highly with the black of interface's background and the text displayed in it. The iButton's black border created contrast between its contents and the title bar. The menu icon was white and the button's color was the maximally saturated blue (i.e., #0000FF). The color cyan was chosen for the two step buttons; the buttons' black outline and grey shading created contrast against the robot base icon. The blue of the menu button and cyan of the step buttons were not used elsewhere in our interface. The Animated Vector Indicator (AVI) emanating from the robot base icon was a dotted white line with each dot and arrowhead outlined in black to ensure sufficient contrast, like the hotspots, against both the interface's background and the gallery's flooring.

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.

5.4 Summary

Our goal was to design an "invisible-to-use" [Takayama, 2011] telepresence robot user interface. We drew from HCI interface design guidelines (i.e., Nielsen's usability heuristics [Nielsen, 1994a]), guidelines for accessible web design [Kurniawan and Zaphiris, 2005; W3C, 2008] and accessible consumer products [Vanderheiden and Vanderheiden, 1992], including our previous work developing heuristics for assistive robotics [Tsui, 2008; Tsui et al., 2009, 2010].

Nielsen's 10 usability heuristics for designing user interfaces are the gold standard in HCI. Due to their generality and simplicity, we largely drew from Nielsen, augmenting his guidelines with ones more specific to accessibility (RQ3). The vertical panoramic camera view and the to-scale robot base icon were the two primary examples of Nielsen's heuristic of "match between system and the real world" (N-2); these two elements were the foundation of our interface. We implemented Nielsen's heuristic regarding feedback to the user (i.e., "visibility of system status" (N-1)) using three levels of status indicators into the interface in addition to seeing the robot's video update.

Our goal was to maximize the number of users from our target population who could perceive, understand, and operate the interface [Vanderheiden and Vanderheiden, 1992]. Regarding the interface's accessibility, we utilized large buttons and high contrast coloring. We minimized the number of elements displayed and the use of plain text, opting for iconic representations where appropriate. Finally, we opted for a single point "touch and release" interaction; both auditory cues and shadowing were employed to provide feedback to the user regarding touch selection.

While our interface was designed for a remote art gallery scenario, these design principles can be applied to any telepresence robot user interface. In Section 8.3.3, we further discuss how our interface might change if the robot were to be used in a larger gallery or museum, or a different type of environment (e.g., home, school, office) altogether.

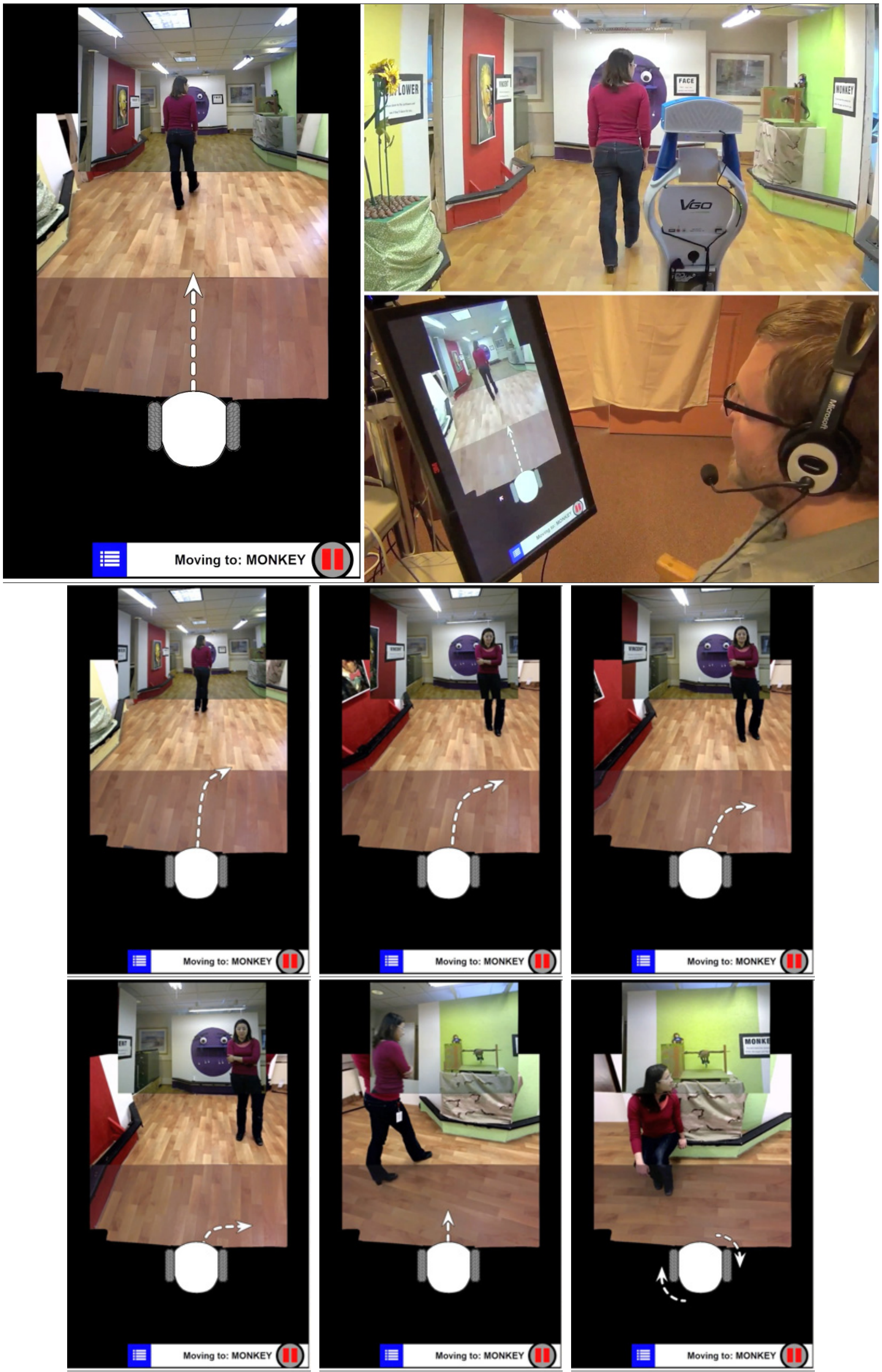


Figure 5-11: Storyboard depicting the Animated Vector Indicator (AVI) states. Each stage is shown as the user and interactant moved from the gallery's entrance to the Monkey exhibit.

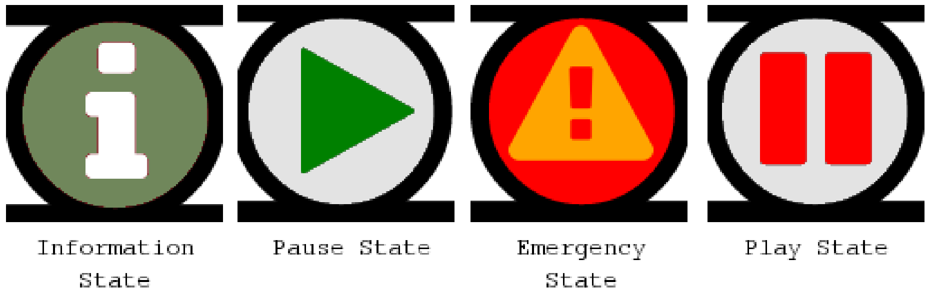


Figure 5-12: Four states of the iButton. **(Left to right) Information State:** Clicking on this state will display the information about the exhibit on screen, in this case Monkey. **Pause State:** While the robot is paused, a green play button is shown while the robot is paused, so that the user may resume movement. **Emergency State:** Displays and disables all click events while the robot is stopped remotely. **Play State:** While the robot is in the play state (moving), a red pause button is shown so that the user may pause movement.

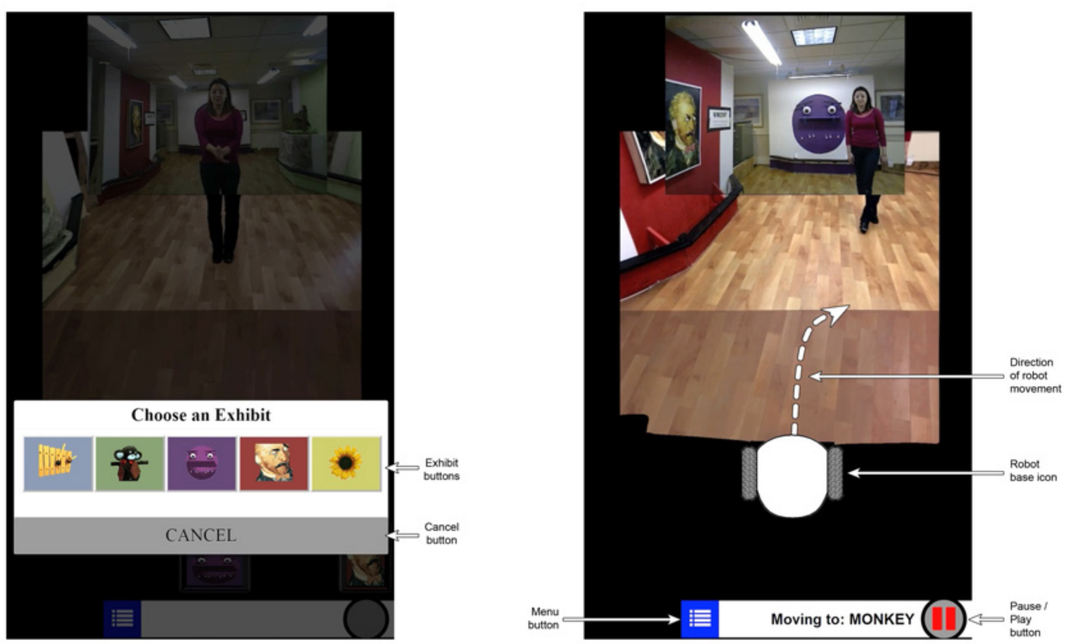


Figure 5-13: Margo's custom user interface; moving to an exhibit. **(Left)** Exhibit menu with all five exhibits shown. **(Right)** To scale robot base icon overlaid on the robot's combined video stream; the Animated Vector Indicator (AVI, white dotted arrow) shows the robot's direction of travel towards selected exhibit. Exhibit menu button and pause/play button are located to the left and right of the title text, respectively.

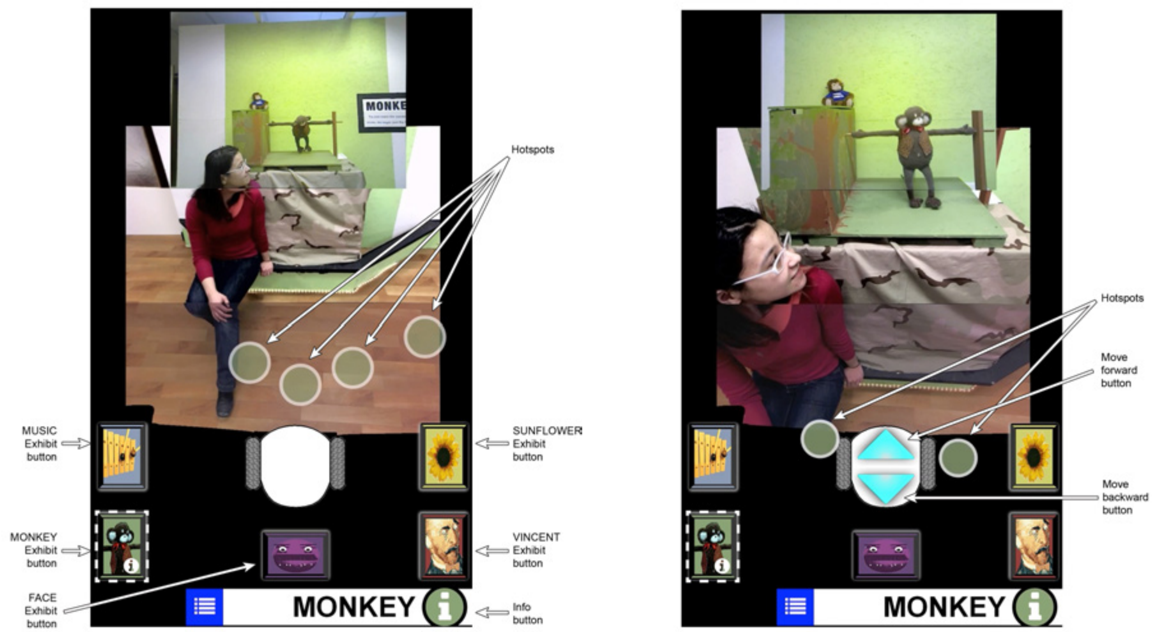


Figure 5-14: Margo's custom user interface; moving within an exhibit. **(Left)** Once at the selected exhibit, exhibit buttons shown around the robot base icon. Pause/play button replaced with information "i" symbol. Exhibit's on-screen information accessible via iButton or corresponding exhibit button, also marked with information "i" symbol in lower right corner and outline with white dotted line. Four hotspots overlaid on the robot's combined 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. **(Right)** Robot arrived at *center* hotspot; adjacent *left* and *right* hotspots appear on either side of the robot base icon. Additionally, the forward/backward step buttons (i.e., two cyan triangle buttons) appear overlaid on robot base icon.

Chapter 6

Movement and Navigation Behaviors

The telepresence robot served a the user's physical embodiment in the remote environment. To support our "invisible" interface [Takayama, 2011], robot's movements and navigation behaviors had to approximate that of a human's as much as possible. Motion-sickness inducing point turns were replaced with wider arcing turns for orientation. Smooth acceleration and deceleration profiles prevented the robot and its cameras from abruptly rocking.

6.1 Requirements

As stated in Chapters 3 and 4, a robot designed for use by our target population had many requirements. The robot must be able to map, navigate, and move through a remote location commanded by the user. To achieve this, all sensors and actuators on the robot were accessible through ROS allowing us to write the mapping and navigation software required. In the following sections, we detail how these systems were designed and implemented.

6.1.1 Levels of User Control

Teleoperation requires the user to pay attention to robot surroundings. Although beneficial for an operator's situation awareness [Endsley, 1988], it can require the user to spend substantial time thinking about how to move the robot, which can reduce telepresence since the robot embodiment becomes visible to the user. As a means to increase telepresence, robots can provide navigational assistance through semi-autonomous or autonomous behaviors. Such navigation behaviors should strive to balance a user's cognitive workload and engagement [Cesta et al., 2013].

We have observed in our own research, as have others [Kim et al., 2010], that people with disabilities dislike systems that do everything for them; people have stated that using fully autonomous systems makes them feel lazy. Additionally, fully autonomous behaviors have traditionally led to the "out-of-the-loop" performance problem for robot operators [Kaber and Endsley, 1997], who may lose all sense of engagement during the robot's execution [Draper et al., 1998; Riley et al., 2004]. The reduction in cognitive load comes at the cost of lost situation awareness and a potential reduction in a sense of telepresence. The user becomes a passenger, and while waiting for the telepresence robot to arrive to the specified destination, the user may switch his or her attention away from the remote environment, negatively impacting telepresence.

Semi-autonomous behaviors are a form of shared control in which the user can identify a high-level goal with the robot controlling the low-level navigation. For example, the Jazz robot has the capacity to move to a local waypoint in the user's video [Dickert, 2011]. The robot's neglect time, or the elapsed time until the robot requires user intervention [Crandall et al., 2005] to receive a new waypoint, is a function of the distance between the robot's current location and the user's selected waypoint. The user has less cognitive workload and engagement when specifying only a final destination waypoint in the global environment than when he or she needs to periodically specify waypoints within a robot's field of view. When using semi-autonomous behaviors or teleoperating the robot, the user may casually observe the remote environment itself and choose to exchange greetings with interactants and bystanders.

As described in Chapter 5, our alternative user interface supports multiple levels of user control over the robot's movements. Users are given high-, mid-, and low-level control of the robot's movement. High-level control is provided by allowing the user to change exhibits. Mid-level control is provided in the form of local, within-exhibit hotspots. Finally, low-level control is possible through the forward/reverse step buttons while at any of an exhibit's hotspots.

6.1.2 Movement Predictability Given Embodiment Constraints

In order for autonomous and semi-autonomous robots to be accepted, a user's confidence must be gained through both technical reliability and reassuring the user that it is working correctly for all scenarios [Norman, 1994]. A robot with any level of autonomy must perform as the user expects from the beginning [Desai, 2012; Desai et al., 2013, 2012]; it is easy for users to lose trust in a system, and once lost, is hard to regain [Desai, 2012; Groom and Nass, 2007; Sanders et al., 2011].

Kruse et al. [2013] performed an extensive survey of human-aware robot navigation. The premise of telepresence robot is that it is an embodiment for its user. Therefore, the robot's movements must approximate that of a human's as much as possible. Our approach was to develop a semi-autonomous robot system to allow users to move in manner similar to in-person visitors [Wineman et al., 2006].

The embodiment of a particular social telepresence robot can constrain its ability to move like a human. Differential drive or skid steer robots, like cars, cannot independently translate from side to side; instead they have to move forward or backward, whereas a person can simply take a step to the side. Holonomic robots can move in any direction regardless of their orientation (e.g., iRobot's Ava [iRobot Corporation, 2011], iRobot and InTouch Health's RP-VITA [InTouch Technologies, Inc., 2012a], iRobot and Cisco's Ava 500 [iRobot, 2014], University of Sherbrook's Telerobot [Michaud et al., 2010]).

The choice of the type of movement impacts how the robot can approach people and objects in the environment. With a holonomic robot, the user can directly command the robot to move, or translate, to the left or right in order to achieve a different viewpoint. With other robot types, the user needs to perform 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. If the telepresence robot has a fixed forward facing camera, both of these techniques yield movements that appear contrary to the user's goal (i.e., turn away, back away). The user's sense of being telepresent may decrease since the nonhuman telepresence robotic embodiment becomes visible. As described in Chapter 3, Margo has two drive wheels and two passive rear casters; additionally, Margo's three hat cameras are fixed (i.e., unable to pan or tilt). It was therefore necessary for us to address this issue.

6.2 Design of the Navigation Behaviors

Our goal was to design two distinct navigation behaviors for the telepresence robot's movements in the gallery. First, when the robot is at an exhibit and the user specifies a movement to another waypoint 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. Second, 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. Using our simulation setup (Section 3.6.8, p. 62), we were able to thoroughly test the navigation behaviors during development before running them on the physical robot.

6.2.1 ROS Navigation Stack

ROS includes an implementation of algorithms for navigation, localization, and mapping, which are contained within the `navigation` stack [Marder-Eppstein, 2014]. Our system uses the stack's `amcl` and `map_server` nodes for localization, previously discussed in Section 3.6.4.

We considered the `navigation` stack's `move_base` node for global and local navigation; however, the `move_base` planner was unable to satisfy the aforementioned requirements. Fundamentally, `move_base` plans a route of waypoints (rather than poses) and then follows the route. The robot rotates in place at the end of the path to meet a specific orientation, if necessary.

The ROS `move_base` node requires a global planner, global costmap, local planner, and local costmap. It allows the use of different implementations for either of the planners via ROS plugin architecture. Costmaps are populated with "cost" values, where each cell's cost is based on the geometry of the robot and obstacles perceived via sensors (e.g., laser rangefinders, IR). The standard implementation for the global planner is provided by `navfn` package. For `navfn`, another component of the cost is the distance between a given cell and the starting location. Once the costs are calculated, `navfn` uses Dijkstra's algorithm to calculate the global plan, which is the shortest path between the robot's current location and the goal [Marder-Eppstein, 2014; ROS Answers forum, 2013].

We were able to approximate keeping the robot in the center of the gallery open space by increasing the inflation radius in global costmap parameters. However, it did not provide consistent approach trajectories when transitioning between the exhibits. The local planner produced erratic velocity commands while attempting to adhere to the global path planned. For example, the robot turned too much in one direction and then turned back in the opposite direction to correct itself. This unpredictable behavior would have had a negative impact on user's perception of system's reliability. It should be noted that the `move_base` planner was designed for holomonic systems, specifically Willow Garage's PR2.

Furthermore, because Margo was unable to translate laterally, we were unable to configure `move_base` in such a way that would allow the robot to maintain a point of interest (i.e., an exhibit) within the camera view when visiting waypoints placed around it. In order to fulfill this requirement, Margo must first back away; the `base_local_planner` does not support backward translation movements beyond performing an escape recovery behavior [Marder-Eppstein and Perko, 2014], which is true of any of the local planners provided with ROS `navigation` stack. With a standard local planner, the robot would turn away from the exhibit, drive forward to the new location, and then turn back toward the exhibit, as previously mentioned.

As an alternative solution, the desired navigation behaviors could have been approximated by pretending that our non-holonomic system is differentially constrained (similar to a car) with a large turning radius. A sufficiently large turning radius would prevent the robot from turning far enough for the exhibit to disappear from the camera view while moving, which could potentially have been accomplished using a technique similar to Pivtoraiko et al. [2009].

Therefore, we created a custom navigation component (movement component of the system, Figure 6-1) to satisfy our requirements given the system constraints.

6.2.2 Custom Navigation Implementation

Like many other planning systems, our navigation software was divided into a global path planner (`/path_manager`) and a local planner (`/nav`). 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 visible to the user. Alignment poses and retreat poses are used to transition between the hotspots, while satisfying the movement requirements.

Terminology

Pose Representation. Each waypoint pose, in addition to specifying position and orientation for robot to match, also needed to have special properties:

Speed. The waypoint's speed is specified as a multiplier to the default speed. Local planner `/nav` would use the multiplier to adjust generated velocity commands while moving towards the waypoint.

Radius. The radius is used to determine how close the robot needs to get to the position specified by the waypoint to clear the waypoint.

Reversibility. We had two types of waypoints: *reversible* (read: can drive in reverse towards it) and *forward-only*. Reverse waypoints allowed backwards movement towards the waypoint (depending on the robot current orientation). Forward-only required the robot to first orient itself to face the waypoint.

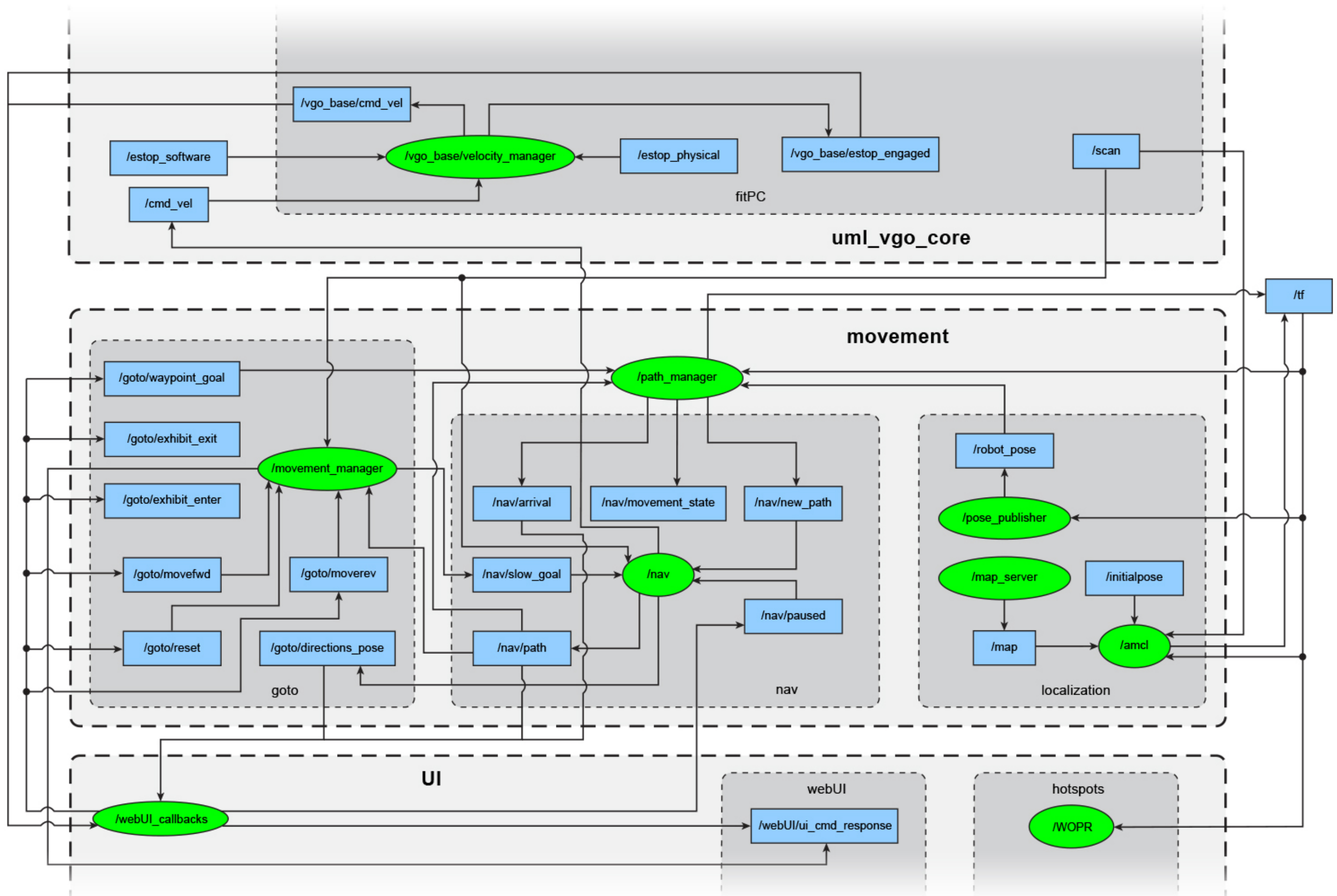


Figure 6-1: System diagram of movement component, implementing global navigation (`/path_manager`) and local navigation (`/nav`); ROS nodes are shown in green and topics in light blue. The full system diagram is shown in Figure 3-11 on p. 56.

Table 6.1: Each navigation waypoint used standard ROS `Pose` message, with orientation and z component of position remapped to represent the waypoint properties (radius, orientation agnosticism, reversibility, speed multiplier).

Position component

x	x	Retains its original meaning of position in Cartesian space.
y	y	Retains its original meaning of position in Cartesian space.
z	radius	Radius required for achieving the waypoint from x, y.

Orientation component (quaternion)

w	orientation agnosticism/yaw	Rotation around z axis (upwards from the ground in a right-handed coordinate system). If $w = 0$, the waypoint disregards orientation (i.e., it is orientation agnostic). If $w \neq 0$, the waypoint's orientation is observed, and w indicates the waypoint's yaw in radians.
x	reversibility	If $x = 0$, the waypoint is reversible and can be approached using forward or backward translation. If $x = 1$, the waypoint is forward-only and can only be approached using forward translation.
y	speed multiplier	If $y \neq 0$ use the value as a factor in the waypoint approach speed calculation. If $y = 0$, do not modify the speed (i.e. $y = 0$ is equivalent to $y = 1$).
z	–	Unused.

Orientation Agnosticism. For an orientation-agnostic waypoint, the orientation is disregarded. If the waypoint is not orientation-agnostic, then upon arrival to the waypoint, the robot should match a specific yaw.

Because Margo only moved on the 2D ground plane, we repurposed unused components of each pose to reflect these properties. Using the standard ROS path message allowed us to reuse available tools for visualization within Rviz. The resulting waypoint encoding is shown in Table 6.1.

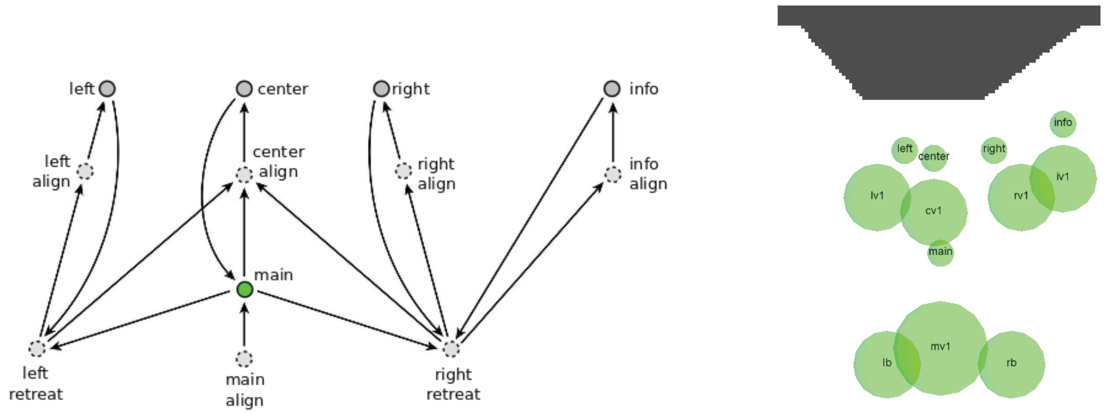


Figure 6-2: **(Left)** Connected graph of the robot's movement within an exhibit via its entry pose *main* and four hotspots (*left*, *center*, *right*, and *info*). **(Right)** Radii associated with poses at the Monkey exhibit. Hotspots are 0.1m, alignment poses are 0.25m, and the main alignment pose is 0.35m.

Named Exhibit Poses. Each exhibit has an entry pose *main*, shown in Figure 6-2. There are four hotspots – *center*, *left*, *right*, and *info* – from which the robot can view an exhibit. The entry pose and each hotspot has a corresponding alignment point, indicated as dotted circles in Figure 6-2. 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, the orientation of the exhibit, and several parameters such as the distance from the hotspot to the exhibit and distance between hotspots and alignment points.

Custom Global Planner

The `/path_manager` node began and remained in its "idle" state, waiting for the user to provide a goal (Figure 6-3). When the user selected a new exhibit or a hotspot, the `/webUI_callbacks` node published the goal to `/path_manager` on the topic `/goto/waypoint_goal`. `path_manager` parsed the goal and if it was valid (i.e., a known, named pose), transitioned from its idle state to its "departing" state. It

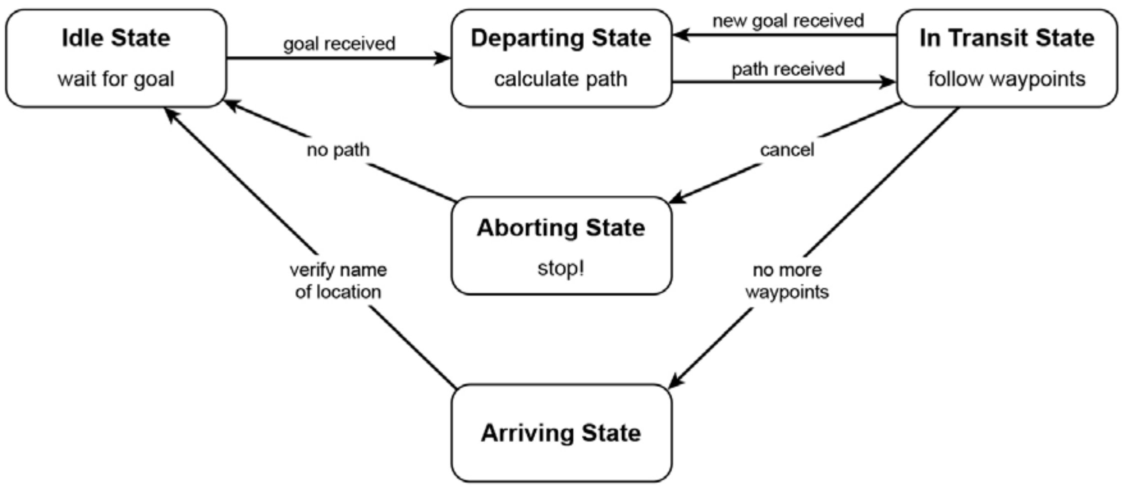


Figure 6-3: State diagram of global path planner.

then planned a path from the robot's current location to the goal. Upon successful planning, `path_manager` changed to its "in transit" state and moved to each of the waypoints in its plan. As the waypoints were reached, they were removed from the plan. When the plan contained no waypoints, the `path_manager` changed to its "arriving" state. Finally, `path_planner` checked that name of the nearest pose (directly below the robot) matched the goal and published it to the `/nav/arrival` topic before transitioning back to its idle state.

There were two basic routing scenarios: intra-exhibit and inter-exhibit. First, the intra-exhibit path planning considered only the graph of waypoints for the current exhibit (Figure 6-2 left). `path_manager` 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. For example, as shown in Figure 6-4 (p. 133), to move from the exhibit's *center* hotspot to its *right* hotspot, the `path_manager`'s global plan would first have the robot move backwards to the *right retreat*, then move forward through the *right align* pose, 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 navigation behavior of always keeping the exhibit in view while changing viewpoints.

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 `path_manager` accomplished this navigation behavior in two stages. First, it computed the exit path from the closest pose from the robot's current location to a corresponding retreat pose at the current exhibit. Next, the `path_manager` computed an entry path from the alignment point to the main entry pose of the destination exhibit. Finally, the resulting path was the exit and entry paths concatenated together.

When the robot was positioned at one exhibit, all other exhibits were available for the user to select (Figure 6-5, p. 134). Additionally, the user could select a new exhibit even while the robot was executing its current plan (Figure 6-3, p. 131).

Custom Local Planner

Overview. The local planner received a path either from global planner or from `/movement_manager` (step buttons movement). Local planner then iterated over the path and calculated velocity commands, published them on `/cmd_vel` topic, making robot traverse the given path according to each of the waypoint's properties, as previously defined in Table 6.1. During the movement the remainder of the path was published on `/nav/path` topic. Each time a waypoint was reached, the waypoint was removed from the path.

The local planner accepted a path via two topics: on `/nav/new_path` (used by `path_manager`), and on `/nav/slow_goal` (used by `/movement_manager`); see Figure 6-1 (p. 128). `/movement_manager` handled requests generated when user interacted with an exhibit via step buttons (stepwise forward and back movement). For each interaction it published a path containing single waypoint. The waypoint was offset forward or backward from the robot by an appropriate step size (13 cm for the Face exhibit, and 16 cm for all other exhibits). For all the paths waypoints received

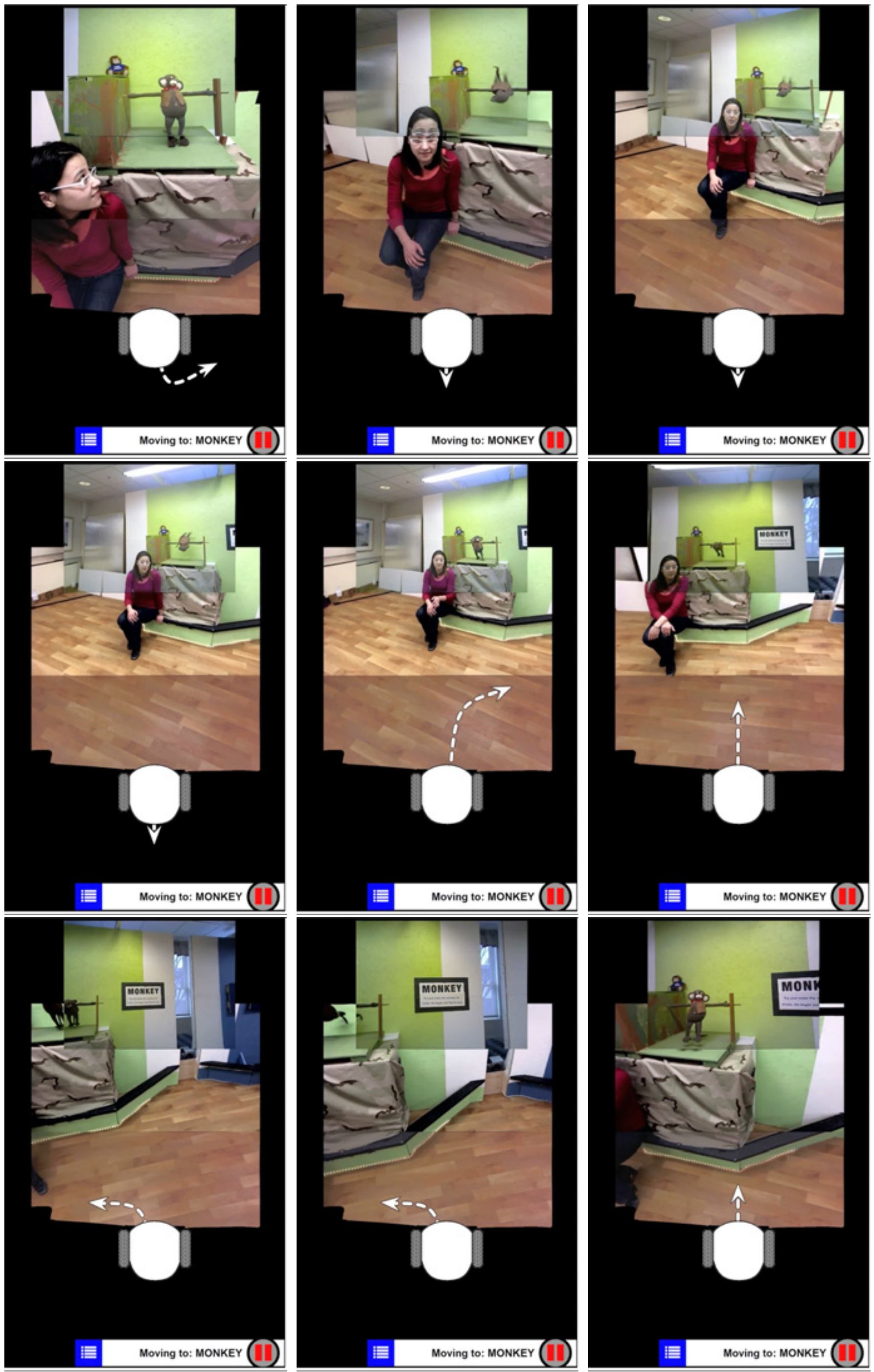


Figure 6-4: When moving from Monkey's *center* hotspot to its *right* hotspot, the robot first backs away to the *right retreat* pose and then moves forward to its alignment point *right align*, and finally arrives at the hotspot *right*. The exhibit remains in the robot's field of view for the duration of this transition.

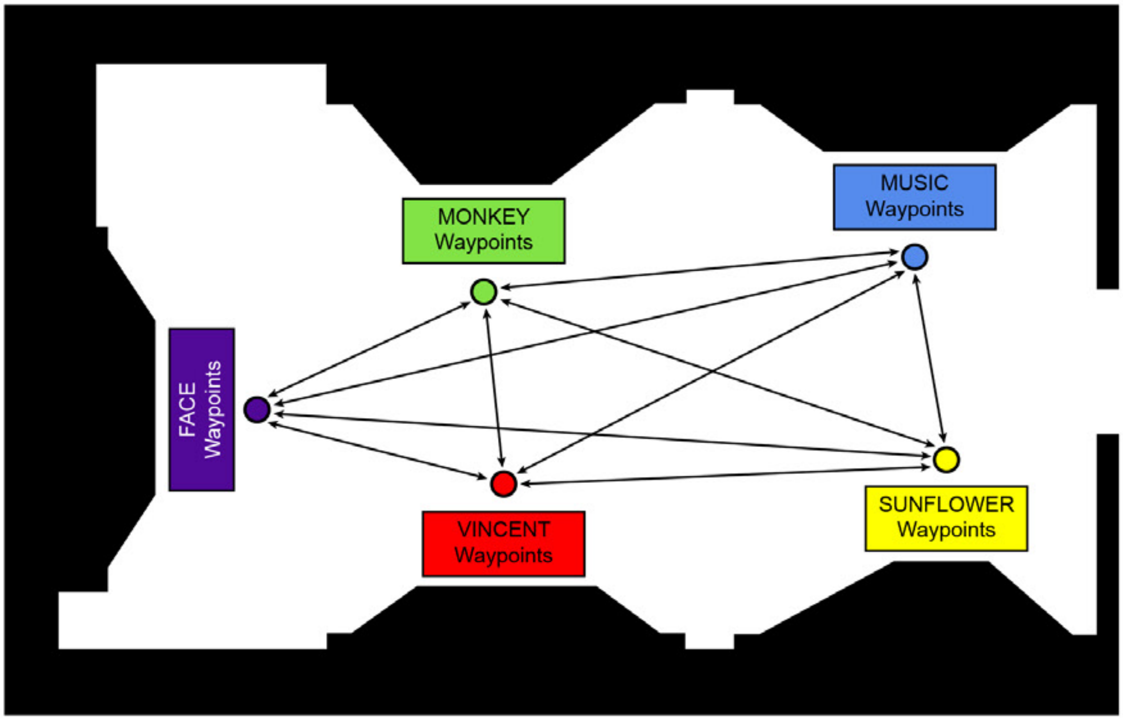


Figure 6-5: Connected graph of the robot's movement between exhibits. A user can choose to move from the robot's current location to any other exhibit, even while in motion.

from `/movement_manager` on `/nav/slow_goal` topic, the local planner used speed multiplier of 0.5, reducing the resulting speed in half. The robot could move 2 steps forward or backward from each hotspot.

Obstacle Avoidance. Robots with sophisticated autonomous behaviors have been less accepted by users (e.g., [Carlson and Demiris, 2010; Lankenau, 2001; Parikh et al., 2005; Viswanathan et al., 2007]), particularly if it modifies a user's direct command given with a joystick. Collision avoidance is a prospective, self-preservation behavior, and is implemented on most autonomous and semi-autonomous robots using a number of techniques (e.g., vector field histogram [Borenstein and Koren, 1991]). There are two issues which attribute to a user's lack of understanding of robot navigation behaviors activating collision avoidance. First, the collision avoidance behavior may be non-deterministic, or an emergent result of the underlying algorithmic implementation and the environmental conditions [Maes, 1994]. Second, Blumberg and Galyean [1997] note that a prospective behavior avoids doing a specified behavior, thus *any other* behavior is technically correct regardless of any expectations the user might have [Desai, 2012; Desai et al., 2013, 2012]. In contrast, a prescriptive behavior only specifies the desired behavior.

The necessary tradeoff for reproducible trajectories was to omit obstacle avoidance. The static configuration of the environment ensured the overall safety of the robot system. We discuss this further in 8.3.1.

Local Planner Implementation. The desired velocity was calculated based on the target waypoint pose and its properties. We used normalized angle difference between robot's yaw and the goal orientation towards the target paired with modified sigmoid functions to calculate both translational and rotational velocities (Figure 6-6, p. 136). The speed curves were tuned so the resulting behavior would be as follows:

- If the current target is mostly ahead, the robot would drive straight toward the target at full speed.
- If the target is to the side, robot would quickly turn and then transition to forward movement.
- If the target is behind, two possibilities exist, depending on the target reversibility: (1) point turn until the target is in front, then use forward driving mode; or (2) drive backwards (Figure 6-7, p. 137).

Figure 6-6 (p. 136) shows an example of velocity calculation for the target waypoint which is mostly to the right, but still in front of the robot. First, angle α toward the target is calculated and mapped to the interval $(-1, 1)$ (middle drawing). This angle is referred to as the *normalized angle*. In the example, the normalized angle $\alpha \approx 0.375$ corresponds to angle of $3^{\pi}/_{8}$ radians. The resulting forward translation speed (top graph) is very small. While resulting rotation speed (bottom graph) is close to the maximum possible rotational speed. For the shown scenario, the generated velocity would cause robot to rotate clockwise, almost in-place. As angle α becomes smaller, the translation speed would gradually increase until it is at its maximum, while the

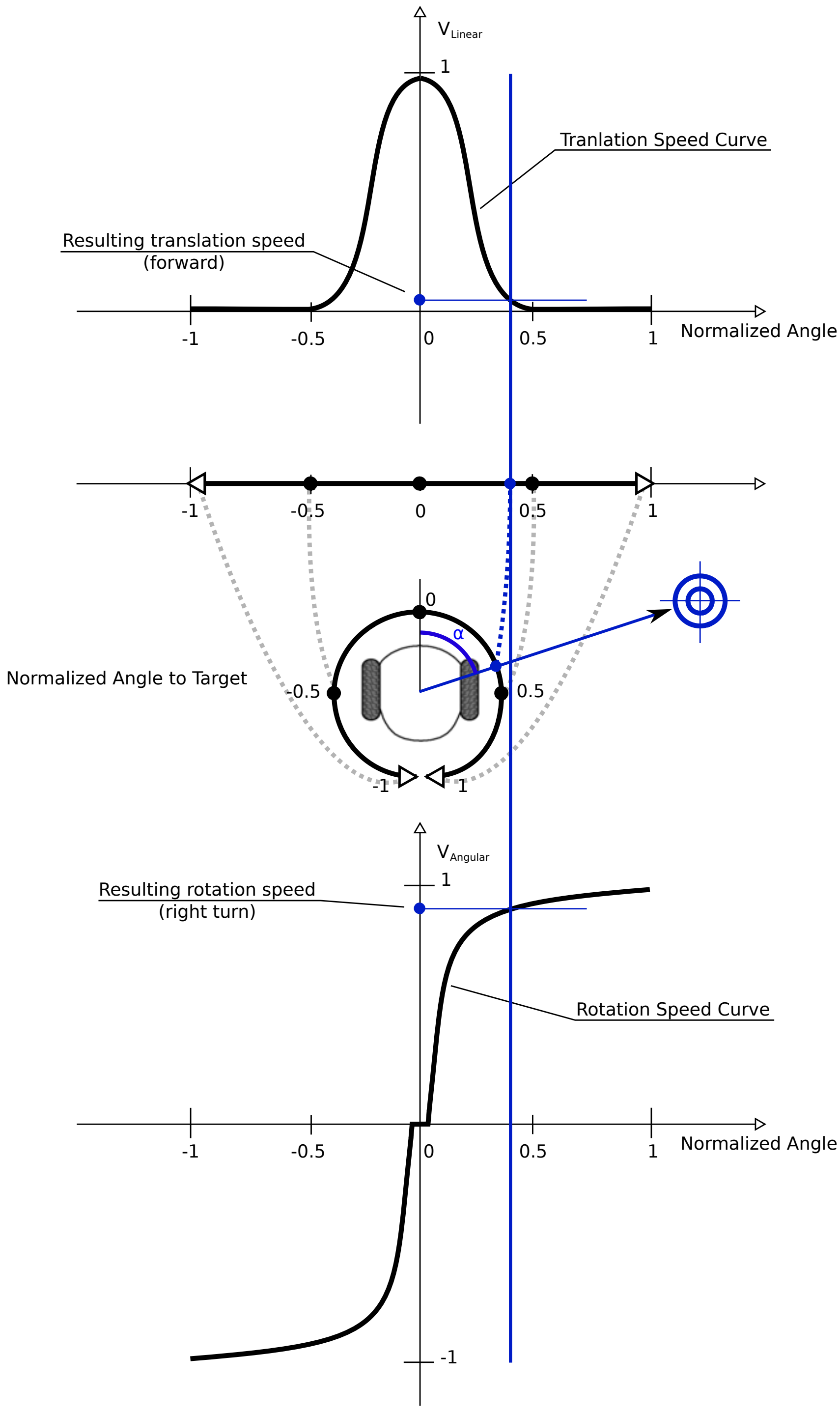


Figure 6-6: Graphical representation of functions used to generate translational **(top)** and rotational **(bottom)** speeds, based on angle to the target waypoint **(middle)**. Note that rotation speed curve is snapped to zero for small angles, helping to avoid oscillations when the robot is close to achieving the target yaw. These graphs are approximations.

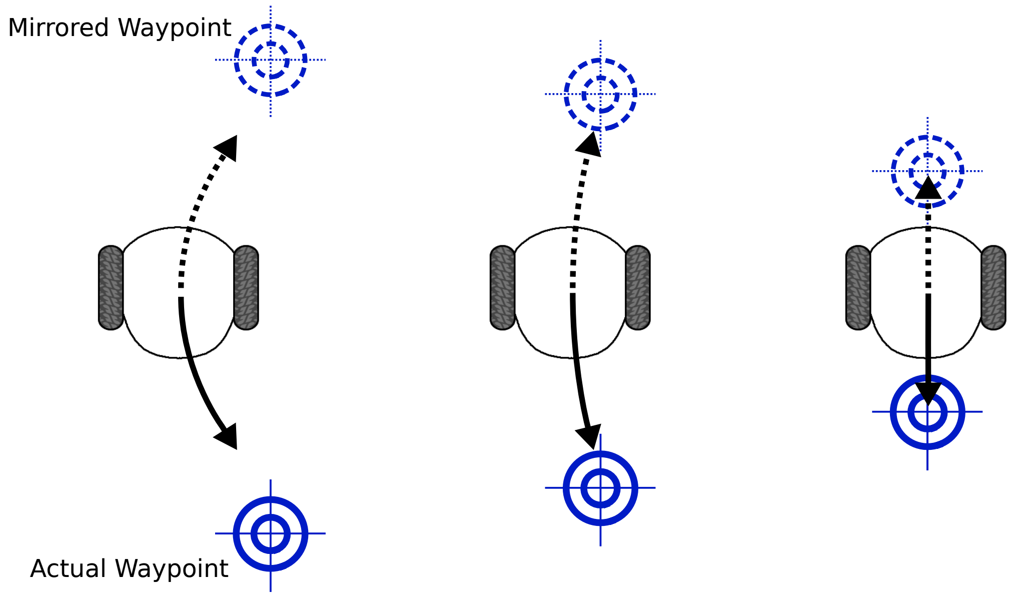


Figure 6-7: Movement sequence for retreat behavior. To drive the robot backward to a reversible waypoint, the local planner used the same algorithm and speed curves (see Figure 6-6, p. 136) as were used for forward-only waypoints. To allow for this behavior, waypoints behind the robot (solid targets) were mirrored to appear in front of the robot (dotted targets). Produced velocity then was reversed to achieve retreat behavior.

rotational speed would decrease. The resulting trajectory would resemble the one shown in Figure 6-8 (middle) (p. 138).

6.3 Updating UI Elements

The robot's movements directly affected elements of the user interface: the exhibit hotspots, the AVI (white dotted movement arrow originating from the robot base icon), and the emergency stop.

6.3.1 Recognition over Recall

The hotspots represented real-world wayposes that the user could select. `/path_manager` instantiates each of the five exhibits. When the robot arrives at an exhibit's *main* entry pose, coordinate frames are created for each of its four hotspots: *left*, *center*,

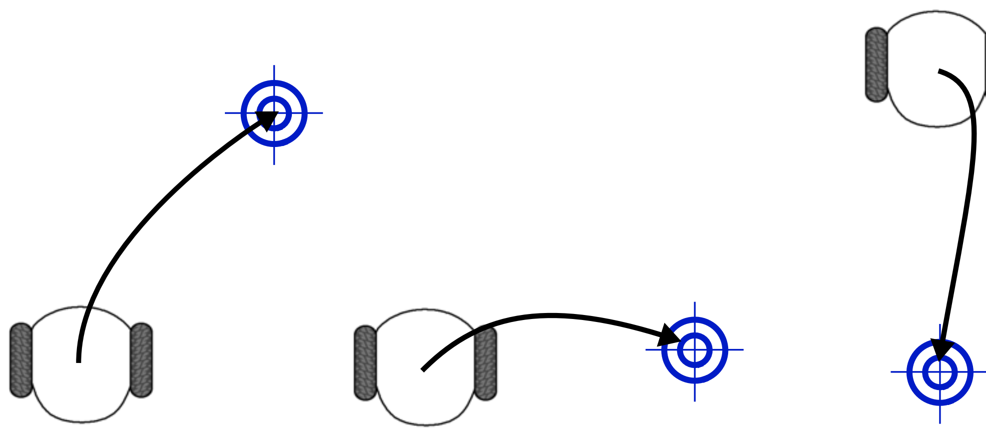


Figure 6-8: Local planner trajectories. **(Left)** and **(Middle)**: Waypoint is in front. **(Right)**: Waypoint is behind and is not reversible (robot first point turns, and then starts moving toward it).

right, and *info*. These coordinate frames are published as timestamped transforms to `/tf` [Foote, 2013]. The transform references the global `/map` coordinate system.

The `WOPR` node from the user interface component (Figure 6-1, p. 128) polls `/tf` at the rate of 10 Hz for new transforms. It looks at the most recent `/tf` messages for the *left*, *center*, *right*, and *info* coordinate frames transformed from the `/map` reference frame to the robot's `/base_link` frame (local coordinates surrounding the robot).

If `/webUI_callbacks` has requested the hotspots, then `/WOPR` converts the hotspots from the physical real-world coordinates (x, y in meters) to be displayed on the user interface (x, y in pixels).

The conversion of the hotspots' locations from meters to pixels requires swapping from Cartesian coordinates, to polar coordinates, and then back Cartesian coordinates. According ROS's REP 103 [Foote, 2010], coordinate frames must be right-handed; that is, $+z$ points upward from the ground plane, $+x$ is forward, and $+y$ is to the left. Thus, coordinate frames are inherently Cartesian. Since the hotspots are to be visualized around the robot base icon, it follows that the robot base icon should become the origin of the coordinate system while the placement of each of the hotspot's visualization is computed. Each visualized hotspot had a radius of 60 pixels. We calculate the distance r and angle φ between the robot (`/base_link`) and a given hotspot by converting x, y to polar coordinates:

$$r = \sqrt{x^2 + y^2} \tag{6.1}$$

$$\varphi = atan2(y, x) \tag{6.2}$$

Additionally, the angle φ is adjusted using the robot's current orientation (yaw). The magnitude m of the vector in pixels from the robot base icon is:

$$m = 2 * (853.02r - 119) \tag{6.3}$$

Equation 6.3 was determined empirically by performing a linear regression of the paths, in the robot's video, of a calibration marker on the floor as the robot rotated. The hotspots were converted back to Cartesian coordinates to be displayed on the screen:

$$x = 471 - (m * \cos(\varphi)) \tag{6.4}$$

$$y = 1142 - (m * \sin(\varphi)) \tag{6.5}$$

where (471, 1141) is the center of the robot base icon. 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 pixels. The visualizations of the hotspots were pushed outside the robot base icon. Overlapping adjacent hotspots were separated by moving both equally from the midpoint of the line between their centers. After `/WOPR` calculated the renderable location of hotspots, the user interface would draw hotspot visualizations at those locations. If the robot were at a hotspot, only the other hotspots would be drawn.

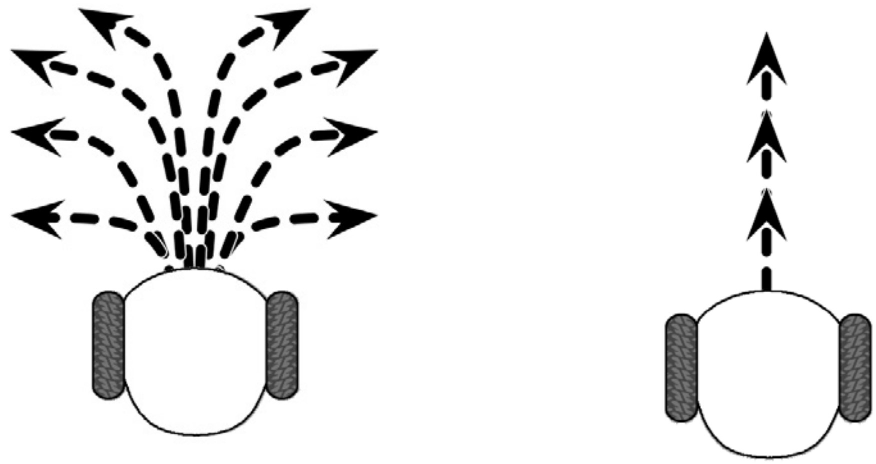


Figure 6-9: Forward Animated Vector Indicators (AVI).

6.3.2 Visibility of System Status

Animated Vector Indicator (AVI)

To animate the AVI originating from the robot base icon on the user interface, `/nav` published an indicator pose to the `/goto/directions_pose` topic. The indicator pose contained Cartesian coordinates relative to the robot's origin; that is, it was transformed into robot's `/base_link` frame. The indicator pose was selected from the remainder of the path currently being traversed as follows. If there was only one waypoint remaining on the path, it was used as the indicator pose. If the first waypoint of the path remainder (active waypoint) was in front of the robot, then the second waypoint (one after the active) was used as the indicator pose. Finally, if the first waypoint of the path remainder was behind of the robot, it was used as the indicator pose. We found that such strategy allowed us to accurately create an appropriate AVI based on the indicator pose.

The indicator pose's x and y components were used to create an AVI. The x value was binned by `/webUI_callbacks` into forward, backward, or stopped, and the y value into left, right, or centered. We utilized short circuit evaluation of these six booleans to change the magnitude and direction of the AVI displayed on the user interface. For the backward AVI, there were three representations (Figure 6-11):

Figure 6-10: Rotation Animated Vector Indicators (AVI).

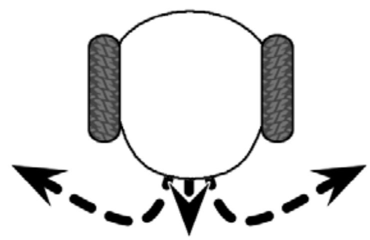


Figure 6-11: Backward Animated Vector Indicators (AVI). AVIs for three representations are overlaid.

one to indicate the robot moving straight back, one to indicate backing up to the right, and another backing up to the left. There were 4 magnitudes for forward translation to the right or to the left (>1.5m, 1.2m, 0.8m, and 0.01m); Figure 6-9 (p. 140). There were 3 magnitudes for straight forward motions (>1.5m, 0.8m, and 0.01m). Finally, /webUI_callbacks monitored /vgo_base/cmd_vel to change the AVI when the robot was turning in place right or left without any forward or backward translation; Figure 6-10 (p. 141).

Emergency Stop

Robot handlers could emergency stop the robot by several methods: from the user interface via a keypress, pulling the physical rip-cord e-stop, or any means of publishing a boolean `true` to the `/estop_software` topic. Regardless, the resulting behavior was that the robot stopped, and `/vgo_base/velocity_manager` ceased publishing velocity commands to `/vgo_base/cmd_vel`. The interface showed a yellow triangle with red exclamation point in the lower right corner in place of the iButton (Figure 5-12, p. 120).

6.4 Indication of Robot Movement to Interactant

Margo's Hawaiian shirt contained tri-color LED flowers, as described in Chapter 3. When the robot was idle (i.e., waiting for a movement command), the flowers were blue. While the robot was moving, the flowers were green; if the robot were paused during its movement, the flowers shone red.

6.5 Summary

It has been largely assumed that the user is always controlling the telepresence robot's movements, regardless of the robot. Thus, the general perception of how to use a telepresence robot has been to provide low level forward, back, left, and right (FBLR) commands. In our early work, we found that able-bodied novice users had difficulty driving telepresence robots straight down a corridor [Desai et al., 2011; Tsui et al., 2011a]. The latency often caused the robot to turn more than the user intended and thus zig zag down the hallway. Teleoperating a robot at this level can be a cognitively taxing task, particularly over long periods of time. We found that continuous robot movement via teleoperation was an issue with our target population's mental model of the robot due to the latency [Tsui et al., 2011b]. Therefore, we changed our drive mode to a discrete style; that is, all robot movement was autonomous.

Autonomous navigation behaviors can free the user from the details of robot navigation, making the driving task easier; consequently, the user can focus on exploring the remote environment and/or engaging in remote social interaction. From Study 2 (scavenger hunt, Chapter 4), we found that users wanted to give directives at the low-level (i.e., forward, back, 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). Thus, high-level control was provided by allowing the user to change exhibits. Mid-level control was provided in the form of local, within-exhibit hotspots. Finally, low-level control was possible through the forward/reverse step buttons while at any of an exhibit's hotspots.

Chapter 7

Case Study: Exploring an Art Gallery

We believe that the 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, described in Chapters 5 and 6. 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 [Nielsen, 1994a]. 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 (Study 3) 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 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.

As noted in Chapter 2, Beer and Takayama [2011] found that seniors wanted to use the Texai robot to attend concerts or sporting events, and visit museums or theatres. 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 found that 51.8% of respondents said that they visited the museum to spend time together as a group or family [Lindgren-Streicher and Reich, 2011]. The content of the art gallery was also inherently engaging; all of the Artbotics exhibits were kinetic and animated differently when approached from each angle.

7.1 Study 3 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 (Figure 5-1, p. 95). 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.

7.1.1 Setup

A Windows 7 Dell XPS laptop (Windows 7 Pro SP1 64bit, Intel Core i7-2820QM @2.3 GHz, 4 GB RAM) ran the VGo Desktop App (version 2.1.0.23574) and our interface in the Chrome browser (version 34.0.1847.116m). The XPS laptop was connected to a 22" 3M multitouch 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.

7.1.2 Procedure

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

In the first session, the experimenters obtained the participant's consent and administered an interview (demographic information and prior experiences) and a training exercise. 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 than one viewpoint, and (3) moving from one exhibit to another; detailed later on p. 146. 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; detailed later on p. 148. 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 (Training 2). The participant visited the gallery using telepresence robot a second time and explored for up to 45 minutes. There were 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 participant was instructed to start at a specific exhibit, which was his or her least favorite exhibit as noted from the first visitation; detailed later on p. 148. 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.

In the post-session interview, we asked the participant to describe his or her

Figure 7-1: During training, only two exhibits were available to the participant. All exhibit and exhibit menu buttons were set to only show the color of the exhibit, not the icons associated with them. **(Left)** shows the exhibit menu buttons in the menu. **(Right)** shows the exhibit buttons.

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

Training

While a telepresence robot system is being learned, a user's telepresence may be hindered by the need to think about the robot's movement. We provided a separate training exercise in which participants were taught about the robot and how to control it in the remote environment during each session. All gallery exhibits and their informational plaques were covered. The participant practiced moving the robot in the red and purple exhibit areas. Only the background color was shown in exhibit or exhibit menu buttons; all exhibits in the gallery were covered at this point. Figure 7-1 shows the interface the participants used for training.

An experimenter read the following script and prompted the participant to click on elements of the user interface:

> You will be operating a robot that we've placed in another room. The size of the robot is 14 inches wide and it is shaped like a circle. *[Show cut out]* It is like a power wheelchair and has 2 motorized wheels and 2 casters for balance. The robot is 4 feet tall. It has video cameras, microphones, and

speakers so you can see and hear what's going on on the other side. There is a screen that will show your Face on the robot.

This is the user interface. The robot knows how to get to key places in the training area we have set up for you to practice. The practice area has 2 exhibits: RED and PURPLE. We will begin with PURPLE and show you how to move to an exhibit.

There are colored buttons around the base of the robot icon. You will push and release the button and the robot will begin moving. *[Prompt for action]* When the robot is moving, dotted arrows will appear. These show where the robot is moving.

[Continue reading while the robot is moving] You can pause the robot by pressing the pause icon on the lower right of the screen. *[Prompt for action]* While the robot is paused, the lower right icon changes to a play icon. You can make the robot resume moving by pressing the play icon. *[Prompt for action]*

When the robot arrives at an exhibit, hotspot circles will appear. The robot can move to these hotspots to get a different view. *[Prompt for action]*

[Continue reading while the robot moves between hotspots] The robot can't move sideways, so it will first back up and then move forward to get to the hotspot. This is similar to how a car moves.

When the robot arrives at a hotspot, triangles will appear over the icon of the robot's base. You can press and release them to make the robot move a little bit forward or backward. *[Prompt for action]*

There is another way to go to an exhibit: using the menu. This is the dark blue button with white horizontal lines. *[Prompt for action]* Notice the red dotted line. This shows you which exhibit you are currently at. You can press the other button, which is the RED button, and the robot will move there. *[Prompt for action]*

Table 7.1: Time spent on training (seconds). Experimenter time is the time it took to go through the training script before the 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

> The robot is now at the RED exhibit. Notice that the red exhibit button also has a dotted line around it. You can press and release it and information about the exhibit will be shown. *[Prompt for action]* Press and release "DONE" to close the information box. *[Prompt for action]* You can also access the exhibit's information by pressing the iButton in the lower right corner of the screen. *[Prompt for action]*

The experimenter solicited questions about the interface and reviewed as necessary. Participants were given 10 minutes to freely practice controlling the robot. Upon completion of the training, participants then used the robot to visit the gallery.

There was one training session before each gallery visit session. During the second training session, two of the participants required little to no retraining; while the other two only required marginally more time to practice (Table 7.1). P3 required more than 10 minutes during his first training session.

Task

The task descriptions for Sessions 1 and 2 were similar. The exhibit menu was opened to start, and the experimenters instructed the participants as follows:

> We've set the robot up in the art gallery and there are <*nubmer_of_exhibits*> exhibits. You will have <*number_of_minutes*> minutes to visit the exhibits. You will start at the <*name_of_exhibit*> exhibit (<*color_of_exhibit*>

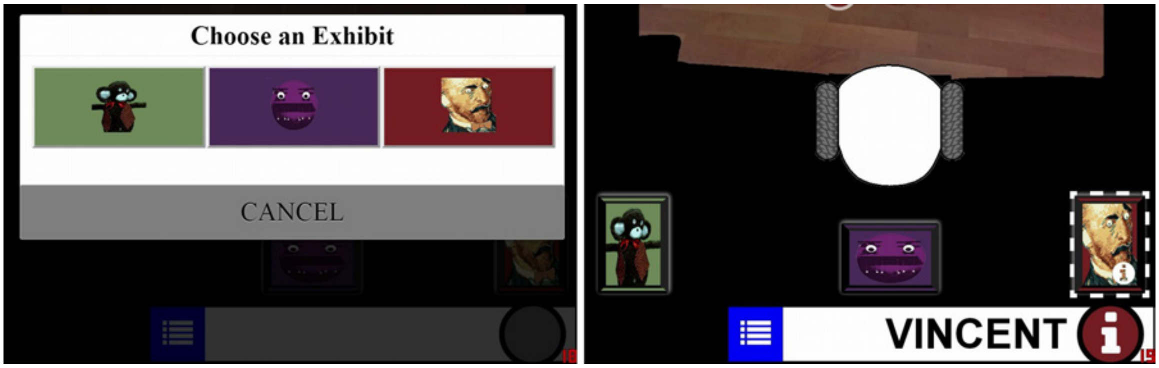


Figure 7-2: During the first session, only three exhibits were available to the participant through exhibit and exhibit menu buttons. **(Left)** shows the exhibit menu buttons in the menu. **(Right)** shows the exhibit buttons.

color). Pay attention to each exhibit's color, motion, sound, etc. When you are done, please let me know.

For Session 1, there were 3 exhibits; the other two exhibits were covered. Figure 7-2 shows the interface the participants used for Session 1. The participants were given 20 minutes to explore. The starting conditions were randomized. P1 and P4 were instructed to begin at the Vincent exhibit (red background), P2 at the purple Face, and P3 at the Monkey (green background).

For Session 2, there were 5 exhibits and the participants were given 45 minutes. Figure 7-3 shows the interface the participants used for Session 2. The first exhibit as instructed by the experimenters was each participant's least favorite exhibit as recorded in Session 1. P1 and P3 were instructed to begin at the green Monkey, and P2 and P4 the purple Face.

Confederate

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 experimenters introduced the confederate to P3 at the end of his first session. The confederate was not given prior training on the use of the user interface the participant was using to control the robot. Aside from being instructed to prompt conversation from the participant about the exhibits, the

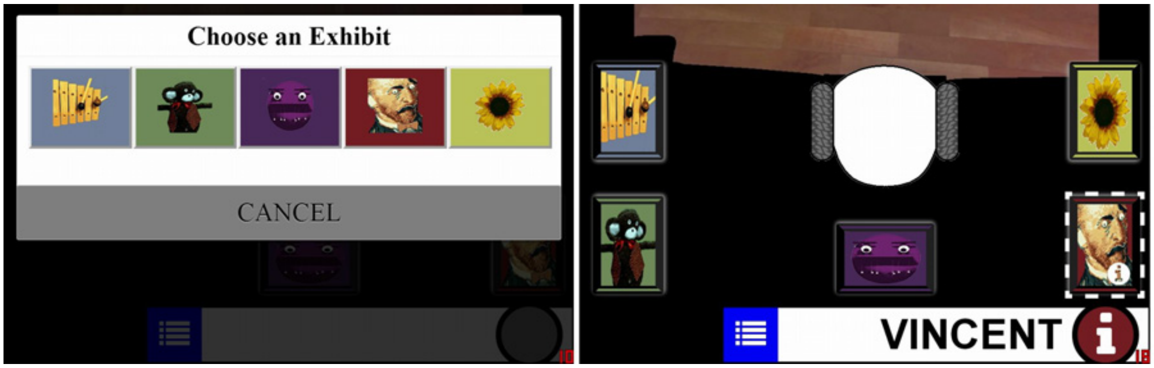


Figure 7-3: During the second session, all exhibits were available to the participant through exhibit and exhibit menu buttons. **(Left)** shows the exhibit menu buttons in the menu. **(Right)** shows the exhibit buttons.

confederate was not given a script, in order to elicit more natural conversation.

The confederate's role was to engage the participant in conversation about the exhibits in the gallery during his or her second visit. As such, her interactions with the participant were semi-structured for the first two exhibits. The confederate was instructed to initially stand at the participant's least favorite exhibit when the participant began to use the robot. The exhibit and its lights were powered on so that the confederate could move back and forth between exhibit's sensors, as if looking at and interacting with the exhibit. When the participant arrived at the first exhibit, the confederate initiated conversation by greeting the participant and engaging in a moment of unscripted small talk. The confederate redirected the conversation to express how much she liked this exhibit, assert that it was her favorite of all the exhibits in the gallery, and embellish with why she liked it so much. To elicit conversation from the participant, the confederate asked the participant what he or she thought of the first exhibit; she prompted the participant about what he or she liked and disliked about the exhibit.

Next, the confederate transitioned the conversation to the participant's favorite exhibit and suggested, "Let's go there now. Let's go see <*name_ of_ exhibit*>." On the way to the next exhibit, the confederate asked the participant why this exhibit was his or her favorite. Then, the confederate spent some time at this exhibit with the participant and prompted the participant to move around the exhibit to demonstrate what features he or she liked.

To transition to the third exhibit, the confederate stated, “There are two new exhibits in the gallery. One exhibit has dancing Sunflowers, and the other is a xylophone that plays Music. Which exhibit are you going to see next?” The confederate then explicitly asked permission to join the participant at the next exhibit (i.e., “May I join you?”) and proceeded to that exhibit if invited. From the third exhibit onward, the confederate was to continue visiting exhibits and accompanying the participant so long as he or she was engaged with the confederate. If the participant was not engaged, the confederate was instructed to remain in the gallery, continue moving between other exhibits, and re-engage if spoken to or called over by the participant.

Finally, 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). The confederate told the participant that she had to meet another client and made social pleasantries when saying goodbye (i.e., “I’m so sorry I have to leave early! Thank you for going through the gallery with me today. I really had fun and I hope you did too. Enjoy the rest of the exhibits.”). The participant was then free to continue exploring the gallery for the remainder of the 45 minute session.

7.1.3 Data Collection

In addition to semi-structured interviews (see Appendix E), 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 [2012]. 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 7.2). Cohen’s kappa for inter-rater reliability between two raters was $\kappa = 0.73$ excluding chance; both raters fully reviewed the categorization of the utterances.

7.1.4 Recruitment and Participants

Inclusion 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. Participants must have been able to fluently speak English but were not required to be a native speaker.

Exclusion criteria

People with blindness, severe cognitive challenges, low arousal levels, or other conditions may not benefit from 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).

Table 7.2: Coding categories for conversation between confederate and participant during participant's second gallery visit ($\kappa = 0.73$)

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

Two participants had spinal cord injury; P2's injury occurred one year prior, and P4's less than one year. 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. Both 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. Both 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 [Suitable Technologies, Inc., 2013]. 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 Study 2 (scavenger hunt study, Chapter 4) and therefore had prior knowledge about our robot system.

7.2 Visiting the Gallery

7.2.1 Session 1

All four participants visited the three exhibits (Vincent, Face, and Monkey) during their first session; P1, P2, and P3 explored the gallery for the full 20 minutes allocated. P2 and P4 interacted with the exhibits more thoroughly than P1 and P3; both P2

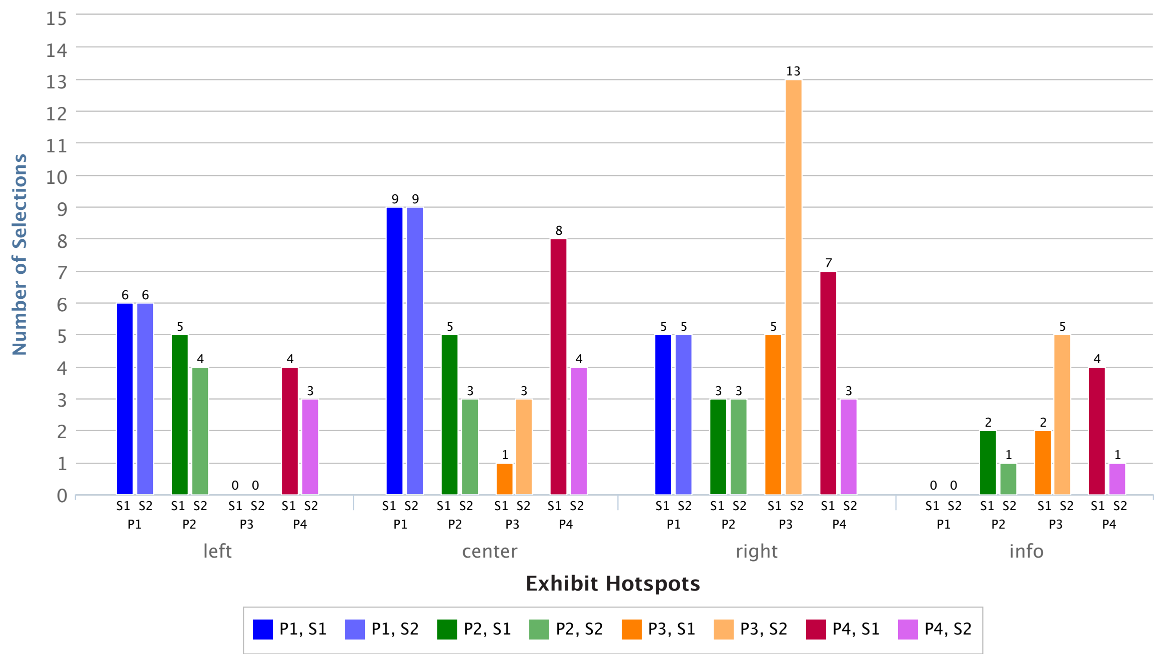


Figure 7-4: Frequency count of the participants selecting hotspots during their two in-robot visitations.

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 ($t = 20\text{m } 29\text{s}$), 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; 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) ($t_1 = 4\text{m}$). Then, P1 interacted with the Monkey from three hotspots ($t_2 = 2\text{m } 35\text{s}$, $t_4 = 3\text{m } 8\text{s}$); 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 ($t_3 = 5\text{m } 36\text{s}$, $t_5 = 2\text{m } 11\text{s}$) 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 Figure 7-6, P2 revisited each of the three exhibits after her first pass during her first in-robot visitation to the gallery. She visited 11 of the 12 hotspots corresponding to the three exhibits; P2 bypassed the information sign on the Face exhibit. P2 noted in her interview that

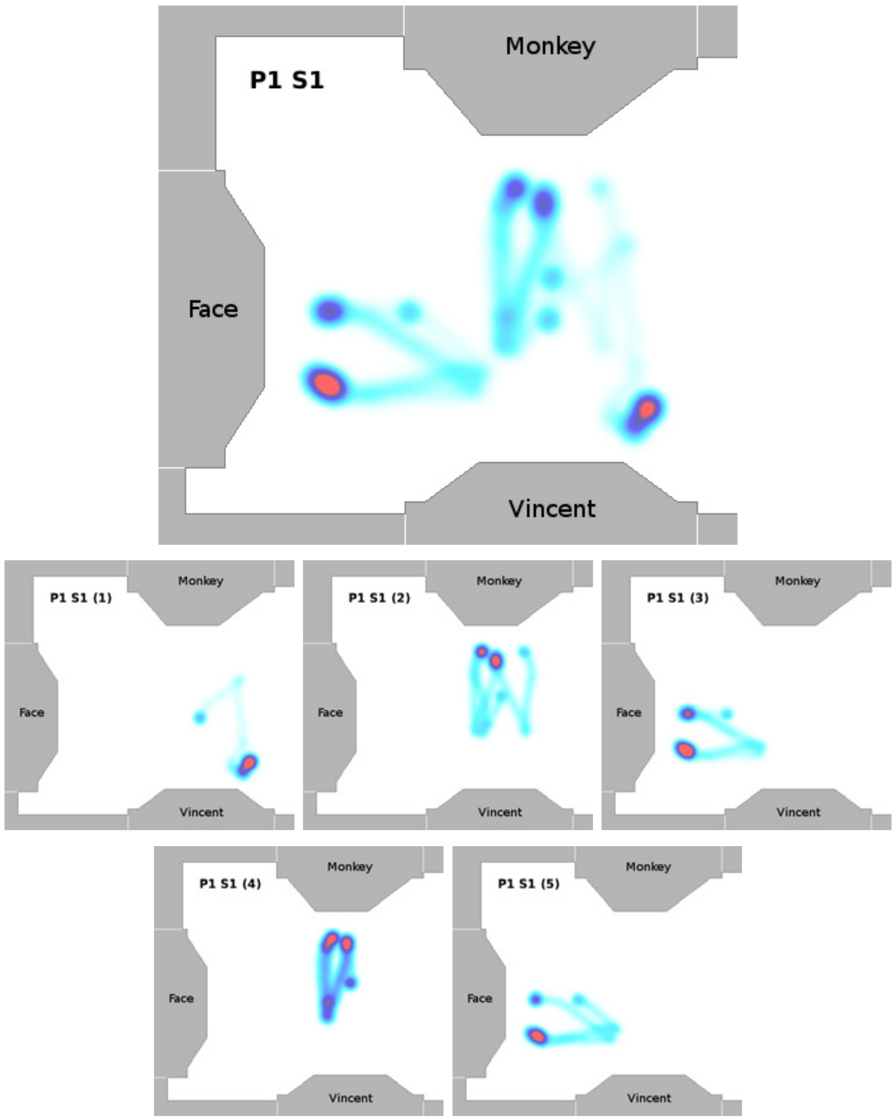


Figure 7-5: Heatmap of P1's first in-robot gallery visitation. The robot's movements within each exhibit area are shown below the large composite image of the robot's movements in all exhibit areas.

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 (t_1 = 2m 19s) prior to Vincent (t_2 = 1m 51s); P2 visited the Monkey exhibit (t_3 = 3m 2s) before making the same loop a second time (t_4 = 2m 11.5s, t_5 = 2m 17s, t_6 = 1m 59s).

P3 explored the gallery for the full amount of time (t = 20m 45s), and spent 12.6% of his time (2m 37s) moving between exhibits. He visited each once and returned to the first exhibit (Monkey; t_1 = 1m 27s, t_2 = 3m 15s). P3 spent the most time at the Face exhibit (t_3 = 10m) and viewed it from two hotspots. He viewed the Monkey and Vincent (t_2 = 3m 26s) primarily from a single hotspot.

Like P2, P4 also revisited each of the exhibits a second time during his first session (t = 18m 8s) and spent 22.7% of his time (4m 7s) moving between exhibits. P4 noted that when visiting his third exhibit (Monkey; t_3 = 4m 12s), 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. Like P2, he revisited the exhibits in the same order: Vincent (t_1 = 2m 32s, t_4 = 1m 45s), Face (t_2 = 2m 8s, $t_5 = 2mm$ 5s), and Monkey (t_3 = 4m 12s, t_6 = 1m 20s).

7.2.2 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. 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 (n_{P3}=332), as shown in Figure 7-9. The fewest utterances occurred between P4 and the confederate (n_{P4}=98), and P4 was the quickest 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 confed-

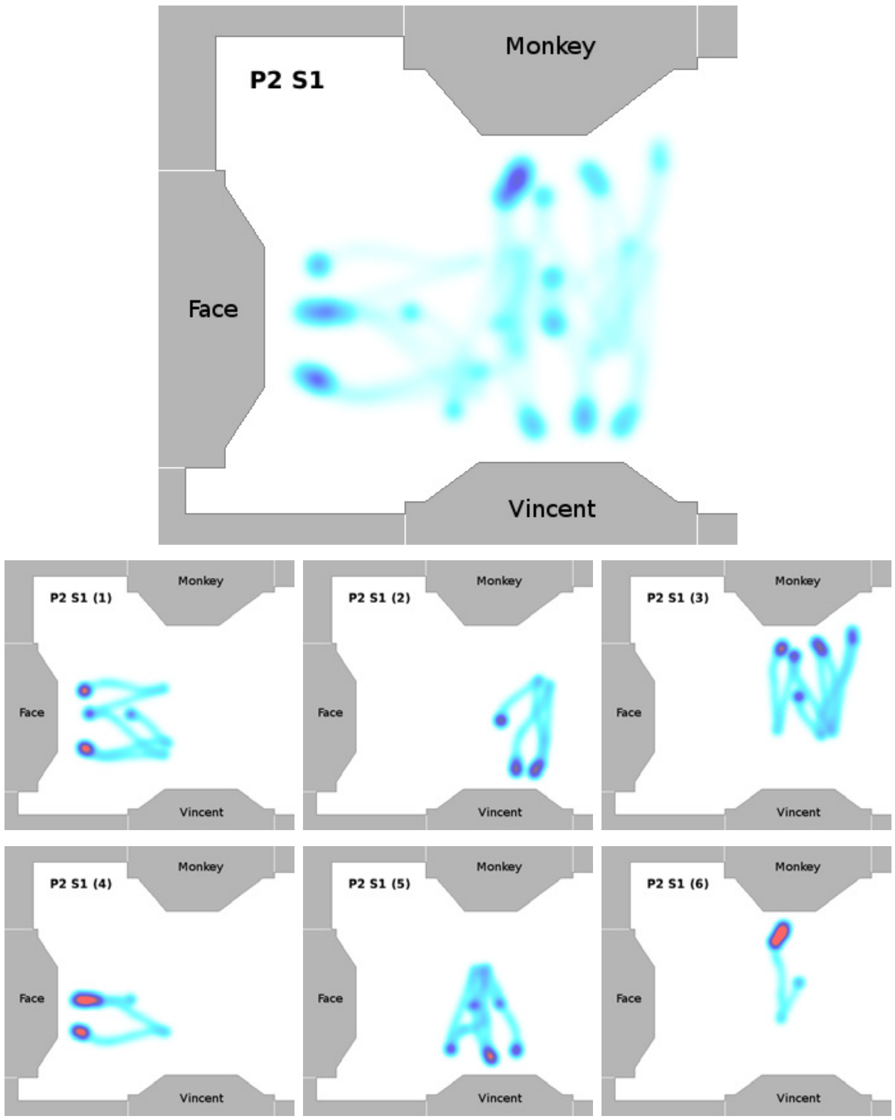


Figure 7-6: Heatmap of P2's first in-robot gallery visitation. The robot's movements within each exhibit area are shown below the large composite image of the robot's movements in all exhibit areas.

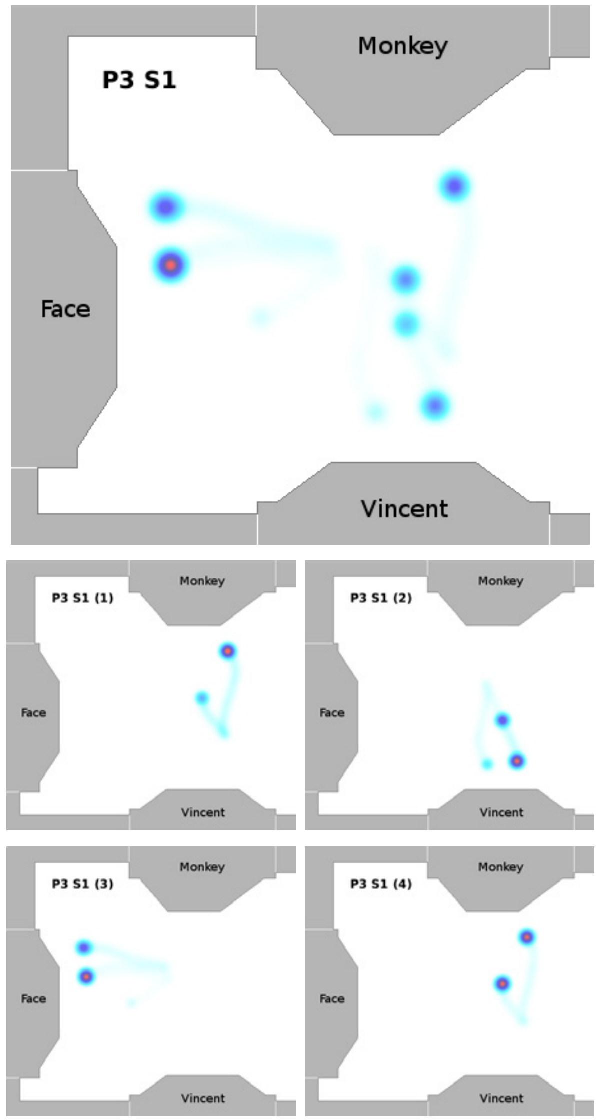


Figure 7-7: Heatmap of P3's first in-robot gallery visitation. The robot's movements within each exhibit area are shown below the large composite image of the robot's movements in all exhibit areas.

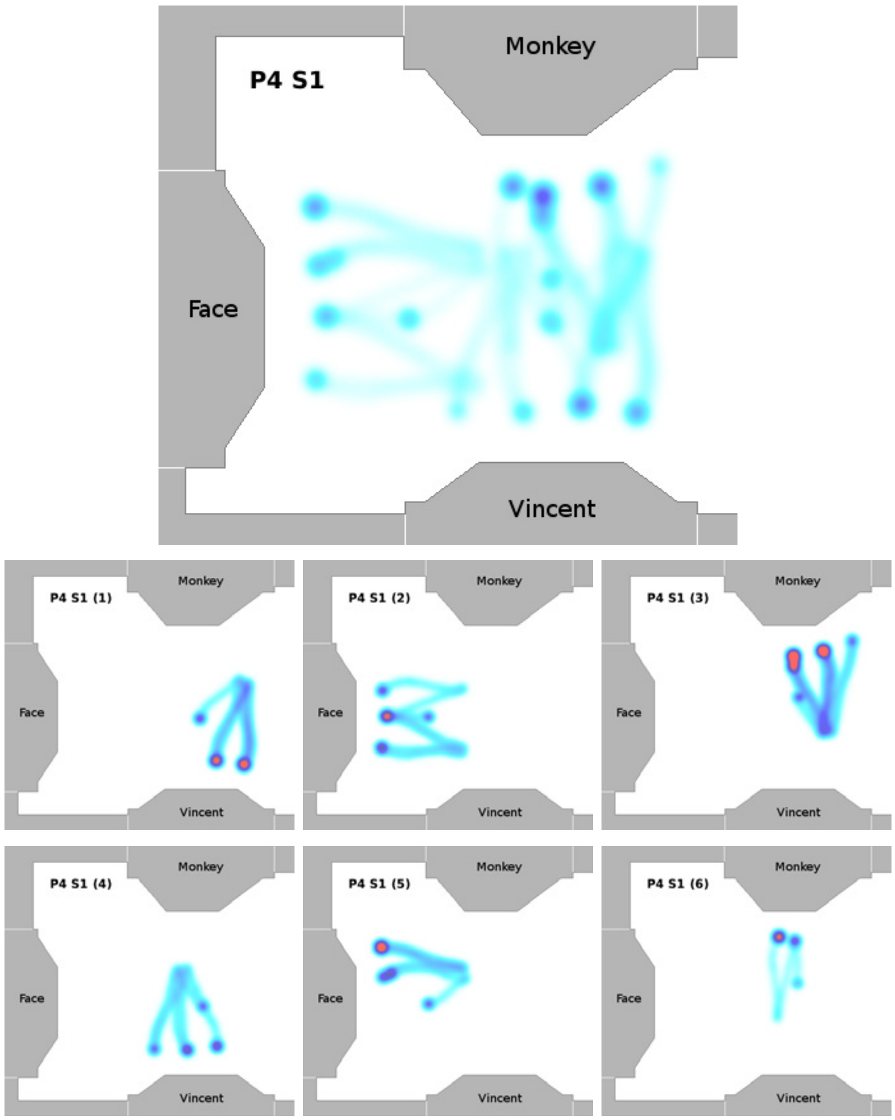


Figure 7-8: Heatmap of P4's first in-robot gallery visitation. The robot's movements within each exhibit area are shown below the large composite image of the robot's movements in all exhibit areas.

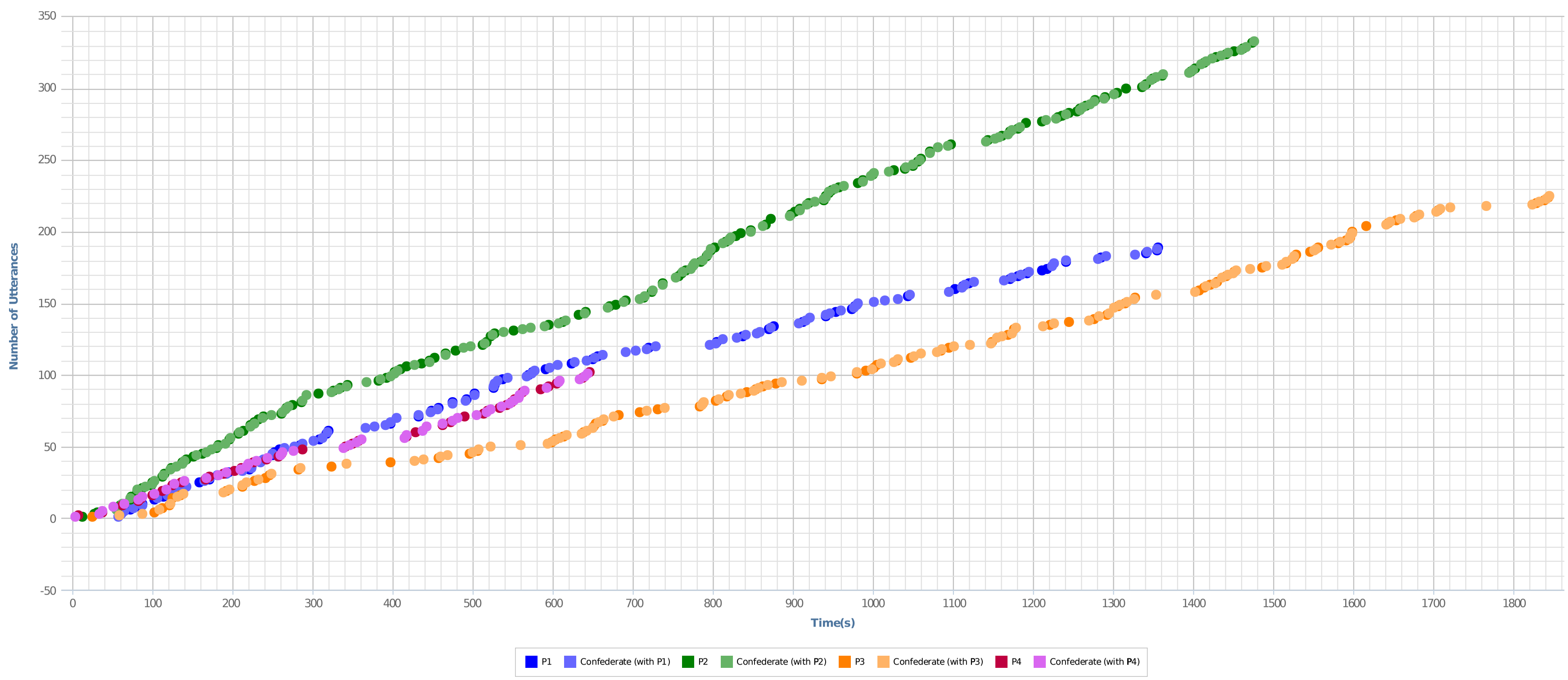


Figure 7-9: Timeline of the participants' overall conversations with the confederate in the second visit to the gallery.

erate at the Monkey exhibit and engaged in conversation for 2m 13s; during this time, P1 did not move the robot. They next visited Vincent (t_2 = 5m 24s) and continued the conversation; P1 moved to the left side of the exhibit. P1 and the confederate went to a new exhibit (Music, t_3 = 5m 43s) 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 (t_2 = 48s) before excusing herself. P1 completed the full session length, revisiting the Music (t_5 = 6m 40s, t_9 = 3m 44s), Monkey (t_7 = 1m 25s), and Vincent (t_8 = 4m 2s). He also explored the Sunflower exhibit (t_6 = 3m 34s) from the left, center, and right exhibit hotspots; P1 visited 6 of the 8 new exhibit hotspots. Over the course of visiting the nine exhibits, P4 spent 7 minutes (15.6%) moving from one exhibit to another.

P2 and the confederate engaged in the most talkative visitation, and P2 immediately concluded her second visitation upon the confederate's departure. Together, they visited each exhibit once (t = 23m 26s). They began at the Face exhibit (t_1 = 1m 43s); 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 (t_2 = 2m 27s) 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 (t_3 = 5m 57s). 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 (t_4 = 5m 59s) before finishing their visitation at the Monkey exhibit (t_5 = 1m 58s). 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.

P3 and the confederate began their visit together at the Monkey exhibit (t_1 = 7m 7s). P3 moved the robot back and forth between the exhibit's right and info hotspot. They then went to the Face exhibit (t_2 = 3m 42s) 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 (t_3 = 5m 45s) and then

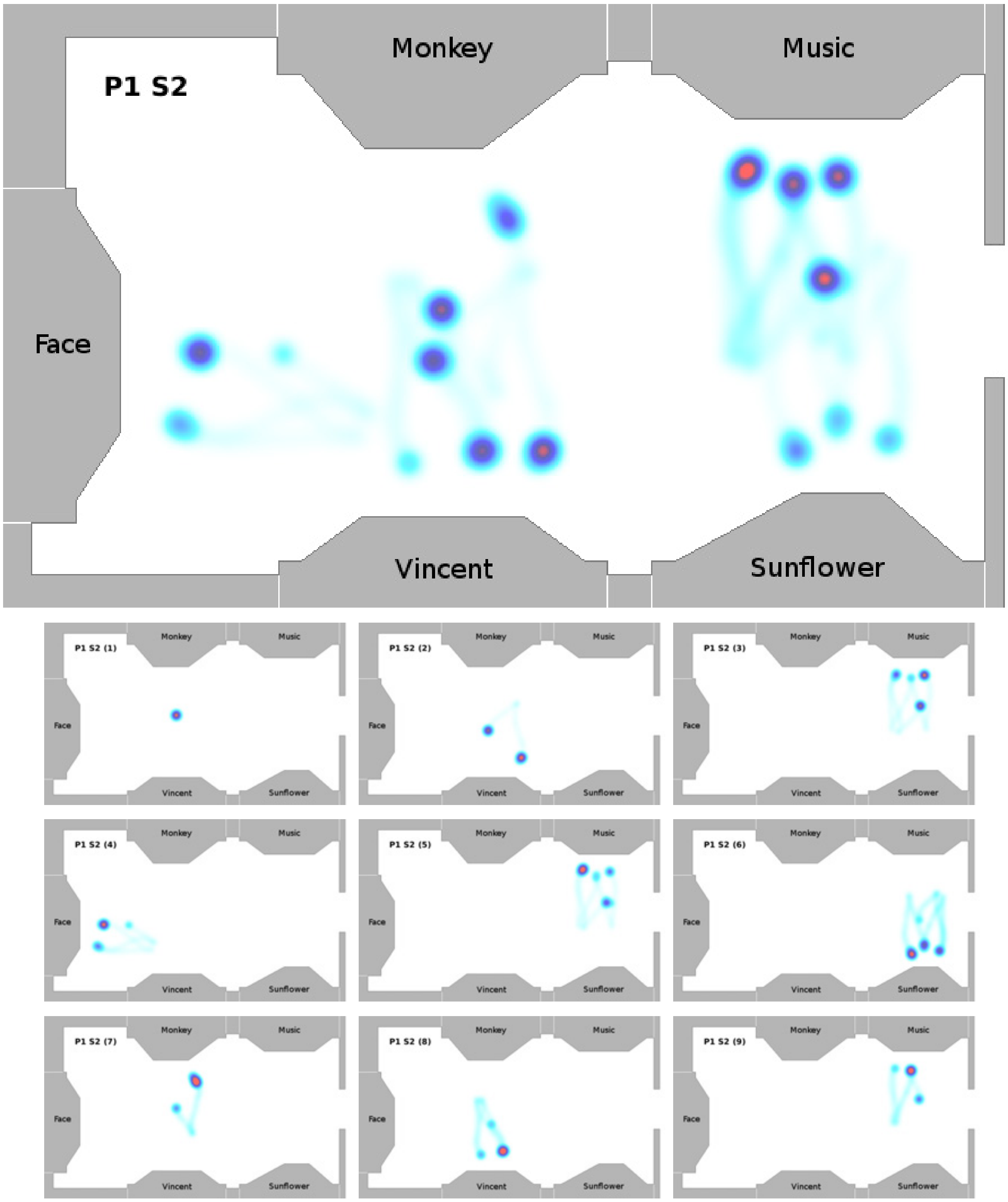


Figure 7-10: Heatmap of P1's second in-robot gallery visitation. The robot's movements within each exhibit area are shown below the larger composite image of the robot's movements in all exhibit areas.

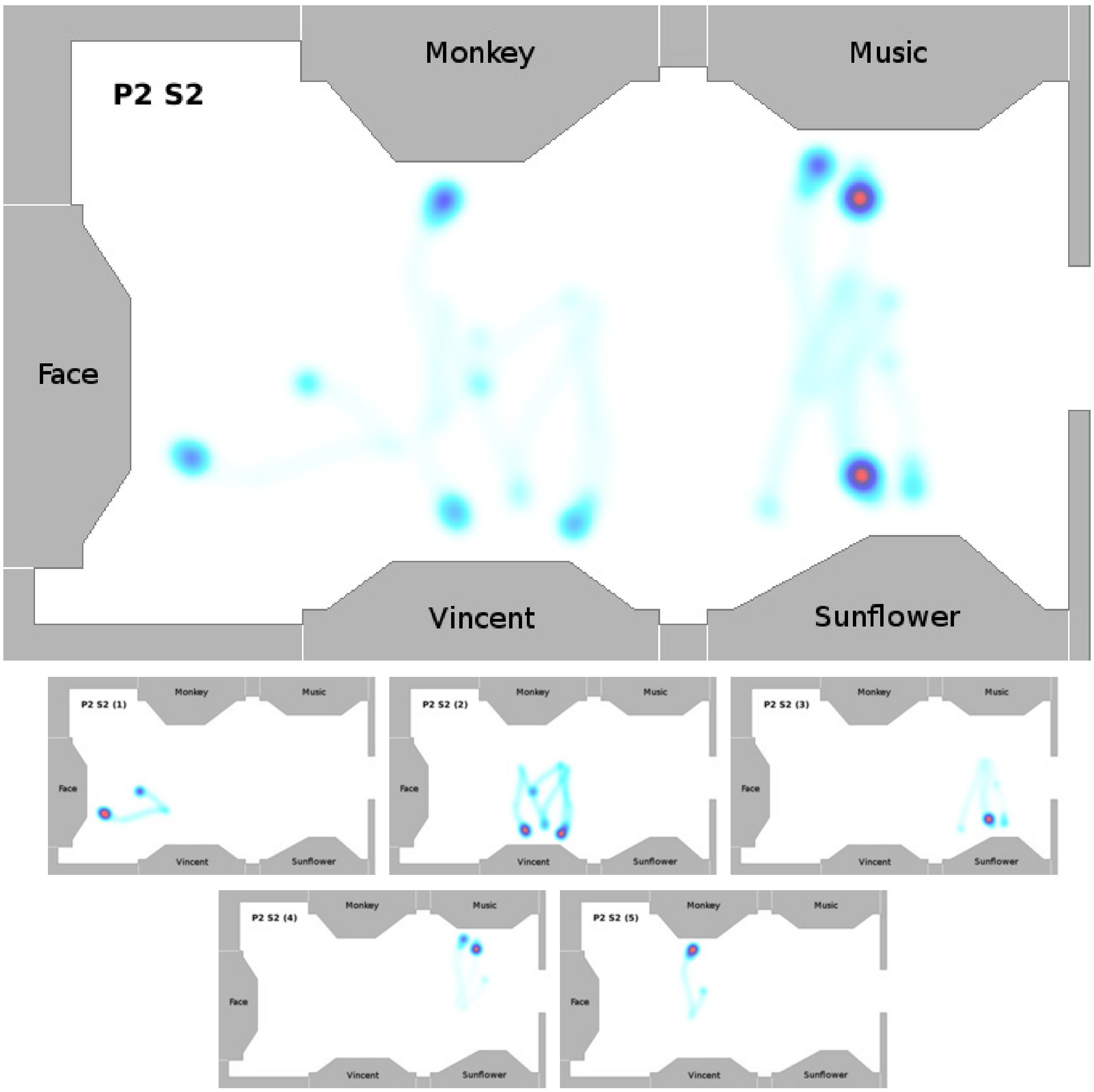


Figure 7-11: Heatmap of P2's second in-robot gallery visitation. The robot's movements within each exhibit area are shown below the larger composite image of the robot's movements in all exhibit areas.

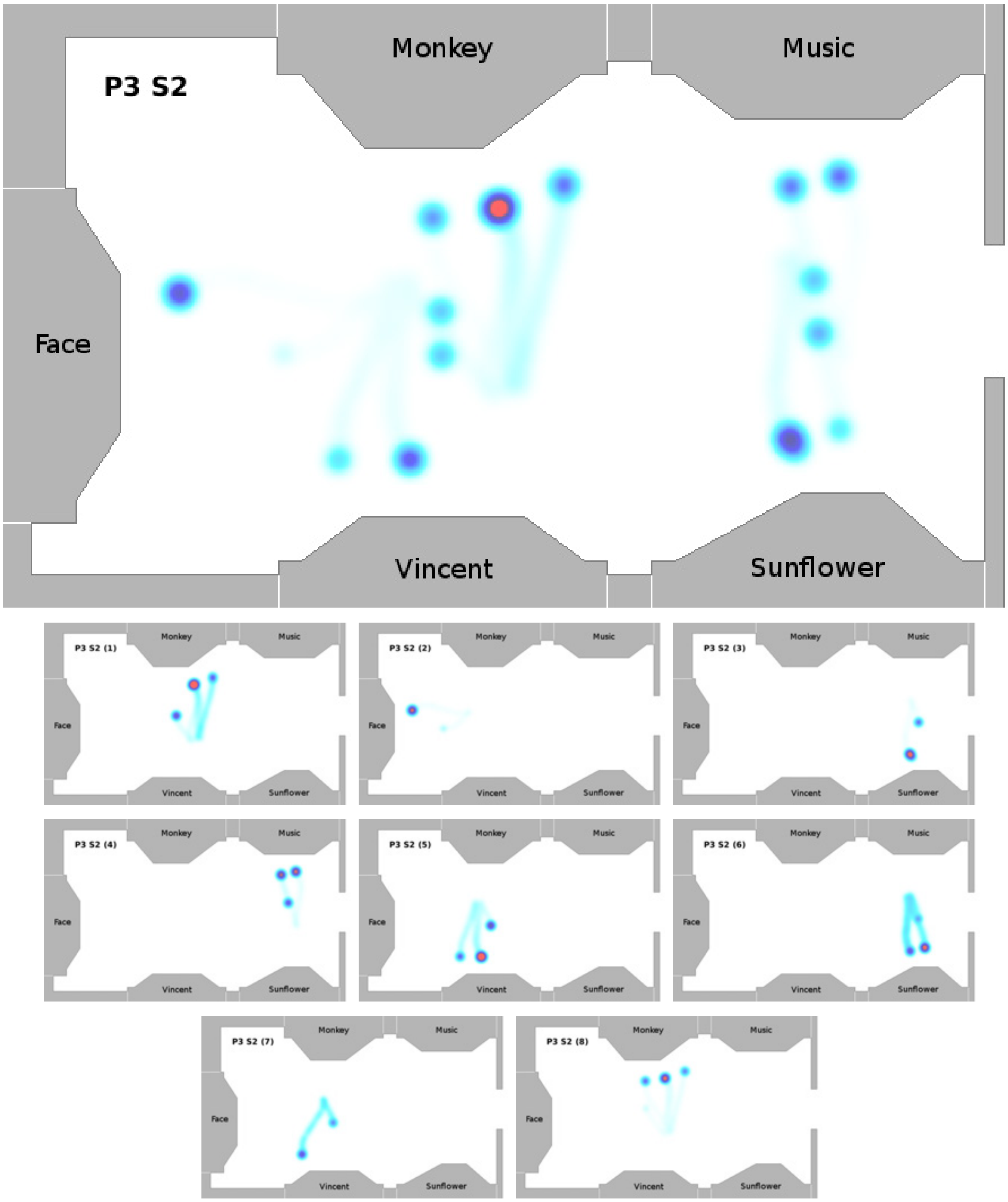


Figure 7-12: Heatmap of P3's second in-robot gallery visitation. The robot's movements within each exhibit area are shown below the larger composite image of the robot's movements in all exhibit areas.

to Music ($t_4 = 5$ 21s). 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 ($t_5 = 5$m 2s) before departing; P3 moved between the info and right hotspots. P3 revisited the Sunflower exhibit ($t_6 = 2$m 5s) viewing it from the right and center hotspots. He also revisited Vincent ($t_7 = 35$s) briefly before turning around to interact with the Monkey exhibit ($t_8 = 8$m 58s) for the remainder of his second session. P3 visited 4 of the 8 hotspots around the two new exhibits; 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, $t_1 = 1$m 36s). P4 also led the movement to each of the next exhibits. P4 and the confederate briefly visited Vincent ($t_2 = 40$s) before moving over to the new Music exhibit ($t_3 = 1$m 20s). 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 ($t_4 = 2$m 35s), which P4 viewed from its left, center, and right hotspots. The confederate excused herself after accompanying P4 to the Monkey exhibit ($t_5 = 46$s), as she had visited all exhibits with the participant. P4 returned briefly to the Face exhibit ($t_6 = 53$s) and then the Music exhibit ($t_7 = 1m$ 55s). P4 moved to the Music exhibit's left hotspot to interact with the xylophone before ending his short visit ($t = 14$m 43s).

7.3 Results and Discussion

7.3.1 Interface Ease of Use

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.

Due to his visual impairment on his left eye, P3 had trouble targeting the hotspots

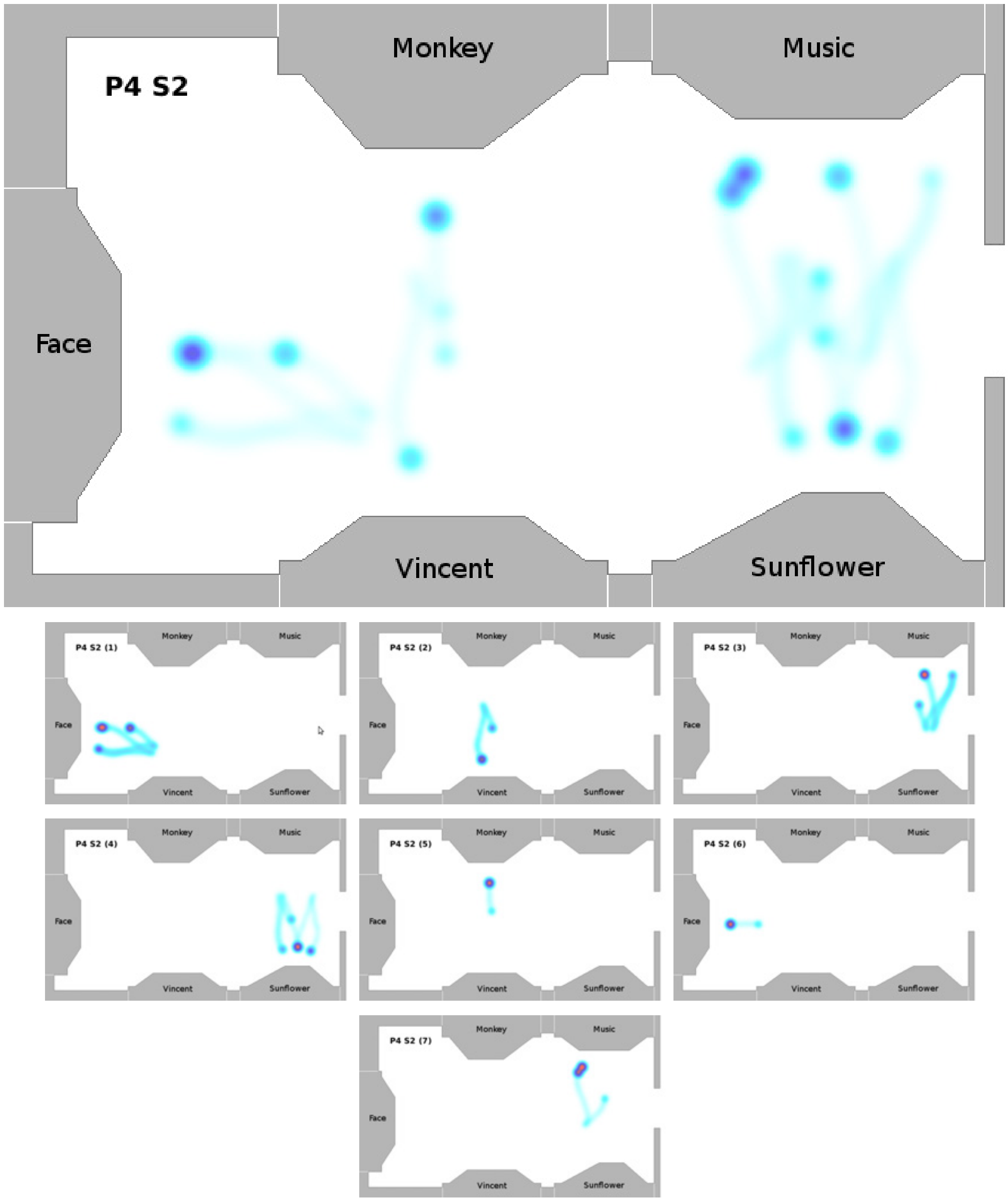


Figure 7-13: Heatmap of P4's second in-robot gallery visitation. The robot's movements within each exhibit area are shown below the larger composite image of the robot's movements in all exhibit areas.

and consistently touched the screen 2 inches to the right during his first training exercise. After his 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." P2 opted of out 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.

P3 also commented on the interface's the overall accessibility, he stated:

> 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, as stated in Chapter 5. First, there were exhibit buttons around the robot base icon with the icon and background color of each exhibit. Second, the participant could also chose a new exhibit by pressing the menu button (bright blue with white horizontal lines); 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.

Figure 7-14 shows that the participants primarily used the first method to direct the robot to a new exhibit. P3 and P4 used the menu 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.

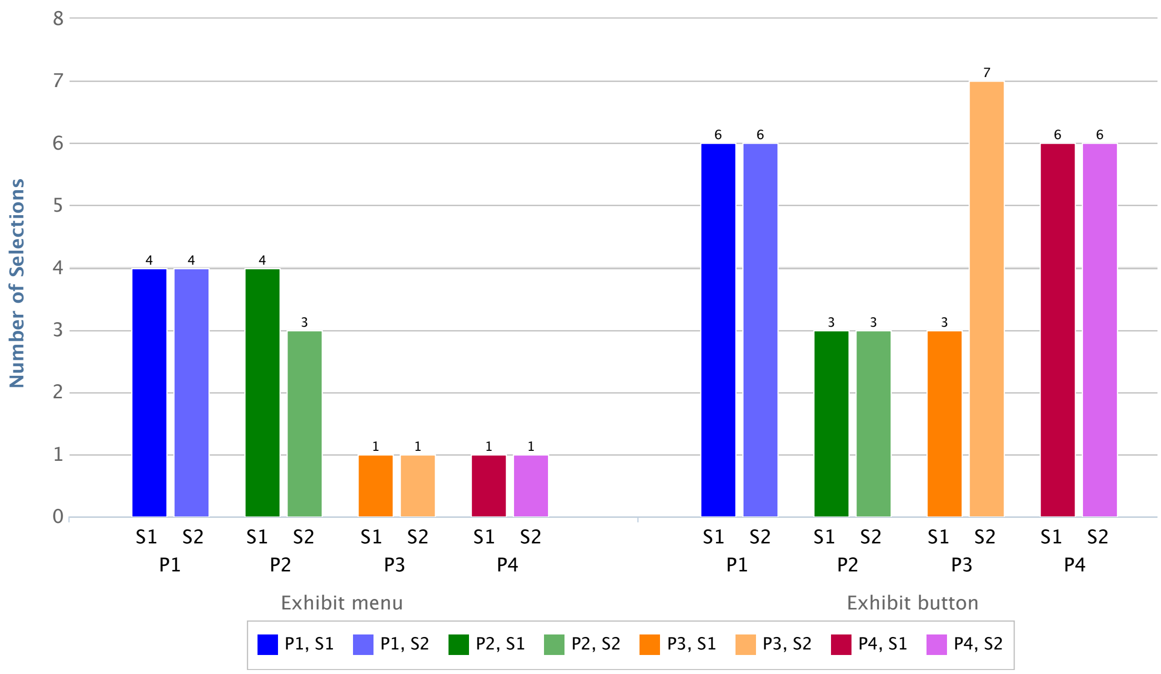


Figure 7-14: Selection of a new exhibit using the menu **(left)** and exhibit buttons around the robot base icon **(right)**.

Accessing an exhibit's information

As stated in Chapter 5, 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. 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.

Figure 7-15 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

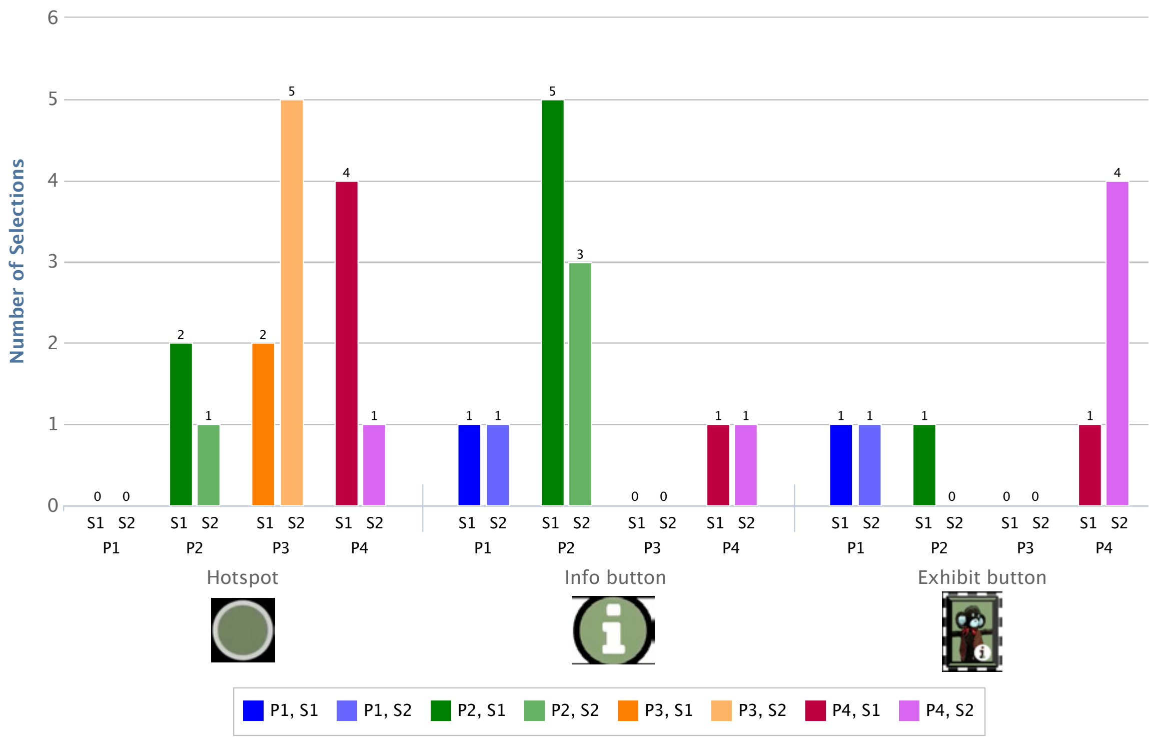


Figure 7-15: Viewing an exhibit's information by physically moving the robot to see the sign on the wall **(left)**. Participant could also open a pop-up on the interface by pressing the iButton on the lower right of the screen **(center)** or selecting the corresponding exhibit button around the robot base icon **(right)**.

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

Each exhibit had four "hotspots" that the participants could use to view the exhibit from different angles.These hotspots were shown as pulsing circles, described in Chapter 5. As noted above, the rightmost hotspot positioned the robot in front of the information sign on the wall. The other three hotspots provided a means to switch view points around the exhibit itself. Figure 7-4 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. Additionally, when the robot arrived at a hotspot, two cyan triangle buttons appeared overlaid on the robot base icon. 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 Figure 7-16, 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 [Norman, 1994], a user may feel dissatisfied due to lack of involvement and/or lack of control (e.g., [Kim et al., 2012; Kotala, 2010]). P1 felt that that it was difficult to plan ahead for where

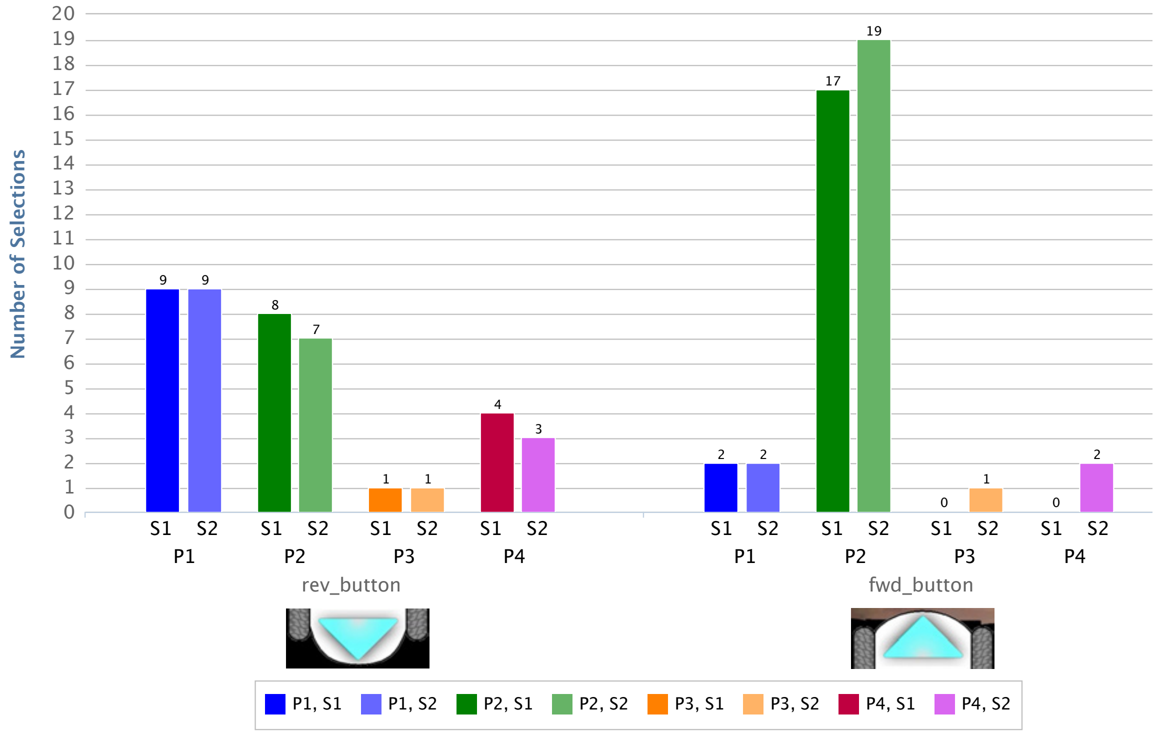


Figure 7-16: Forward and backward translation commands given by pressing the cyan triangle buttons overlaid on the robot base icon (see Figure 5-14 right).

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).

7.3.2 Interface Transparency

If a telepresence robot system has been designed well – for the user, interactants, and bystanders – the technology should just disappear [Draper et al., 1998; Takayama, 2011]. 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. In Appendix F, we provide an overview of qualitative and quantitative performance measures for the audio signal, video signal, and human-human communication.

By and large, 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. Figure 7-17 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. In both his visitations, P3 gave suggestions for how to augment different exhibits (specifically, Face, Sunflower, and Monkey). In his second visitation, P4 commented aloud that he was able to see the robot handler in the reflection of Face's eye.

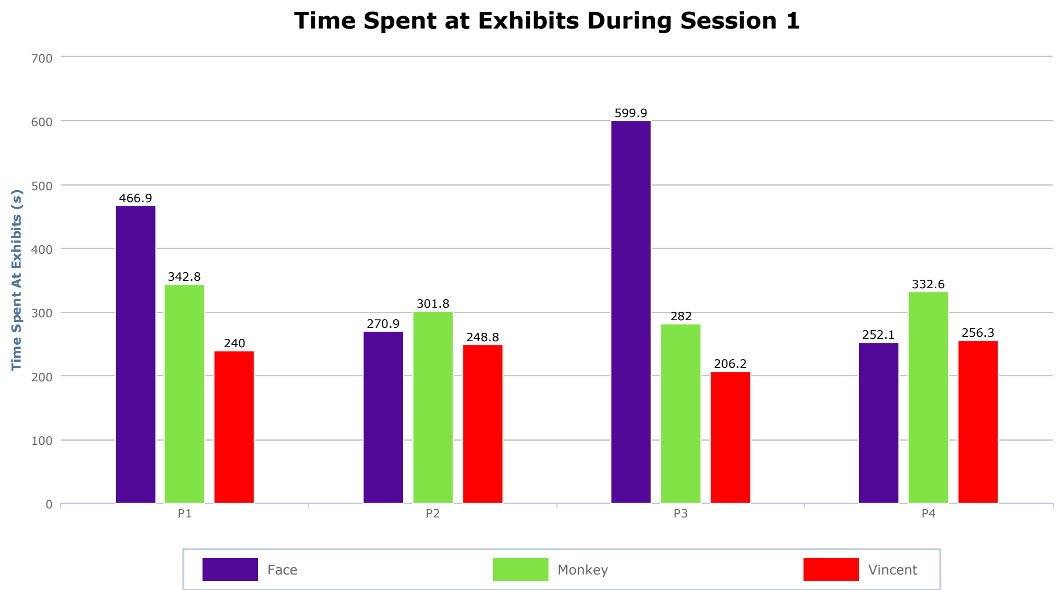
Movement and Navigation Strategies

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 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



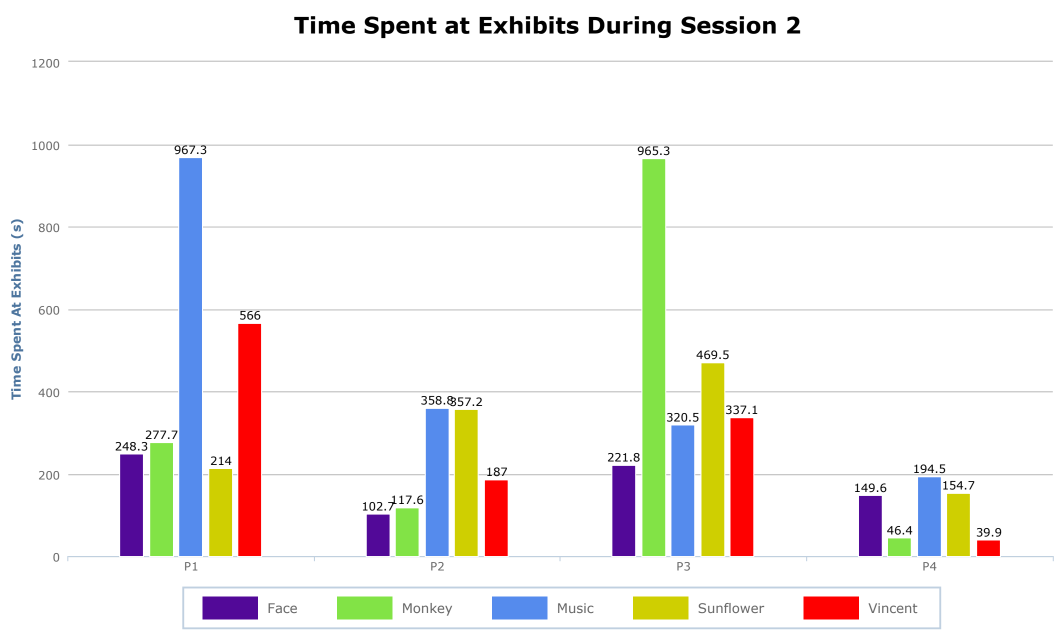


Figure 7-17: **(Top)** shows the time spent at each exhibit by each participant during Session 1. **(Bottom)** shows the time spent at each exhibit by each participant during Session 2.

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:

> 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.

Figure 7-18 shows that the confederate and participants at times 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. 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

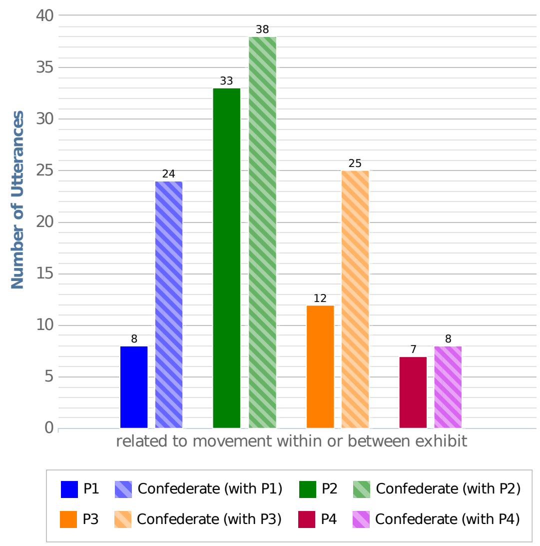


Figure 7-18: The participants and confederate verbally negotiated how to move within an exhibit to switch sides and between exhibits.

also increases [Rosenberg, 2010]. As noted in Chapter 3, 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., [Haas, 1972; Miller and Licklider, 1950]). 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 [Rubin et al., 2009] 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,

Table 7.3: Rephrasings of Yarosh and Markopoulos's Affective Benefits and Cost of Communication Technologies (ABCCT) questionnaire items

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)	How was <*confederate's_name*> feeling today? What was his or her mood?
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 <*confederate's_name*> took too long to respond to something you said? **Yes** **No** How did it make you feel? Describe the situation.
I could tell over the medium how **X** was feeling that day. (emotional expressiveness)	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 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.

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 7.3). Like Kiesler et al. [2008], 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. For the second visitation, 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

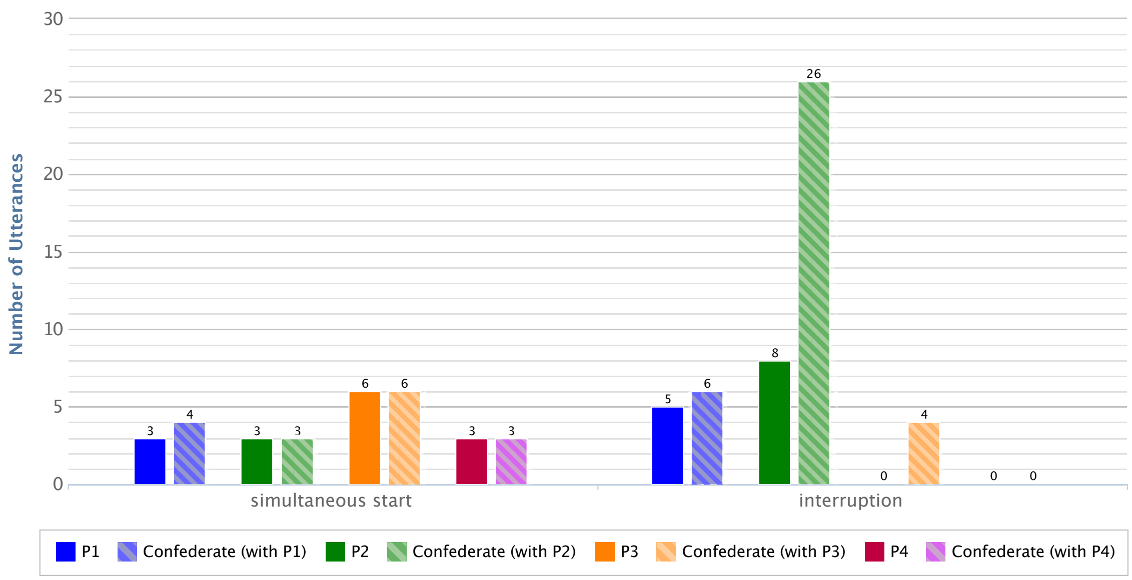


Figure 7-19: Frequency of crosstalk during the second session.

partner and even initiated a joke with her ("Is that big spider next to your shoulder supposed to be there?").

Like Fish et al. [1992], we looked at the content of the conversations between the participants and confederate in their second in-robot visitation. We discerned three salient categories: discussion of an exhibit, discussion of an aspect related to the exhibit, and off-topic conversation (Figure 7-20). Figure 7-19 shows 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

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. 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} = 4m 10s, t_{P2} = 8m 30s, t_{P3} = 6m 35s, and t_{P4} = 7m.

Unlike Study 2 (scavenger hunt, Chapter 4), participants were not given a map on

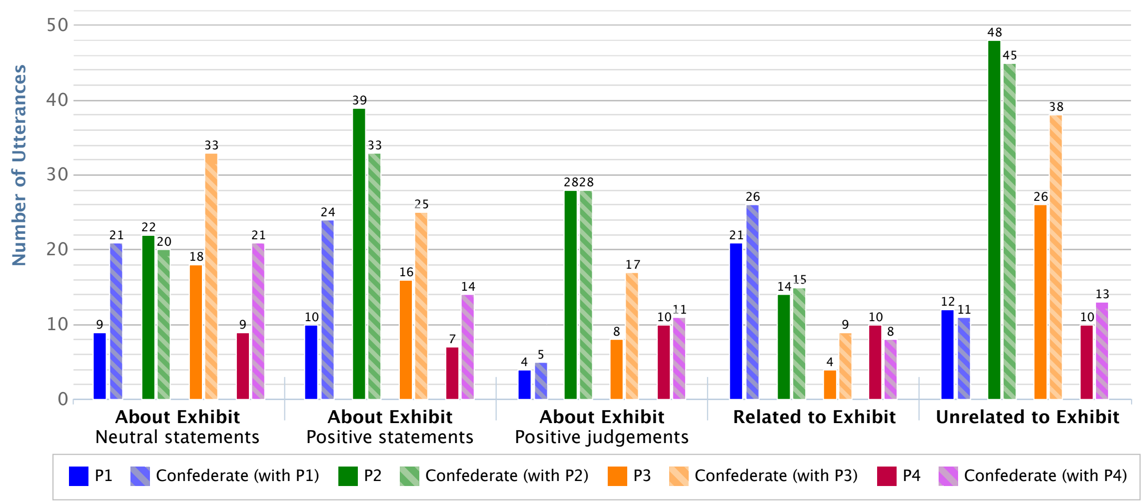


Figure 7-20: Frequency count of the participants' conversations with the confederate in the second visit to the gallery.

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 she expected.

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.

7.4 Summary

Telepresence is a multifaceted continuum of user, task, system, and environmental factors [Tsui and Yanco, 2013]. 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 the 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 (RQ4) in this case study (Study 3). The quality of an interaction via telepresence robot can be measured both quantitatively and qualitatively (Appendix F), 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.

We indirectly measured the degree to which the participants experienced the remote environment by we 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 [Kiesler et al., 2008] – 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.

Chapter 8

Conclusions and Future Work

8.1 Summary

This thesis began with four research questions. The responses to these questions depend upon the cognitive and physical abilities and other individual characteristics of the people who may use telepresence robot systems, which is one of the reasons why it is so challenging to design technologies for people who have disabilities. For the population we studied at Crotched Mountain Rehabilitation Center, we can summarize the answers to the research questions as follows.

RQ1: What levels of abstraction and autonomy are needed for people with disabilities to effectively control a telepresence robot system in a remote environment? As noted in Section 4.4, prior to this research, it was unknown how our target audience would want to direct robots in a remote environment. The people who participated in our formative assessments (Studies 1 and 2) wished to command robots using three levels of abstraction:

low-level : forward, back, left, right, and stop (FBLR commands);

mid-level : requests to send the robot to places within the robot's camera view; and

high-level : requests to send the robot to places beyond its current camera view.

A telepresence robot's autonomy must support these three levels, as discussed in

Section 6.5. Additionally, the robot's movements must be predictable and approximate that of a human's as much as possible.

RQ2: What are the essential components of a research platform needed to inform the design of future telepresence robots for the target population? We developed a set of nine requirements that we believe will apply to future telepresence robots in general, as discussed in Section 3.7. Specifically, a useful social telepresence research robot will:

- Maintain a friendly appearance,
- Remain stable when stationary and moving,
- Map an unknown environment,
- Localize within the map,
- Utilize semi-autonomous navigation,
- Have advanced sensing for interacting with people in the remote environment,
- Have a dedicated camera for driving,
- Have forward- and downward-facing camera views, and
- Have a wide field of view, both horizontally and vertically.

RQ3: Which design principles facilitate the development of telepresence robot interfaces for use by the target population? We surveyed the design guidelines and heuristics that have been developed for both general-purpose user interfaces and interfaces for assistive technologies, as described in Section 5.4. When designing our interface, we found the most relevant guidance to be the following:

- Ensure a match between the system and the real world,
- Provide visibility of system status,
- Prevent errors,

- Facilitate recognition rather than requiring recall, and

- Aid perception.

We developed a more complete set of heuristics, of which the five listed above are a subset, by pulling applicable heuristics from the literature and using the Model Human Processor [Card et al., 1983] as a theory-based framework; this set was documented in Tsui et al. [2009] and can be found in Appendix A.

Study 3 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.

RQ4: To what degree can the target population experience remote social interaction and the remote environment itself? We surveyed quantitative and qualitative performance measures needed to assess a communication by leveraging work from the fields of HCI, computer supported cooperative work (CSCW), communications, and psychology, as described in Appendix F. The quality of interaction through telepresence robots consists of 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 concepts for Study 3 indirectly as open-ended interview questions, as described in Section 7.4. We measured the degree to which the case study participants experienced the remote environment by asking them to describe their most favorite and least favorite exhibits after each of the two in-robot visitations. 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.

Overall, Study 3 demonstrated that four people from our target population were able to use our telepresence robot system to twice visit an art gallery. When visiting the gallery for the first time, they would laugh and make comments about the exhibits to themselves or to the experimenters. For both gallery visitations, every case study participant was to develop an informed opinion about their favorite exhibit and provide reasoning as to why they liked it. All participants had "normal" conversations with the confederate, and none felt obligated to. They felt co-present with the confederate and were able to discern the confederate's mood. Three participants formulated movement strategies for touring the gallery: one clockwise, one counter-clockwise, and one moving to the furthest point in and working back toward the entrance. This qualitative data indicates that the four case study participants were able experience the remote environment and engage in social interaction.

Knowledge about assistive technologies is being built incrementally as more case studies and empirical investigations are undertaken with the active participation of individuals with disabilities. In this way, a more complete picture is slowly being built that shows the relative value of different approaches for designing assistive technologies. Thus, the overarching contribution of this thesis is its addition to the body of knowledge of assistive technology design based on three years of working with the target population.

8.2 Contributions

The research presented in this dissertation is the culmination of three years of work, and has resulted in three major contributions. First and foremost was pursuing the use case of people with disabilities the taking active role of operating a telepresence robot. There has been considerable research already done in the traditional use case where the person with special needs is visited by a healthcare professional, family member, or friend operating a telepresence robot (i.e., passive role), discussed in Chapter 2. A person's quality of life is impacted when he or she is no longer able to participate in everyday activities. Hopps et al. [2001] found that for people with disabilities, there was a negative correlation between loneliness and physical independence. A telepresence robot provides a physical embodiment for its user, and in the active role, can be used to support social engagement for our target population. Designing a telepresence robot system to maximize the user experience, for a person with disabilities, requires expertise in the domains of human-robot interaction and assistive robotics and the understanding of how to balance the two.

Our second major contribution was the design, development, and architecture of our social telepresence robot research platform, Margo. A social telepresence robot is unlike a traditional robot, and must sufficiently pass as a human proxy. We conducted usability studies on prototype telepresence robots in 2010 [Desai et al., 2011; Tsui et al., 2011c], and synthesized performance measures for quality of interaction through a telepresence robot [Tsui et al., 2012]. We have identified issues related to using telepresence robots for social interaction between people in different locations, by examining this area from three perspectives: (1) designing for the robot user, who is in a remote location; (2) designing for people near the robot, who are interacting with the user; and (3) designing so that the conversation is not hampered by the technology. We discussed the design challenges and guidelines for social interaction in [Tsui and Yanco, 2013], which are relevant for all telepresence robots.

Finally, our third major contribution was an example of an "invisible to use" [Takayama, 2011] telepresence user interface designed for users to explore a remote art gallery. Our approach in designing HRI systems has been an iterative process, involving the target population (primary stakeholders), caregivers (secondary stakeholders), and clinicians from the beginning, formative stages through the summative evaluations. It was unknown how our target audience would want to direct robots in a remote environment. Two formative assessments yielded a data set of first-hand accounts of users from our target population giving spatial navigation commands to a telepresence robot. We drew a critical insight from this data set that users would command the robots using multiple levels of abstraction; that is, all users gave directives at the low-level (i.e., forward, back, 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 synthesized user interface design guidelines and principles from the domains of HRI, HCI, and assistive technology with this insight, and designed and implemented our telepresence robot system. The robot's movement and autonomous navigation behaviors supported these three levels and allowed the case study participants to examine exhibits from multiple viewpoints and move between exhibits; the manner in which the robot moved was reminiscent of how a person would explore a gallery space, keeping an exhibit in view until moving to the next one. The user interface provided a first-person, video-centric view with these three levels of control presented at the top level in a meaningful way. Our telepresence robot system was sufficiently transparent and allowed the case study participants to experience the Artbotics exhibits and sustain conversation with another person in the gallery.

8.3 Future Work and Open Research Questions

Robot systems designed for general use do not typically consider people with disabilities. Our research has demonstrated a critical first step towards having our target population take the active role of the telepresence robot operator. It should be noted that Study 3 represents the best case scenario with respect to its content and environment. 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, who were residents at CMRC, were already familiar with the confederate, who was an occupational therapy intern at the center.

After a short time using the robot, the case study participants had sufficient telepresence experience for them to imagine its use in situations beyond what was tested in our case study, which is consistent with Beer and Takayama's findings [2011]. Our participants imagined using a telepresence robot to visit galleries and museums in other states and countries (e.g., Nascar Hall of Fame in Charlotte, NC; the Smithsonian museums in Washington DC; and the Louvre in Paris, France). They also wanted to use a telepresence robot to go outside (e.g., Boston Common, Great Wall of China). P2 reported wanting to use a telepresence robot to accompany her son to the grocery store to pick produce, and her daughter to an upcoming concert. We look ahead to scaling our system to function in these larger public venues, and there a number of challenges which must be addressed. Social telepresence is in its infancy, and as such, this research has opened more questions than it has closed.

8.3.1 Autonomy

We have demonstrated that autonomous and semi-autonomous navigation behaviors, when properly implemented, can assist a person with special needs in operating telepresence robots. Autonomous navigation behaviors freed the users from the details of robot navigation, making the driving task easier; consequently, the users were able to focus on exploring the art gallery and conversing with the confederate.

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 [Tsui and Yanco, 2013]. 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.

Collision avoidance is a prospective, self-preservation behavior, and is implemented on most autonomous and semi-autonomous robots using a number of techniques (e.g., vector field histogram [Borenstein and Koren, 1991]). Blumberg and Galyean [1997] note that a prospective behavior avoids doing a specified behavior, thus *any other* behavior is technically correct regardless of any expectations the user might have [Desai, 2012; Desai et al., 2013, 2012]. These collision and obstacle avoidance behaviors may be non-deterministic, or an emergent result of the underlying algorithmic implementation and the environmental conditions [Maes, 1994]. These trajectories are not likely to emulate how a person moves through environments with other people present, and the lack of direct control noted in [Carlson and Demiris, 2010; Lankenau, 2001; Parikh et al., 2005; Viswanathan et al., 2007] may negatively impact a user's sense of telepresence.

8.3.2 System Design

The robot must be able to function sufficiently well as a human proxy. As noted in Chapters 3 and 6, we have found that a user's telepresence can be impacted by the robot's field of view, articulation of cameras, and locomotion technique. Margo's three hat cameras combined into a vertical panoramic video with a portion of the base in view positively impacted the four case study participants' feelings of telepresence. However, Margo's embodiment negatively impacted telepresence when it moved from one exhibit hotspot to another. Due to its non-holonomic base in conjunction with its fixed cameras, the robot had to first back away before approaching a new hotspot. A person would step to the side, or keep focus on the exhibit while walking to another vantage point.

In future work, we must determine how much of a telepresence robot's functionality must mimic a person's. Articulated cameras (i.e., pan and tilt) are similar to a human's shaking or nodding of his or her head. Although it is desirable to be able to control camera viewpoint, more degrees of freedom (DOFs) in robot control often translate to higher user learning requirements and effort in monitoring the status of the robot. The method of control of each additional DOF will impact a user's feeling of telepresence, as increased focus on control detracts from time on social interaction. It is also the case that a user's awareness (or lack thereof) of the pose of each DOF will impact telepresence. For example, if the forward camera is off-center when the user is driving the robot forward, the actual robot movement will be misaligned from the user's viewpoint. In a study of robots designed for urban search and rescue, it was observed that people drove with their camera off-center without realizing it, resulting in navigation errors and dangerous situations [Yanco et al., 2004]. Draper et al. [1998] found that the robot need not be a humanoid robot, looking exactly like its user; therefore, we posit that holonomic locomotion can improve a telepresence robot's ability to move in human-like trajectories.

There is also the open research question regarding telemanipulation; that is, if it is necessary for the use case of social telepresence. If it is necessary, what form should it take? Telemanipulation represents an important capability that may contribute to telepresence experiences for users of social telepresence robots. The inability to open a closed door causes a user to depend on the interactants and bystanders for assistance. Lee and Takayama [2011] reported that one user felt as if he were disabled when operating the telepresence robot; interactants have referred to this comparison as well [Takayama and Go, 2012]. Further, gesticulation is a natural part of spoken language and occurs in conversations between people, even when they cannot see each other [Goldin-Meadow et al., 1999; Gullberg and Holmqvist, 1999]. The addition of a human-like manipulator will increase both the complexity of the interface as noted above and also people's expectations of a telepresence robot. The natural next step in this line of research would be to examine full body social cues and apply them to a robot's planar movement. For example, one full body gesture for taking the floor on a hallway conversation is to step forward.

8.3.3 User Interface

The step buttons overlaid on the robot base icon and the hotspots overlaid on the camera view in our telepresence robot user interface allowed the user to directly manipulate [Shneiderman, 1993] the robot's low- and mid-level autonomous behaviors. From Study 3, 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). A similar interaction method for P1's first suggestion is utilized by the Giraff [Coradeschi et al., 2011; Giraff Technologies AB, 2011] and Texai [Guizzo, 2010]; the user drags the pointer from a central point.

The Jazz interface utilizes a 3D pointer in the robot's video [Dickert, 2011], which is similar to P1's second suggestion. When Jazz's 3D pointer is projected into the scene and is moved to an area with open floor space, the user can double-click to send the robot there autonomously; this local waypoint style control is known as telerobotic control [Sheik-Nainar et al., 2005].

Both suggestions by P1 are possible, but increase the complexity of the touch interaction. To accommodate the largest number of potential users, we selected a single point "touch and release" user interaction, and activated the robot's movement upon release without additional confirmation; that is, we did not make any assumptions about our target population's manual dexterity, finger isolation, or ability to sustain and/or repeat a touch. Moving from a single touch and release interaction to a gesture-base one warrants the addition of the user's explicit confirmation; for example, the gesture grammar by Micire [2010] terminated with a double tap of a single finger. Regardless, it holds that errors might be minimized if visual feedback of the trajectory the robot is about to embark upon is provided prior to the user's explicit confirmation.

Our interface provided the user the ability to move the robot from one exhibit to another (i.e., high-level autonomous behavior). Five exhibit buttons were placed around the robot base icon and were all visible from the top level. 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 (e.g. pictureSET [Special Education Technology, 2011]). The environment in Study 3 was quite small. 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. This HRI technique primarily exists in research (e.g., [Coltin et al., 2012; González-Jiménez et al., 2012; Michaud et al., 2010; Yanco et al., 2007a]). By incorporating a map widget, the buttons around the robot base icon could instead show exhibits to the robot's immediate left or right or behind it (utilizing a compass rose metaphor). If our system were placed in a much larger gallery or museum, for example the Museum of Science, the buttons might instead show nearby exhibit halls instead of individual exhibits.

8.3.4 Interaction Quality

Our case study was primarily focused on the user's experience (Section 7.4). There is much research to be done to address the interactant's experience. In Study 3, the confederate asked the participants on several occasions, "Can you see me?" Another open question is what cues are needed by interactants and bystanders, as well as how those cues can be provided. To improve user telepresence experiences, interactants need to know if the user can hear them and if they are in the camera view presented to the user without showing a picture-in-picture on the robot's display.

P2 reported a compensatory strategy for moving the robot to the opposite side from the confederate to keep both the exhibit and confederate in view. P3, however, noted difficulty with the lack of non-verbal social cues:

> It's like you're trying hard to focus on something to say to somebody if you're not right next to them... As you're having a conversation a lot of the times with the people, you can tell the interest on the subject just by looking at their eyes. Looking them straight in the eyes when they're telling you about something. You can see if they're interested or not. If someone keeps talking about something and they keep looking away, it's like they're not paying attention to you. You can see intensity in somebody if they're very excited about something. With the Monkey, my eyes would have probably been a lot different than Van Gogh. It would be like "there is an old guy." The Monkey was cool.

As social interaction is the primary goal of social telepresence robots, it will be necessary for telepresence robots to understand social cues. Failure to design for eye gaze, facial expressions, and nonverbal gestures will result in systems that hinder the ability to achieve telepresence for the user and/or the interactant.

8.4 Final Thoughts

There is great potential for our research to impact people with special needs both in personal and work-related scenarios. People with special needs have already begun to adopt this new technology. For example, Lyndon Baty, a Knox City, TX student who has polycystic kidney disease, has been using a VGo robot to remotely attend school since January 2011 [Richards, 2011; Robertson, 2011]. A child's job is to go to school, and in that sense, Lyndon is participating in regular school work via his telepresence robot. Each child with a disability has an individualized education program (IEP) which includes a his or her current academic level, annual academic and functional goals, and documentation of special education services and aids [US Department of Education, Office of Special Education Programs, 2004]. A student's IEP may include assistive technologies such as sensory enhancers, adaptive computer controls, environmental controls, mobility devices, and self-care devices [Virginia Assistive Technology System, 2009]. In [Virginia Assistive Technology System, 2009], robotic devices have been listed under the categories of environmental controls and self-care devices. As telepresence robots prove their value in this use case, we may find telepresence robots explicitly listed in students' IEPs in the near future. As noted in Chapter 2, to date, 50 VGo robots have been sold for classroom use, and other telepresence robots are also being used so that students can attend their classes instead of being home-schooled.

For adults with disabilities, telepresence robots could be used to engage in telecommuting or remote work. The National Council on Disability [2003] specifically notes that the "unavailability of supported employment" is a barrier for integrating people with special needs into the community. According to the American with Disabilities Act (ADA), employees and potential employees with disabilities[1] can request reasonable accommodations to the work environment or processes related to their job [US Equal Employment Opportunity Commission, 2005]. Employers are not required

[1] The Americans with Disabilities Act (ADA) Amendment Act of 2008 provided a broader definition of "disability" [US Equal Employment Opportunity Commission, 2011]. These changes have been implemented by the Equal Employment Opportunity Commission (EEOC) and active as of March 25, 2011.

to accommodate requests that incur a large expense or are difficult to implement. However, companies have already begun to investigate telepresence robots for ad-hoc conversations beyond the conference room for remote employees to be better connected. As telepresence robots become part of the corporate culture, it will become feasible for more adults with disabilities to telecommute from their residence.

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 Margo [Vanderheiden and Vanderheiden, 1992]. 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 Margo. Vanderheiden and Vanderheiden [1992] 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 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 robot user interfaces that people with disabilities (and without) will be able to use easily to visit remote people and places.

Bibliography

Abascal, J. (2002). Human-computer interaction in assistive technology: from "patchwork" to "universal design". In *IEEE International Conference on Systems, Man and Cybernetics*, volume 3, page 6. IEEE.

Abascal, J. and Azevedo, L. (2007). Fundamentals of inclusive HCI design. In *Universal Access in Human Computer Interaction. Coping with Diversity*, pages 3–9. Springer.

Abascal, J. and Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting with Computers*, 17(5):484–505. Social Impact of Emerging Technologies.

Ackerman, E. (2009). [CES 2009] Anybots QA Demo. `http://www.botjunkie.com/2009/01/14/ces-2009-anybots-qa-demo`. Accessed July 2, 2012.

Ackerman, E. and Guizzo, E. (2011). Mystery Robot Revealed: RoboDynamics Luna is Fully Programmable Adult-Size Personal Robot. IEEE Spectrum. `http://spectrum.ieee.org/automaton/robotics/home-robots/robodynamics-luna-fully-programmable-adult-size-personal-robot`. Accessed July 24, 2012.

Acosta Calderon, C., Mohan, R., and Zhou, C. (2011). Elderly Tele-care in the Singapore Context. In *Proceedings of the HRI Workshop on Social Robotics Telepresence*.

Adalgeirsson, S. and Breazeal, C. (2010). MeBot: A Robotic Platform for Socially Embodied Telepresence. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 15–22. IEEE.

Adams, D. (2012). iRobot's In-hospital Robots to Link Doctors, Patients. The Boston Globe. http://www.bostonglobe.com/business/2012/07/23/new-robot-will-connect-doctors-patients/aYB9pBvEPnel8AUuxrL1zL/story.html?s_campaign=8315. Accessed July 24, 2012.

Agarwal, R., Levinson, A., Allaf, M., Makarov, D., Nason, A., and Su, L. (2007). The RoboConsultant: Telementoring and Remote Presence in the Operating Room during Minimally Invasive Urologic Surgeries Using a Novel Mobile Robotic Interface. *Urology*, 70(5):970–974.

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

Akkeer, H. O. D. (2009). On addressee prediction for remote hybrid meeting settings. Master's thesis, University of Twente.

Allen, M., Leung, R., McGrenere, J., and Purves, B. (2008). Involving domain experts in assistive technology research. *Universal Access in the Information Society*, 7(3):145–154.

Amazon (2009). Amazon Mechanical Turk: Artificial Artificial Intelligence. http://www.mturk.com, accessed Dec. 2009.

Amer. Natl. Standards Institute (1997). S3. 5-1997, methods for the calculation of the speech intelligibility index.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H., and Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, 34(4):351–366.

Anybots (2011). Anybots – Your Personal Avatar. Webpage. http://www.anybots.com/, accessed March 2011.

Apple (2003). Apple Introduces iChat AV and iSight. Press release. http://www.apple.com/pr/library/2003/jun/23ichat.html, accessed Sept. 2010.

Associated Press (2006). Robot Rep Goes to School. Wired. http://www.wired.com/techbiz/media/news/2006/06/71078, accessed March 2011.

Barnett, P. and Knight, R. (1996). The common intelligibility scale. *Inst. of Acoustics*, 17(7):201–206.

Bartelmus, C. (2011). LIRC - Linux Infrared Remote Control. http://www.lirc.org. Accessed September 2008.

Bartneck, C. (2002). *EMuu: An Embodied Emotional Character for the Ambient Intelligent Home*. Technische Universiteit Eindhoven.

Bartneck, C. and Okada, M. (2001). Robotic user interfaces. In *Proceedings of the Human and Computer Conference*, pages 130–140. Citeseer.

Beer, J. and Takayama, L. (2011). Mobile Remote Presence Systems for Older Adults: Acceptance, Benefits, and Concerns. In *Proc. of the 6th Intl. Conf. on Human-Robot Interaction*, pages 19–26. ACM.

Bekker, M., Olson, J., and Olson, G. (1995). Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *Proc. of the 1st Conf. on Designing Interactive Systems: Processes, Practices, Methods, & Techniques*, pages 157–166. ACM.

Bennet, I. (2010). Through Robot's Eyes: Leukemia Patient Goes to School Without Leaving Home. http://www.rt.com/news/prime-time/robot-school-leukemia-patient/. Accessed July 25, 2012.

Beranek, L. (1947). Airplane quieting ii specification of acceptable noise levels. *Trans. Amer. Soc. Mech. Engrs*, 69:97–100.

Bergman, E. and Johnson, E. (2008). Towards Accessible Human-Computer Interaction. Webpage. http://www.sun.com/accessibility/docs/access_hci.jsp. Accessed August 2009.

Billard, A., Robins, B., Nadel, J., and Dautenhahn, K. (2006). Building Robota, a Mini-humanoid Robot for the Rehabilitation of Children with Autism. *RESNA Assistive Technology Journal*, 19(1):37–49.

Björkman, P. and Hedman, L. (2010). Development of a Collision Avoidance System for a Videoconferencing Robot. Master's thesis, Mälardalen University.

Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C., and Lee, D. (2006). The Art of Designing Robot Faces: Dimensions for Human-Robot Interaction. In *ACM SIGCHI/SIGART Conf. on Human-Robot Interaction (HRI)*, pages 331–332.

Blumberg, B. and Galyean, T. (1997). Multi-level control for animated autonomous agents: Do the right thing... oh, not that... *Creating Personalities for Synthetic Actors*, pages 74–82.

Borenstein, J. and Koren, Y. (1991). The vector field histogram-fast obstacle avoidance for mobile robots. *IEEE Transactions on Robotics and Automation*, 7(3):278–288.

Boyle, E., Anderson, A., and Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37(1):1–20.

Braddock, D. (2002). The state of the states in developmental disabilities: 2002 study summary. Coleman Institute for Cognitive Disabilities and the Department of Psychiatry, University of Colorado.

Bradley, N. and Poppen, W. (2003). Assistive technology, computers and Internet may decrease sense of isolation for homebound elderly and disabled persons. *Technology and Disability*, 15(1):19–25.

Broadbent, E., Stafford, R., and MacDonald, B. (2009). Acceptance of Healthcare Robots for the Older Population: Review and Future Directions. *International Journal of Social Robotics*, 1(4):319–330.

Broekens, J., Heerink, M., and Rosendal, H. (2009). Assistive Social Robots in Elderly Care: A Review. *Gerontechnology*, 8(2):94–103.

Brown, R. (2013). A swiveling proxy that will even wear a tutu. http://www.nytimes.com/2013/06/08/education/for-homebound-students-a-robot-proxy-in-the-classroom.html. Accessed Apr. 2014.

Bugmann, G., Klein, E., Lauria, S., and Kyriacou, T. (2004). Corpus-based robotics: A route instruction example. In *Proc. of Intelligent Autonomous Sys.*, pages 96–103.

Bugmann, G., Lauria, S., Kyriacou, T., Klein, E., Bos, J., and Coventry, K. (2001). Using verbal instructions for route learning: Instruction analysis. *Proc. TIMR*, pages 1–6.

Burgard, W., Cremers, A., Fox, D., Hähnel, D., Lakemeyer, G., Schulz, D., Steiner, W., and Thrun, S. (1998). The Interactive Museum Tour-Guide Robot. In *Proceedings of the National Conference on Artificial Intelligence*, pages 11–18. JOHN WILEY & SONS LTD.

Burt, J. (2014). iRobot's Cisco-based Ava 500 telepresence robot goes on sale. http://www.techweekeurope.co.uk/news/irobots-cisco-based-ava-500-telepresence-robot-now-available-141804. Accessed Mar. 2014.

Campbell, M. (2011). Knox City Batybot Meets the Governor and the Rest of the World. Webpage. http://www.esc9.net/vnews/display.v/ART/4d66824d3c401, accessed March 2011.

Card, S. K., Moran, T. P., and Newell, A. (1983). The psychology of human computer interaction Hillsdale. *NJ: LEA*.

Carlson, T. and Demiris, Y. (2010). Increasing robotic wheelchair safety with collaborative control: Evidence from secondary task experiments. In *2010 IEEE*

International Conference on Robotics and Automation (ICRA), pages 5582–5587. IEEE.

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

Casper, J. and Murphy, R. (2003). 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.

CastingWords (2012). Audio transcription services: MP3s, video, CD/DVDs. `http://castingwords.com`. Accessed Sept. 2012.

Cesta, A., Coradeschi, S., Cortellessa, G., Gonzalez, J., Tiberio, L., and Von Rump, S. (2011). Enabling Social Interaction Through Embodiment in ExCITE. In *Proc. of Human-Robot Interaction Workshop on Social Robotic Telepresence*.

Cesta, A., Cortellessa, G., Orlandini, A., and Tiberio, L. (2013). Evaluating telepresence robots in the field. In *Agents and Artificial Intelligence*, pages 433–448. Springer.

Cheetham, A., Young, C., and Fels, D. (2000). Interface Development for a Child's Video Conferencing Robot. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 44, pages 380–383. Human Factors and Ergonomics Society.

Chellali, R. and Baizid, K. (2011). 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.

Chen, J. Y., Haas, E. C., and Barnes, M. J. (2007). 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.

Chernova, S., Orkin, J., and Breazeal, C. (2010). Crowdsourcing HRI through online multiplayer games. In *Proc. Dialog with Robots: AAAI Fall Symp.*

Choi, Y. M. (2011). Managing input during assistive technology product design. *Assistive Technology®*, 23(2):65–75.

Chung, K., Grathwohl, K., Poropatich, R., Wolf, S., and Holcomb, J. (2007). Robotic Telepresence: Past, Present, and Future. *J. of Cardiothoracic and Vascular Anesthesia*, 21(4):593–596.

Cohen, B., Lanir, J., Stone, R., and Gurevich, P. (2011a). Requirements and design considerations for a fully immersive robotic telepresence system. In *Proccedings of HRI 2011 Workshop on Social Robotic Telepresence*, pages 16–22.

Cohen, I., Looije, R., and Neerincx, M. A. (2011b). Child's recognition of emotions in robot's face and body. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 123–124. ACM.

Coltin, B., Biswas, J., Pomerleau, D., and Veloso, M. (2012). Effective semi-autonomous telepresence. *RoboCup 2011: Robot Soccer World Cup XV*, pages 365–376.

CompuLab (2012). fit-PC2 specifications. Webpage. `http://www.fit-pc.com/web/fit-pc/fit-pc2-specifications`. Accessed Nov. 2012.

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

Cooper, R. A., Grindle, G., Vazquez, J., Xu, J., Wang, H., Candiotti, J., Chung, C., Salatin, B., Houston, E., Kelleher, A., et al. (2012). Personal mobility and manipulation appliance – Design, development, and initial testing. *Proceedings of the IEEE*, 100(8):2505–2511.

Coradeschi, S., Loutfi, A., Kristoffersson, A., Von Rump, S., Cesta, A., and Cortellessa, G. (2011). Towards a Methodology for Longitudinal Evaluation of Social Robotic

Telepresence for Elderly. In *Proc. of Human-Robot Interaction Workshop on Social Robotic Telepresence*.

Crandall, J. W., Goodrich, M. A., Olsen Jr, D. R., and Nielsen, C. W. (2005). Validating human-robot interaction schemes in multitasking environments. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 35(4):438–449.

Cronbach, L. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3):297–334.

CtrlWorks (2013). Introducing the Puppet. `http://www.ctrlworks.com/products/puppet`. Accessed Mar. 2014.

Deml, B. (2007). Human factors issues on the design of telepresence systems. *Presence: Teleoperators and Virtual Environments*, 16(5):471–487.

Desai, M. (2012). *Modeling Trust to Improve Human-Robot Interaction*. PhD thesis, University of Massachusetts Lowell.

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., and Yanco, H. (2013). Impact of robot failures and feedback on real-time trust. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 251–258. IEEE Press.

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., Steinfeld, A., and Yanco, H. (2012). Effects of changing reliability on trust of robot systems. In *Proceedings of HRI*. ACM/IEEE.

Desai, M., Tsui, K., Yanco, H., and Uhlik, C. (2011). Essential Features of Telepresence Robots. In *Proc. of the IEEE Conf. on Technologies for Practical Robot Applications (TePRA)*.

Dickert, S. (2011). From Russia With Love – R.BOT Telepresence Robot. `http://www.pilotpresence.com/2011/01/`

from-russia-with-love-r-bot-telepresence-robot/. Accessed February 14, 2012.

Ding, D., Cooper, R. A., and Pearlman, J. (2007a). Incorporating participatory action design into research and education. In *International Conference on Electrical Engineering*.

Ding, X., Erickson, T., Kellogg, W., Levy, S., Christensen, J., Sussman, J., Wolf, T., and Bennett, W. (2007b). An empirical study of the use of visually enhanced voip audio conferencing: the case of ieac. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 1019–1028. ACM.

DiSalvo, C., Gemperle, F., Forlizzi, J., and Kiesler, S. (2002). All Robots Are Not Created Equal: The Design and Perception of Humanoid Robot Heads. In *Conf. on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pages 321–326. ACM.

Double Robotics (2013). Double Robotics: Work from anywhere. http://www.doublerobotics.com. Accessed Apr. 2014.

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

Drury, J. L., Klein, G., Pfaff, M. S., and More, L. D. (2009). Dynamic decision support for emergency. In *Technologies for Homeland Security, 2009. HST'09. IEEE Conference on*, pages 537–544. IEEE.

Duncan, J. (2012). Students raise funds to purchase VGo for homebound classmate. http://ivc.esc6.net/?p=40. Accessed Apr. 2014.

Ebben, P. (2014). Eye on education: Tyngsboro High student uses robot to keep up in school. http://boston.cbslocal.com/2014/03/03/eye-on-education-tyngsboro-high-student-uses-robot-to-keep-up-in-school.

Eberhard, K. M., Nicholson, H., Kúbler, S., Gundersen, S., and Scheutz, M. (2010). The Indiana "cooperative remote search task" (CReST) corpus. In *Proc. of the Intl. Conf. on Language Resources and Evaluation (LREC)*.

Endsley, M. (1988). Design and evaluation for situation awareness enhancement. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 32(2), pages 97–101. Human Factors and Ergonomics Society.

Evercool USA (2010). FAN-EC8010LL05E. Website. `http://http://www.evercoolusa.com/?p=1422`. Accessed Aug. 2012.

Farrand, M. (2014). Robot assists homebound students. `http://www.standard-journal.com/news/free/article_a2e64bc2-c0a4-11e3-9488-0019bb2963f4.html`. Accessed Apr. 2014.

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

Fels, D., Williams, L., Smith, G., Treviranus, J., and Eagleson, R. (1999). Developing a Video-mediated Communication System for Hospitalized Children. *Telemedicine Journal*, 5(2):193–208.

Findlay, R. (2003). Interventions to Reduce Social Isolation Amongst Older People: Where Is the Evidence? *Ageing & Society*, 23(05):647–658.

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

Fishman, S. (2014). School's video spotlighted in White House student film festival. `http://florida.icito.com/schools-video-spotlighted-in-white-house-student-film-festival`. Accessed Apr. 2014.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381.

Fitzgerald, J. (2011). After Surgery, A Robot May Be at Your Side. The Boston Globe. http://www.boston.com/business/technology/articles/2011/12/12/the_robot_that_makes_house_calls/. Accessed July 26, 2012.

Fliesler, N. (2011). A Roving Eye: Home Health Monitoring with Robotic Systems. http://vectorblog.org/2011/12/a-roving-eye-home-health-monitoring-with-robotic-systems/. Accessed July 26, 2012.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166.

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

Foote, T. (2010). Standard units of measure and coordinate conventions. http://www.ros.org/reps/rep-0103.html. Accessed Apr. 2014.

Foote, T. (2013). tf: the transform library. In *IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, pages 1–6. IEEE.

Fox, W. (2005). *I, Robot Entrepreneur*. America First Books.

Fraunhofer IPA (2012a). Care-O-bot® I. http://www.care-o-bot.de/english/Care-O-bot_1.php. Accessed July 28, 2012.

Fraunhofer IPA (2012b). Care-O-bot® I. http://www.care-o-bot.de/english/Care-O-bot_2.php. Accessed July 28, 2012.

Fraunhofer IPA (2012c). Care-O-bot®: We create your future! http://www.care-o-bot.de/english/. Accessed July 28, 2012.

Gargett, A., Garoufi, K., Koller, A., and Striegnitz, K. (2010). The GIVE-2 corpus of giving instructions in virtual environments. In *Proc. of LREC*.

Geelhoed, E., Parker, A., Williams, D., and Groen, M. (2009). Effects of Latency on Telepresence. Technical Report HPL-2009-120, Hewlett-Packard Labs.

Gergle, D., Kraut, R., and Fussell, S. (2004). Action as language in a shared visual space. In *Proc. of Computer Supported Cooperative Work*.

Gerkey, B., Vaughan, R., and Howard, A. (2003). The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems. In *Proc. of Intl. Conf. on Advanced Robotics*, pages 317–323.

Gerkey, B. P. (2011). amcl - ROS Wiki. `http://www.ros.org/wiki/amcl`. Accessed August 8, 2012.

Gerkey, B. P., Leibs, J., and Gassend, B. (2011). hokuyo_node - ROS Wiki. `http://www.ros.org/wiki/hokuyo_node`. Accessed July 31, 2012.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

Giraff Technologies AB (2011). Giraff: Confidence for Independent Living. Webpage. `http://giraff.org/`, accessed March 2011.

Gist, Y. and Hetzel, L. (2004). We the People: Aging in the United States. Census 2000 Special Reports. CENSR-19.

Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine de Gruyter.

Goetz, J., Kiesler, S., and Powers, A. (2003). Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation. In *IEEE Intl. Wkshp. on Robot and Human Interactive Communication (ROMAN)*, pages 55–60.

Goldin-Meadow, S., Kim, S., and Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, 91(4):720.

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

González-Jiménez, J., Galindo, C., and Ruiz-Sarmiento, J. (2012). Technical improvements of the Giraff telepresence robot based on users' evaluation. In *RO-MAN*, pages 827–832. IEEE.

Gonzalez-Jorge, H., Riveiro, B., Vazquez-Fernandez, E., Martínez-Sánchez, J., and Arias, P. (2013). Metrological evaluation of Microsoft Kinect and Asus Xtion sensors. *Measurement*.

Gorman, M. (2012). Suitable technologies introduces beam, the remote presence device. http://www.engadget.com/2012/09/26/beam-telepresence-device-suitable-technologies. Accessed April 2014.

Gostai (2011). Jazz. Webpage. http://www.gostai.com/, accessed March 2011.

Graf, B., Parlitz, C., and Hägele, M. (2009). Robotic Home Assistant Care-O-bot® 3 Product Vision and Innovation Platform. *Lecture Notes in Computer Science. Human-Computer Interaction. Novel Interaction Methods and Techniques*, pages 312–320.

Grayson, D. and Coventry, L. (1998). The effects of visual proxemic information in video mediated communication. *ACM SIGCHI Bulletin*, 30(3):30–39.

Gridley, C. M., Retik, A. B., Cilento, B., and Nguyen, H. T. (2012). 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*. Session abstract. https://aap.confex.com/aap/2012/webprogram/Paper17310.html. Accessed Nov. 2013.

Groom, V. and Nass, C. (2007). Can robots be teammates? benchmarks in humanrobot teams. *Interaction Studies*, 8(3):483–500.

Guizzo, E. (2010). When My Avatar Went to Work. IEEE Spectrum. http://spectrum.ieee.org/robotics/industrial-robots/when-my-avatar-went-to-work/, accessed Sept. 2010.

Guizzo, E. (2011). Can Japan Send In Robots To Fix Troubled Nuclear Reactors? IEEE Spectrum. `http://spectrum.ieee.org/automaton/robotics/industrial-robots/japan-robots-to-fix-troubled-nuclear-reactors`, accessed April 2011.

Gullberg, M. and Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, 7(1):35–63.

Ha-Won, J. (2010). English Teachers Replaced by Robots in Korea Institute of Science and Technology Project. Seoul/AFP. Accessed July 24, 2012.

Haas, H. (1972). The Influence of a Single Echo on the Audibility of Speech. *J. Audio Eng. Soc.*, 20(2):146–159.

Hadzipetros, P. (2007). Dr. Robot. CBC News. `http://www.cbc.ca/news/background/tech/robotics/robot-doctor.html`, accessed Sept. 2010.

Hans, M., Graf, B., and Schraft, R. (2002). Robotic Home Assistant Care-O-bot: Past–Present–Future. In *Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication*, pages 380–385. IEEE.

Hassenzahl, M. (2011). Encyclopedia Chapter on User Experience and Experience Design. Webpage. `http://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html`, accessed April 2011.

Heinzmann, J. and Zelinsky, A. (2003). Quantitative safety guarantees for physical human-robot interaction. *The International Journal of Robotics Research*, 22(7-8):479–504.

Helal, A. and Abdulrazak, B. (2006). TeCaRob: Tele-Care using Telepresence and Robotic Technology for Assisting People with Special Needs. *Intl. J. of Human-friendly Welfare Robotic Systems*, 7(3).

Helal, A. and Abdulrazak, B. (2007). Toward a Scalable Home-Care Delivery for Frail Elders and People with Special Needs. In *Proceedings of the IEEE 10th International Conference on Rehabilitation Robotics (ICORR)*, pages 994–998. IEEE.

Hetzel, L. and Smith, A. (2001). The 65 Years and Over Population: 2000. Census 2000 Brief. C2KBR/01-10.

Hine, B., Stoker, C., Sims, M., Rasmussen, D., Hontalas, P., Fong, T., Steele, J., Barch, D., Andersen, D., Miles, E., et al. (1994). The Application of Telepresence and Virtual Reality to Subsea Exploration. In *Proc. of Workshop on: Mobile Robots for Subsea Environments, Intl. Advanced Robotics Program*, pages 117–126.

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

Hooker, M. (2011). Robot Goes to Class for Student with Dangerous Allergies. CBS Denver.

Hoover, R. (2011). The Force is Strong with NASA's Smartphone-Powered Satellite. `http://www.nasa.gov/mission_pages/station/main/spheres_smartphone.html`. Accessed Apr. 2014.

Hopps, S., Pepin, M., Arseneau, I., Frechette, M., and Begin, G. (2001). Disability Related Variables Associated with Loneliness among People with Disabilities. *Journal of Rehabilitation*, 67(3).

Hornyak, T. (2011a). CES: Pilot iRobot's AVA Telepresence Bot with iPad. Webpage. `http://ces.cnet.com/8301-32254_1-20027701-283.html`, accessed March 2011.

Hornyak, T. (2011b). Robots Help Sick Kids Go to School. CNET. `http://news.cnet.com/8301-17938_105-20030156-1.html`. Accessed July 28, 2012.

Hornyak, T. (2012). School Buys Telepresence Robot to Help Sick Student. CNET News. `http://news.cnet.com/8301-17938_105-57434925-1/school-buys-telepresence-robot-to-help-sick-student/`. Accessed July 25, 2012.

Ideas On Board (2012). Linux UVC driver and tools – FAQ. Webpage. http://www.ideasonboard.org/uvc/faq. Accessed Aug. 2013.

IJsselsteijn, W., Baren, J., Markopoulos, P., Romero, N., and Ruyter, B. (2009). Measuring affective benefits and costs of mediated awareness: Development and validation of the abc-questionnaire. *Awareness Systems*, pages 473–488.

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

Intl. Electrotechnical Commission (1998). 60268-16. objective rating of speech intelligibility by the speech transmission index.

Intl. Telecommunication Union (1999). Rec. p.910, subjective video quality assessment methods for multimedia applications.

Intl. Telecommunication Union (2001). Rec. P.805, Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs.

Intl. Telecommunication Union (2002). Rec. bt.500-11,methodology for the subjective assessment of the quality of television pictures.

Intl. Telecommunication Union (2007). Rec. P.805, Subjective Evaluation of Conversational Quality.

Intl. Telecommunication Union (2008). Rec. j.247, objective perceptual multimedia video quality measurement in the presence of a full reference.

Intl. Telecommunication Union (2009). G.711: Pulse Code Modulation (PCM) of Voice Frequencies. http://www.itu.int/rec/T-REC-G.711/en, accessed Nov. 2011.

Intl. Telecommunication Union (2011). G.729: Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP). http://www.itu.int/rec/T-REC-G.729/en, accessed Nov. 2011.

InTouch Health (2010). Remote Presence endpoint devices. http://www.intouchhealth.com/products_remote_presence_endpoint_devices.html, accessed Nov. 2010.

InTouch Health. (2011). FAQ. http://www.intouchhealth.com/ITHFAQs.pdf, accessed Nov. 2011.

InTouch Health (2013). InTouch Health receives FDA clearance for the RP-VITA remote presence robot. Press release. http://www.intouchhealth.com/company/press-releases/01-08-2013.

InTouch Technologies (2011). Newsroom: Media Kit. Webpage. http://www.intouchhealth.com/mediakit.html, accessed April 2011.

InTouch Technologies, Inc. (2003a). InTouch Health and Rehabilitation Institute at Santa Barbara Collaborate to Evaluate World's First Mobile Remote Presence Robot for Healthcare. Press release. http://www.intouchhealth.com/pr1-15-03.html, accessed Sept. 2010.

InTouch Technologies, Inc. (2003b). World's First Remote Presence Robot Used by Healthcare Experts in Elder Care Begins Clinical Testing. Press release. http://www.intouchhealth.com/pr2-24-03.html, accessed Sept. 2010.

InTouch Technologies, Inc. (2004). Robots at Sentara General Hospital Boosting Doctors' Efficiency. Press release. http://www.intouchhealth.com/pr4-27-04.html, accessed Sept. 2010.

InTouch Technologies, Inc. (2012a). InTouch Health and iRobot unveil the RP-VITA telemedicine robot at clinical innovations forum. Press release.

InTouch Technologies, Inc. (2012b). InTouch Health and iRobot Unveil the RP-VITA^{TM} Telemedicine Robot at Clinical Innovations Forum. Press release. http://www.intouchhealth.com/pr7_24_2012_InTouch_Health_and_iRobot_to_Unveil_the_RP-VITA_Telemedicine_Robot_at_Clinical_Innovations_Forum.html. Accessed July 24, 2012.

iRobot (2007). iRobot Unveils Two New Robots at DigitalLife. Press release. http://www.irobot.com/sp.cfm?pageid=86\&id=366\&referrer=169, accessed Sept. 2010.

iRobot (2014). iRobot Ava 500: Video Collaboration Robot. http://www.irobot.com/us/learn/commercial/ava500.aspx. Accessed Apr. 2014.

iRobot Corporation (2011). Ava mobile robotics platform. http://www.irobot.com/filelibrary/pdfs/hrd/ava/Ava-Brochure-May2-2011.pdf. Accessed Oct. 2012.

Isaacs, E. and Tang, J. (1994). What video can and cannot do for collaboration: a case study. *Multimedia Systems*, 2(2):63–73.

Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2011). Do Elderly People Prefer a Conversational Humanoid as a Shopping Assistant Partner in Supermarkets? In *Proceedings of the 6th Intl. COnf. on Human-Robot Interaction*, pages 449–456. ACM.

Jay, G. T. and Crick, C. (2012). gscam - ROS Wiki. http://www.ros.org/wiki/gscam. Accessed July 31, 2012.

Jokinen, K., Nishida, M., and Yamamoto, S. (2009). Eye-gaze experiments for conversation monitoring. In *Proc. of the 3rd Intl. Universal Communication Symp.*, pages 303–308. ACM.

Kaber, D. B. and Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3):126–131.

Kac, E. (1998). Live from Mars. *Leonardo*, 31(1):1–2.

Kelley, J. (1984). An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. *ACM Transactions on Information Systems (TOIS)*, 2(1):26–41.

Keyes, B., Casey, R., Yanco, H. A., Maxwell, B. A., and Georgiev, Y. (2006). Camera placement and multi-camera fusion for remote robot operation. In *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*, pages 22–24.

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

Kim, D., Hazlett, R., Godfrey, H., Rucks, G., Portee, D., Bricout, J., Cunningham, T., and Behal, A. (2010). On the relationship between autonomy, performance, and satisfaction: Lessons from a three-week user study with post-SCI patients using a smart 6DOF assistive robotic manipulator. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 217–222. IEEE.

Kim, D., Hazlett-Knudsen, R., Culver-Godfrey, H., Rucks, G., Cunningham, T., Portee, D., Bricout, J., Wang, Z., and Behal, A. (2012). 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.

Kirbis, D. S. (2013). Logitech c920 and c910 fields of view for RGBDtoolkit. `http://therandomlab.blogspot.com/2013/03/logitech-c920-and-c910-fields-of-view.html`. Accessed Feb. 2014.

Kirk, D., Sellen, A., and Cao, X. (2010). Home video communication: mediating 'closeness'. In *Proc. of the 2010 ACM Conf. on Computer Supported Cooperative Work (CSCW)*, pages 135–144. ACM.

Koenig, N. and Howard, A. (2004). Design and use paradigms for gazebo, an open-source multi-robot simulator. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 3, pages 2149–2154. IEEE.

Koffka, K. (2013). *Principles of Gestalt psychology*. Routledge, 4 edition.

Kotala, Z. G. (2010). Robotic Arm's Big Flaw: Patients Say It's 'Too Easy'. Webpage. http://today.ucf.edu/robotic-arms-big-flaw-patients-say-its-too-easy, accessed Nov. 2012.

Koulouri, T. and Lauria, S. (2009). A WOz Framework for Exploring Miscommunication in HRI. In *Procs. of the AISB Symposium on New Frontiers in Human-Robot Interaction*.

Kruse, T., Pandey, A. K., Alami, R., and Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743.

Kurniawan, S. and Zaphiris, P. (2005). 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.

Kwon, O.-H., Koo, S.-Y., Kim, Y.-G., and Kwon, D.-S. (2010). Telepresence Robot System for English Tutoring. In *IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 152 –155.

KWWL (2011). Boy Learns Via Skype & Robot. http://www.kwwl.com/story/14753944/boy-learns-via-skype-ro. Accessed February 14, 2012.

Lachapelle, S. (2008). Talk face-to-face right from within Gmail. Official Google Blog. http://googleblog.blogspot.com/2008/11/talk-face-to-face-right-from-within.html, accessed Sept. 2010.

Lai, F. (2008). Robotic Telestroke Program: Transforming Stroke Care. *Endovascular Today Supplement, Current Concepts in Comprehensive Stroke Care*, 11:8–11.

Lankenau, A. (2001). Avoiding mode confusion in service robots-the bremen autonomous wheelchair as an example. In *Proceedings of the 7th International Conference on Rehabilitation Robotics (ICORR 2001)*, pages 162–167.

Lee, K. M., Yates, D., Clark, J., and Sawy, O. E. (2010). Value creation of mobile services through presence: Designing mobile information and entertainment applications with presence in mind. *Presence: Teleoperators and Virtual Environments*, 19(3):265–279.

Lee, M., Forlizzi, J., Rybski, P., Crabbe, F., Chung, W., Finkle, J., Glaser, E., and Kiesler, S. (2009). The Snackbot: Documenting the Design of a Robot for Long-Term Human-Robot Interaction. In *ACM/IEEE Intl. Conf. on HRI*, pages 7–14.

Lee, M. and Takayama, L. (2011). Now, I Have a Body: Uses and Social Norms for Mobile Remote Presence in the Workplace. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 33–42. ACM.

Leibs, J. and Gassend, B. (2011). microstrain_3dmgx2_imu - ROS Wiki. `http://www.ros.org/wiki/microstrain_3dmgx2_imu`. Accessed July 31, 2012.

Likert, R. (1974). The method of constructing an attitude scale. *Scaling: a sourcebook for behavioral scientists*, pages 233–243.

Lindgren-Streicher, A. and Reich, C. (2011). Visitor experience monitoring project: Fiscal year 2010 report.

Lo, C. (2010). Hospital Robots of the Future. Webpage. `http://www.hospitalmanagement.net/features/feature83720/`, accessed March 2011.

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

Lopez, S. (2013). VGo robot keeps MISD students present in school. `http://ivc.esc6.net/?p=774`. Accessed Apr. 2014.

Loutfi, A. (2010). ExCITEt. Webpage. `http://www.oru.se/excite`, accessed April 2011.

Ma, R. and Kaber, D. B. (2006). Presence, workload and performance effects of synthetic environment design factors. *International Journal of Human-Computer Studies*, 64(6):541–552.

Mace, J. (2013). Rosbridge – ROS wiki. Webpage. http://www.ros.org/wiki/rosbridge_suite. Accessed Aug. 2013.

Maes, P. (1994). Modeling adaptive autonomous agents. *Artificial Life*, 1:135–162.

Manning, N. (2012). 9th Sense Robotics: Enriching the Human Experience. Investor Presentation. Accessed July 24, 2012.

Marder-Eppstein, E. (2014). ROS Navigation Stack. http://wiki.ros.org/navigation. Accessed Mar. 2014.

Marder-Eppstein, E. and Perko, E. (2014). base_local_planner. http://wiki.ros.org/base_local_planner. Accessed Apr. 2014.

Marge, M. and Rudnicky, A. (2011). The TeamTalk corpus: Route instructions in open spaces. In *Proc. of RSS Wksp. on Grounding Human-Robot Dialog for Spatial Tasks*.

Marti, P., Bacigalupo, M., Giusti, L., Mennecozzi, C., and Shibata, T. (2006). Socially Assistive Robotics in the Treatment of Behavioural and Psychological Symptoms of Dementia. In *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 483–488. IEEE.

Massie, K. (2014). Headline classroom robots make school possible for home-bound students. http://www.simpsonstreetfreepress.org/science-and-technology/robot-substitution. Accessed Apr. 2014.

Mast, M., Burmester, M., Krüger, K., Fatikow, S., Arbeiter, G., Graf, B., Kronreif, G., Pigini, L., Facal, D., and Qiu, R. (2012). User-Centered Design of a Dynamic-Autonomy Remote Interaction Concept for Manipulation-Capable Robots to Assist Elderly People in the Home. *J. of Human-Robot Interaction*, 1(1).

Michalisin, M., Karau, S., and Tangpong, C. (2004). The Effects of Performance and Team Cohesion on Attribution: A Longitudinal Simulation. *J. of Business Research*, 57(10):1108–1115.

Michaud, F., Boissy, P., Labonté, D., Brière, S., Perreault, K., Corriveau, H., Grant, A., Lauria, M., Cloutier, R., Roux, M., et al. (2010). Exploratory Design and Evaluation of a Homecare Teleassistive Mobile Robotic System. *Mechatronics*, 20(7):751–766.

Michaud, F., Boissy, P., Labonté, D., Corriveau, H., Grant, A., Lauria, M., Cloutier, R., Roux, M., Iannuzzi, D., and Royer, M. (2008). A Telementoring Robot for Home Care. In Mihailidis, A. et al., editors, *Technology and Aging: Selected Papers from the 2007 International Conference on Technology and Aging*, volume 21. IOS Press.

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

Micire, M. J. (2010). *Multi-Touch Interaction for Robot Command and Control*. PhD thesis, University of Massachusetts Lowell.

Miller, G. and Licklider, J. (1950). The Intelligibility of Interrupted Speech. *J. of the Acoustical Soc. of Amer.*

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Miller, G. A., Galanter, E., and Pribram, K. H. (1971). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston.

Mine, M. R., Brooks Jr, F. P., and Sequin, C. H. (1997). 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.

Minsky, M. (1980). Telepresence. *Omni*, 2(9):45–52.

Monk, A. and Gale, C. (2002). A look is worth a thousand words: full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3):257–278.

Muller, M. J. (2003). Participatory design: The third space in HCI. *Human-computer interaction: Development process*, pages 165–185.

Muller, M. J. and Kuhn, S. (1993). Participatory design. *Commun. ACM*, 36(6):24–28.

National Council on Disability (2003). OLMSTEAD: Reclaiming Institutionalized Lives.

National Museum Australia (2014). Robot tours. Meet the Museum robots. `http://www.nma.gov.au/engage-learn/robot-tours`. Accessed Mar. 2014.

Newell, A. F. and Gregor, P. (2000). "User sensitive inclusive design" – In search of a new paradigm. In *Proceedings on the 2000 conference on Universal Usability*, pages 39–44. ACM.

News Channel 25 (2011). Baylor Using Robot to Aid School Districts. `http://www.kxxv.com/story/16056909/baylor-using-robot-to-aid-school-districts`. Accessed February 14, 2012.

Nguyen, H. (2012). Health care robots for the home: Why kids and parents prefer "embodied" robots. `http://vectorblog.org/2012/01/health-care-robots-for-the-home-why-kids-and-parents-prefer-embodied-robots`. Accessed Nov. 2013.

Nielsen, J. (1993). *Usability Engineering*. Chestnut Hill, MA: Academic Press.

Nielsen, J. (1994a). Enhancing the Explanatory Power of Usability Heuristics. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 152–158.

Nielsen, J. (1994b). *Usability engineering*. Elsevier.

Nielsen, J. (1995). 10 usability heuristics for user interface design. `http://www.nngroup.com/articles/ten-usability-heuristics`. Accessed Apr. 2014.

Nielsen Norman Group Report (2001). Beyond ALT Text: Making the Web Easy to Use for Users with Disabilities. `http://www.nngroup.com/reports/accessibility`, accessed October 2009.

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

Norman, D. A. (2002). *The design of everyday things*. Basic books.

Nunnaly, J. C. (1978). Psychocometric theory.

Obrenovic, Z., Abascal, J., and Starcevic, D. (2007). Universal accessibility as a multimodal design issue. *Commun. ACM*, 50(5):83–88.

O'Conaill, B., Whittaker, S., and Wilbur, S. (1993). Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human-Computer Interaction*, 8(4):389–428.

O'Gorman, L. (2010). Latency in Speech Feature Analysis for Telepresence Event Coding. In *Intl. Conf. on Pattern Recognition (ICPR)*, pages 4464–4467. IEEE.

Orbis Robotics (2014). Telepresence robots for factory support. `http://www.orbisrobotic.com/robots/cngp`. Accessed Mar. 2014.

Osada, J., Ohnaka, S., and Sato, M. (2006). The Scenario and Design Process of Childcare Eobot, PaPeRo. In *Proceedings of the 2006 ACM SIGCHI Intl. Conf. on Advances in Computer Entertainment Technology*. ACM.

Otsuka, K., Yamato, J., Takemae, Y., and Murase, H. (2006). Quantifying interpersonal influence in face-to-face conversations based on visual attention patterns. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 1175–1180. ACM.

Parikh, S., Grassi Jr, V., Kumar, V., and Okamoto Jr, J. (2005). Usability study of a control framework for an intelligent wheelchair. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 4745–4750. IEEE.

Park, J. and Kuipers, B. (2011). A Smooth Control Law for Graceful Motion of Differential Wheeled Mobile Robots in 2D Environment. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 4896–4902. IEEE.

Parlitz, C., Hägele, M., Klein, P., Seifert, J., and Dautenhahn, K. (2008). Care-O-bot 3 – Rationale for Human-Robot Interaction Design. In *Intl. Symp. on Robotics (ISR)*, pages 275–280.

Parmar, A. (2012). Children's Hospital Boston Testing Three Innovative Telemedicine Pilots. http://medcitynews.com/2012/07/childrens-hospital-boston-testing-three-innovative-telemedicine-pilots/. Accessed July 26, 2012.

Paulos, E. and Canny, J. (1998). PRoP: Personal roving presence. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 296–303. ACM Press/Addison-Wesley Publishing Co.

Paxton, K. (2011). Homebound Student Sends Robot to School. http://wcfcourier.com/lifestyles/homebound-student-sends-robot-to-school/article_d1185574-d781-5653-9e0f-68f45c71f655.html. Accessed February 14, 2012.

Perse, E. (2009). *Presence Questionnaire*, pages 276–283. Routledge, Taylor & Francis.

Pivtoraiko, M., Knepper, R. A., and Kelly, A. (2009). Differentially Constrained Mobile Robot Motion Planning In State Lattices. *Journal of Field Robotics*, 26(3):308–333.

Pollack, M., Brown, L., Colbry, D., Orosz, C., Peintner, B., Ramakrishnan, S., Engberg, S., Matthews, J., Dunbar-Jacob, J., McCarthy, C., Thrun, S., Montemerlo, M., Pineau, J., and Roy, N. (2002). Pearl: A Mobile Robotic Assistant for the Elderly. In *AAAI Workshop on Automation as Eldercare.*

Quigley, M., Gerkey, B., Watts, K., and Gassend, B. (2012). amcl - ROS Wiki. http://ros.org/wiki/joy. Accessed August 8, 2012.

Rabaud, V. (2012). openni_launch - ROS Wiki. http://www.ros.org/wiki/openni_launch. Accessed August 8, 2012.

Rbot (2012). R.BOT 100. http://promo.rbot.ru/?lang=en. Accessed July 28, 2012.

Reece, K. (2013). Robot named 'Blue Deuce' links 9-year-old to teachers, classmates. http://www.khou.com/news/local/Texas-student-sends-robot-to-school-in-his-place-229036671.html. Accessed Apr. 2014.

Region 6 Education Service Center (2014a). Millie is a exhibit docent! http://ivc.esc6.net/?p=835. Accessed Apr. 2014.

Region 6 Education Service Center (2014b). Morgan's angels. http://ivc.esc6.net/?page_id=35. Accessed Apr. 2014.

Reiser, U., Connette, C., Fischer, J., Kubacki, J., Bubeck, A., Weisshardt, F., Jacobs, T., Parlitz, C., Hagele, M., and Verl, A. (2009). Care-O-bot® 3 – Creating a Product Vision for Service Robot Applications by Integrating Design and Technology. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1992–1998. IEEE.

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

Richards, L. (2011). I, Student. Homebound Teen is Attending Classes via Wheel-bound Robot. Webpage. http://www.timesrecordnews.com/news/2011/jan/19/i-student/, accessed March 2011.

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

Riek, L. D. (2012). Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *J. of Human-Robot Interaction*, 1(1).

Riley, J. M., Kaber, D. B., and Draper, J. V. (2004). Situation awareness and attention allocation measures for quantifying telepresence experiences in teleoperation. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 14(1):51–67.

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

Robertson, R. (2011). Knox City Freshman Goes to Class. Webpage. `http://texomashomepage.com/search-fulltext?nxd_id=114136`, accessed March 2011.

Robins, B., Dautenhahn, K., and Dickerson, P. (2009). From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot. *Second International Conferences on Advances in Computer-Human Interactions*, pages 205–211.

RoboDynamics (2011). Introducing the World's First Commercial Telepresence Robot, TiLR. Webpage. `http://robodynamics.com/`, accessed March 2011.

ROS Answers forum (2013). Why navfn is using Dijkstra? `http://answers.ros.org/question/28366/why-navfn-is-using-dijkstra`. Accessed Apr. 2014.

Rosenberg, D. (2002). The "presence" of video. *Journal of Organizational Change Management*, 15(1):81–92.

Rosenberg, J. (2010). Quality Matters. `http://www.tmcnet.com/ucmag/columns/articles/99344-quality-matters.htm`, accessed Nov. 2011.

Rothenberg, S., Yoder, S., Kay, S., and Ponsky, T. (2009). Initial Experience with Surgical Telementoring in Pediatric Laparoscopic Surgery Using Remote Presence Technology. *J. of Laparoendoscopic & Advanced Surgical Techniques*, 19(S1):219–222.

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

Runyon, M. (2013). Valor and VGo carry St. Jude student into classrooms. `http://www.wauwatosanow.com/news/valor-and-vgo-carry-st-jude-student-into-classrooms-on93tof-197514421.html`. Accessed Apr. 2014.

Ryden, T. (2011). Practical Applications for Telepresence Robots. Keynote talk at IEEE Technologies for Practical Robot Applications (TePRA) conference.

Ryerson University (2011). PEBBLES. Webpage. `http://www.ryerson.ca/pebbles/index.html`, accessed March 2011.

Sabelli, A., Kanda, T., and Hagita, N. (2011). A Conversational Robot in an Elderly Care Center: An Ethnographic Study. In *Proc. of the 6th Intl. Conf. on Human-Robot Interaction*, pages 37–44. ACM.

Saenz, A. (2011). Another Robot Teacher Enters Korean Classrooms. `http://singularityhub.com/2011/02/26/another-robot-enters-korean-classrooms-as-automated-teaching-advances/`. Accessed July 24, 2012.

Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., and Hagita, N. (2007). Android as a Telecommunication Medium with a Human-like Presence. In *Proc. of Intl. Conf. on Human-Robot Interaction*, pages 193–200.

Sanders, T., Oleson, K., Billings, D., Chen, J., and Hancock, P. (2011). A model of human-robot trust theoretical model development. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1):1432–1436.

Saynor, J. (2001). Testimony of Janet Saynor, former Commissioner of Aging, CIty of New York, National Listening Session. PM testimony at 186–187.

Schaeffer, C. and May, T. (1999). Care-O-bot^{TM}: A System for Assisting Elderly or Disabled Persons in Home Environments. *Assistive Technology on the Threshold of the New Millenium*, pages 340–345.

Schulz, D., Burgard, W., Fox, D., Thrun, S., and Cremers, A. (2000). Web Interfaces for Mobile Robots in Public Places. *IEEE Robotics & Automation Magazine*, 7(1):48–56.

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

Sellen, A. (1995). Remote Conversations: The Effects of Mediating Talk with Technology. *Human-Computer Interaction*, 10(4):401–444.

Sheehy, K. and Green, A. A. (2011). Beaming children where they cannot go: telepresence robots and inclusive education: an exploratory study. *Ubiquitous Learning: an international journal*, 3(1):135–146.

Sheik-Nainar, M. A., Kaber, D. B., and Chow, M.-Y. (2005). Control gain adaptation in virtual reality mediated human–telerobot interaction. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 15(3):259–274.

Shibata, T., Mitsui, T., Wada, K., Touda, A., Kumasaka, T., Tagami, K., and Tanie, K. (2001). Mental Commit Robot and Its Application to Therapy of Children. In *Proc. of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, volume 2, pages 1053–1058. IEEE.

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

Simmons, R., Goodwin, R., Koenig, S., O'Sullivan, J., and Armstrong, G. (2002). Xavier: An Autonomous Mobile Robot on the Web. *Beyond Webcams: an introduction to online robots*, page 81.

Singer, E. (2010). The Slow Rise of the Robot Surgeon. MIT Technology Review. `http://www.technologyreview.com/biomedicine/24850/?a=f`, accessed Sept. 2010.

Singer, P. (2009). *Wired for War: The Robotics Revolution and Conflict in the Twenty-first Century*. Penguin Pr.

Sirkin, D., Venolia, G., Tang, J., Robertson, G., Kim, T., Inkpen, K., Sedlins, M., Lee, B., and Sinclair, M. (2011). Motion and Attention in a Kinetic Videoconferencing Proxy. *Human-Computer Interaction*, pages 162–180.

Skype (2006). Skype Introduces Video Calling for Macintosh Users. Press release. http://about.skype.com/2006/09/skype_introduces_video_calling.html, accessed Sept. 2010.

Skype S.A. (2011). Amendment No. 2 to Form S-1 Registration Statement. Prospectus. http://www.sec.gov/Archives/edgar/data/1498209/000119312511056174/ds1a.htm#rom83085_3a, accessed Nov. 2011.

Slater, M. (2005). Notes on presence. Virtual Environment course notes at the University College London. http://www0.cs.ucl.ac.uk/teaching/VE/Notes/presencenotes.pdf. Accessed Apr. 2014.

Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Presence*, 3(2):130–144.

Slater, M. and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616.

Sloan, K. (2012). Ailing Russian Teen Gets Robot Help. Associated Press. http://www.youtube.com/watch?v=ow4arKZIwXg. Accessed July 25, 2012.

Smith, C. and Skandalakis, J. (2005). Remote Presence Proctoring by Using a Wireless Remote-Control Videoconferencing System. *Surgical Innovation*, 12(2):139.

Special Education Technology (2011). pictureSET. Webpage. http://www.setbc.org/pictureset/, accessed March 2011.

Steeneken, H. (2006). Standardisation of performance criteria and assessments methods for speech communication. In *6th European Conf. on Speech Communication and Technology*, volume 1, pages 255–258.

Steinfeld, E. and Smith, R. (2012). Universal design for quality of life technologies. *Proceedings of the IEEE*, 100(8):2539–2554.

Stephanidis, C. (2001). User interfaces for all: New perspectives into human-computer interaction. *User Interfaces for All-Concepts, Methods, and Tools*, 1:3–17.

Stephanidis, C., Akoumianakis, D., Sfyrakis, M., and Paramythis, A. (1998). Universal accessibility in HCI: Process-oriented design guidelines and tool requirements. In *Proceedings of the 4th ERCIM Workshop on User Interfaces for all, Stockholm, Sweden*, pages 19–21.

Stockton, N. (2013). Artificial (emotional) intelligence. `http://scienceline.org/2013/01/artificial-emotional-intelligence`. Accessed Nov. 2013.

Stoia, L., Shockley, D., Byron, D., and Fosler-Lussier, E. (2008). SCARE: A situated corpus with annotated referring expressions. In *Proc. of LREC*, pages 28–30.

Suitable Technologies (2012). How beam works. `http://www.suitabletech.com/beam`. Accessed April 2014.

Suitable Technologies, Inc. (2013). Shelbot (aka Shelbot). `http://blog.suitabletech.com/2013/01/09/shelbot-aka-shel-bot`. Accessed Mar. 2014.

Suitable Technologies, Inc. (2014a). Beam+. `https://suitabletech.com/beam-plus`. Accessed Mar. 2014.

Suitable Technologies, Inc. (2014b). Beam pro. `https://www.suitabletech.com/beam`. Accessed Mar. 2014.

Suitable Technologies, Inc. (2014c). Introducing Beam+, pre-order now for $995. `http://blog.suitabletech.com/2014/01/15/introducing-beam-pre-order-now-for-995`. Accessed Mar. 2014.

Sulaiman, S., Saei, S. N. S. M., Mean, F. O., and Hasbullah, H. (2010). Understanding domain expert's perspectives and expectations in assistive technology. In *International Symposium in Information Technology (ITSim)*, volume 3, pages 1164–1167. IEEE.

Taggart, W., Turkle, S., and Kidd, C. (2005). An Interactive Robot in a Nursing Home: Preliminary Remarks. *Towards Social Mechanisms of Android Science*, pages 56–61. Stresa, Italy: Cognitive Science Society.

Takayama, L. (2011). Toward making robots invisible-in-use. *New Frontiers in Human-Robot Interaction*, 2.

Takayama, L. and Go, J. (2012). Mixing metaphors in mobile remote presence. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pages 495–504. ACM.

Takayama, L., Marder-Eppstein, E., Harris, H., and Beer, J. (2011). Assisted Driving of a Mobile Remote Presence System: System Design and Controlled User Evaluation. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1883–1889. IEEE.

Tejima, N. and Stefanov, D. (2005). Fail-safe components for rehabilitation robots-a reflex mechanism and fail-safe force sensor. In *9th International Conference on Rehabilitation Robotics (ICORR)*, pages 456–460. IEEE.

Telebotics (2005). PEBBLES. `http://www.telbotics.com/pebbles.htm`, accessed March 2011.

Thiel, S., Habe, D., and Block, M. (2010). Co-operative Robot Teams in a Hospital Environment. *VIMation Journal*, 1:13–20. Special Issue on Interactive Systems in Healthcare.

Thrun, S., Beetz, M., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Haehnel, D., Rosenberg, C., Roy, N., et al. (2000). Probabilistic Algorithms and the Interactive Museum Tour-Guide Robot Minerva. *The International Journal of Robotics Research*, 19(11):972–999.

Toris, R. (2013). The Standard ROS JavaScript Library. `http://wiki.ros.org/roslibjs`. Accessed Apr. 2014.

Trafton, J., Cassimatis, N., Bugajska, M., Brock, D., Mintz, F., and Schultz, A. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 35(4):460–470.

Trahanias, P., Argyros, A., Tsakiris, D., Cremers, A., Schulz, D., Burgard, W., Haehnel, D., Savvaides, V., Giannoulis, G., Coliou, M., et al. (2000). Tourbot-interactive museum tele-presence through robotic avatars. In *Proc. of the 9th International World Wide Web Conference*.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136.

Tsai, T., Hsu, Y., Ma, A., King, T., and Wu, C. (2007). Developing a Telepresence Robot for Interpersonal Communication with the Elderly in a Home Environment. *Telemedicine and e-Health*, 13(4):407–424.

Tsui, K., Desai, M., Yanco, H., and Uhlik, C. (2011a). Exploring Use Cases for Telepresence Robots. In *Proc. of 6^{th} ACM/IEEE Intl. Conf. on Human-Robot Interaction*.

Tsui, K., Norton, A., Brooks, D., McCann, E., Medvedev, M., and Yanco, H. (2013a). Design and development of two generations of semi-autonomous social telepresence robots. In *IEEE Conference on Technologies for Practical Robot Applications (TePRA)*.

Tsui, K., Norton, A., Brooks, D., Yanco, H., and Kontak, D. (2011b). Designing Telepresence Robot Systems for Use by People with Special Needs. In *Proc. of the Intl. Symp. on Quality of Life Technologies 2011: Intelligent Systems for Better Living, held in conjunction with RESNA 2011 as part of FICCDAT*.

Tsui, K. and Yanco, H. (2010). Prompting Devices: A Survey of Memory Aids for Task Sequencing. In *QoLT International Symposium: Intelligent Systems for Better Living, held in conjunction with RESNA*.

Tsui, K., Yanco, H., Kontak, D., and Beliveau, L. (2008). Development and evaluation of a flexible interface for a wheelchair mounted robotic arm. In *Proceedings of the Third Annual Conference on Human-Robot Interaction*.

Tsui, K. M. (2008). 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.

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

Tsui, K. M., Abu-Zahra, K., Casipe, R., M'Sadoques, J., and Drury, J. L. (2010). Developing Heuristics for Assistive Robotics. In *Proc. of the 5th Annual ACM/IEEE Intl. Conf. on Human-Robot Interaction (HRI)*. Late breaking paper.

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

Tsui, K. M., Desai, M., Yanco, H. A., and Uhlik, C. (2011c). Exploring use cases for telepresence robots. In *6th Annual Human-Robot Interaction Conference*, pages 11–18.

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

Tsui, K. M., Kim, D.-J., Behal, A., Kontak, D., and Yanco, H. A. (2011d). "I Want That": Human-in-the-Loop Control of a Wheelchair-Mounted Robotic Arm. *Journal of Applied Bionics and Biomechanics*, 8(1):127–147.

Tsui, K. M., Norton, A., Brooks, D. J., McCann, E., Medvedev, M. S., Allspaw, J., Suksawat, S., Dalphond, J. M., Lunderville, M., and Yanco, H. A. (2014).

Iterative design of a semi-autonomous social telepresence robot research platform: a chronology. *Intelligent Service Robotics*, 7(2):103–119.

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

Ulanoff, L. (2012). The iRobot Doc Will See You Now: Roomba Maker Enters the ER. `http://mashable.com/2012/07/24/irobot-healthcare-rp-vita/`. Accessed July 24, 2012.

Urrutia, K. (2012). Creating a CSS3 pulsating circle. `http://kevinurrutia.tumblr.com/post/16411271583/creating-a-css3-pulsating-circle`. Accessed Apr. 2014.

US Department of Education, Office of Special Education Programs (2004). Individualized Education Program (IEP).

U.S. Department of Justice (1994). ADA Standards for Accessible Design. `http://www.ada.gov/stdspdf.htm`. Accessed Nov. 2009.

US Equal Employment Opportunity Commission (2005). Work at Home/Telework as a Reasonable Accommodation. Webpage. `http://www.eeoc.gov/facts/telework.html`, accessed March 2011.

US Equal Employment Opportunity Commission (2011). Notice Concerning the Americans with Disabilities Avct (ADA) Amendments Act of 2008. Webpage. `http://www.eeoc.gov/laws/statutes/adaaa_notice.cfm`, accessed March 2011.

Van Baren, J., IJsselsteijn, W., Markopoulos, P., Romero, N., and de Ruyter, B. (2004). Measuring affective benefits and costs of awareness systems supporting intimate social networks. In *CTIT Workshop Proc. Series*, volume 2, pages 13–19.

van der Kleij, R., Maarten Schraagen, J., Werkhoven, P., and De Dreu, C. (2009). How conversations change over time in face-to-face and video-mediated communication. *Small Group Research*, 40(4):355–381.

Vanderheiden, G. and Vanderheiden, K. (1992). Guidelines for the Design of Consumer Products to Increase Their Accessibility to Persons with Disabilities or Who Are Aging. http://trace.wisc.edu/docs/consumer_product_guidelines/toc.htm, accessed October 2009.

Venkatesh, V. and Brown, S. (2001). A Longitudinal Investigation of Personal Computers in Homes: Adoption Determinants and Emerging Challenges. *MIS Quarterly*, pages 71–102.

Venolia, G., Tang, J., Cervantes, R., Bly, S., Robertson, G., Lee, B., and Inkpen, K. (2010). Embodied Social Proxy: Mediating Interpersonal Connection in Hub-and-Satellite Teams. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 1049–1058. ACM.

Verizon (2012). Making Education Accessible to Those Who Need It Most. http://responsibility.verizon.com/education/vgo. Accessed July 25, 2012.

Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 301–308. ACM.

Vertegaal, R., van der Veer, G., and Vons, H. (2000). Effects of gaze on multiparty mediated communication. In *Graphics Interface*, pages 95–102.

Vespa, P. (2005). Multimodality Monitoring and Telemonitoring in Neurocritical Care: From Microdialysis to Robotic telepresence. *Current Opinion in Critical Care*, 11(2):133.

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

Vess, J. (2014). Robots help link homebound students to school. http://www.kvue.com/news/Robots-helping-link-homebound-students-to-school-254904691.html. Accessed Apr. 2014.

VGo Communications (2011). Introducing VGo. From anywhere. Go anywhere. Webpage. http://vgocom.com/, accessed March 2011.

VGo Communications (2012a). Frequently Asked Questions | VGo robotic telepresence for healthcare, education and business. Webpage. http://www.vgocom.com/faq. Accessed August 10, 2012.

VGo Communications (2012b). Software Releases | VGo robotic telepresence for healthcare, education and business. Webpage. http://www.vgocom.com/software-releases. Accessed August 10, 2012.

VGo Communications (2012c). VGo User Guide v1.5.0. Manual. http://www.vgocom.com/sites/default/files/vgo_user_guide_v1.5.0.pdf. Accessed August 10, 2012.

VGo Communications (2014). Homebound student case study & ROI. http://www.vgocom.com/homebound-remote-student-case-study-roi. Accessed Apr. 2014.

Vincent, G. K. and Velkoff, V. A. (2010). The Next Four Decades: The Older Population in the United States: 2010 to 2050 – Population Estimates and Projections. P25-1138. http://www.census.gov/prod/2010pubs/p25-1138.pdf. Accessed July 26, 2012.

Virginia Assistive Technology System (2009). Assistive Technology in the Student's Individualized Education Program: A Handbook for Parents and School Personnel.

Viswanathan, P., Boger, J., Hoey, J., Elinas, P., and Mihailidis, A. (2007). The future of wheelchairs: Intelligent collision avoidance and navigation assistance. *Geriatrics and Aging*, 10(4):253–256.

W3C (1994). User Agent Accessibility Guidelines (UAAG) 2.0. http://www.w3.org/TR/UAAG20/#Guidelines. Accessed Apr. 2014.

W3C (2008). Web Content Accessibility Guidelines (WCAG) 2.0. http://www.w3.org/TR/WCAG20/. Accessed Apr. 2014.

Wada, K., Shibata, T., Saito, T., Sakamoto, K., and Tanie, K. (2005). Psychological and Social Effects of One Year Robot Assisted Activity on Elderly People at a Health Service Facility for the Aged. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2785–2790. IEEE.

Wagstaff, K. (2012). New Advances in Robotics Help the Disabled See the World Around Them. Time. http://techland.time.com/2012/05/02/new-advances-in-robotics-help-the-disabled-see-the-world-around-them/. Accessed July 25, 2012.

Washington, F. (2014). 'Keithbot' allows homebound kennard isd student to attend class from living room. http://www.ktre.com/story/24484454/keithbot-allows-homebound-kennard-isd-student-to-attend-class-from-living-room. Accessed Apr. 2014.

Weaver, R. (2012). New Castle Teen Uses Robot to Continue Learning from Home. Pittsburgh Tribune-Review.

Wiedemann, K. (2012). Robot Allows Edgewood Boy to Attend Class. http://www.kcrg.com/news/local/Robot-Allows-Edgewood-Boy-To-Attend-Class-139417063.html. Accessed July 25, 2012.

Willow Garage (2011a). Texai design principles. http://www.willowgarage.com/pages/texai/principles. Accessed Apr. 2014.

Willow Garage (2011b). Texai Remote Presence System. Webpage. http://www.willowgarage.com/pages/texai/overview, accessed March 2011.

Willow Garage (2012). ROS Fuerte Turtle. `http://ros.org/wiki/fuerte`. Accessed July 31, 2012.

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

Winters, R. (2012). Utah Teens Help Send Robot to Primary Children's Medical Center. The Salt Lake Tribune.

Witmer, B., Jerome, C., and Singer, M. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 14(3):298–312.

Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., and Froehlich, J. (2011). Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)*, 3(3):9.

Woods, S., Dautenhahn, K., and Schulz, J. (2004). The Design Space of Robots: Investigating Children's Views. In *Intl. ROMAN Wkshp.*, pages 47–52. IEEE.

Yanco, H., Drury, J., and Scholtz, J. (2004). Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Human-Computer Interaction*, 19(1):117–149.

Yanco, H. A., Keyes, B., Drury, J. L., Nielsen, C. W., Few, D. A., and Bruemmer, D. J. (2007a). Evolving interface design for robot search tasks. *Journal of Field Robotics*, 24:779–799.

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

Yarosh, S. and Markopoulos, P. (2010). 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.

Appendix A

The Inner Workings of Margo

A.1 VGo App User Interface

VGo App is VGo Communications' video conferencing software.It supports both robots calls (i.e., from a laptop/desktop computer to a robot) and also desktop calls (i.e. between two laptops/desktop computers). VGo App is available for Windows 7/Vista/XP with SP3 or MacOS 10.6.x or higher [VGo Communications, 2012b, c]. The robot call user interface is a sophisticated, video-centric user interface; the user interface is primarily a view of the robot's live camera stream. A small video of the robot driver is shown to the top right; this local video can also be disabled. There are a number of icons and buttons located below the video screen, which we subsequently describe.

Status information about the call and the robot is primarily located above and below the driver's video, as shown in Figure A-1. The VGo App's window title shows the name of the VGo robot and the duration, or elapsed time, of the current call [VGo Communications, 2012c]. Sensor status messages are shown in yellow on a semi-transparent black bar at the top of the video screen. For example, the two forward facing IR distance sensors on the left and right sides are angled slightly downward and provide cliff detection or step down information; the sensor message area text will read "Right/Left/Rear Cliff" Another one faces upward to detect tables; a similar status message appears with the text "Tall Object." When the front bumper

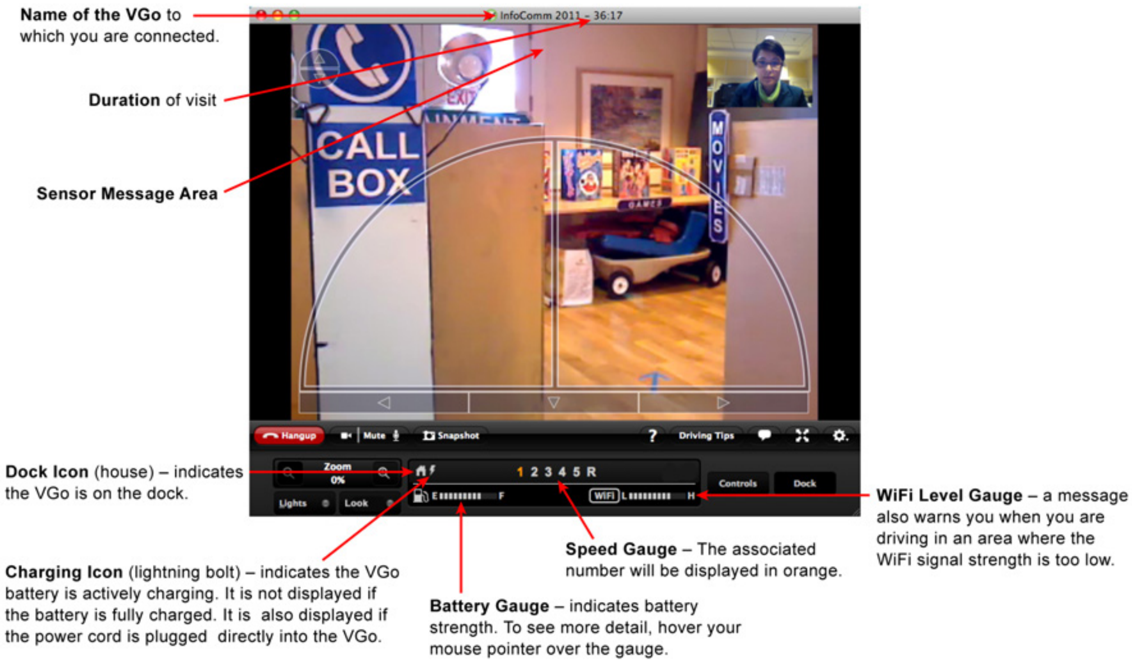


Figure A-1: VGo App user interface showing the location of call and robot status information. Screenshot as of Aug. 2012.

depresses a switch on the left and/or right side of the robot's base, a yellow starburst icon with the text "Bump" in black is shown in the center of the video screen.

The dock icon, a house, appears only if the VGo robot is located on docking station [VGo Communications, 2012c]. The charging icon appears only if the VGo battery is being charged or the VGo is directly plugged into the wall charger. The battery gauge indicates the level of battery remaining; additional information is provided when hovered over with a mouse pointer [VGo Communications, 2012c]. Similarly, the WiFi gauge indicates the WiFi signal strength. If the robot approaches the edge of a WiFi area, a semi-transparent, grey box appears in the center of the video screen with white text saying "Searching for WiFi." The speed gauge is similar to showing the gear progression in a car. When the robot is stopped the speed gauge is shown in a grey tone. As the robot begins to move, the associated number is highlighted in orange; first gear has a low speed, fifth gear has a high speed, and "R" indicates reverse.

Figure A-2 highlights the call related functionality and settings adjustments. From

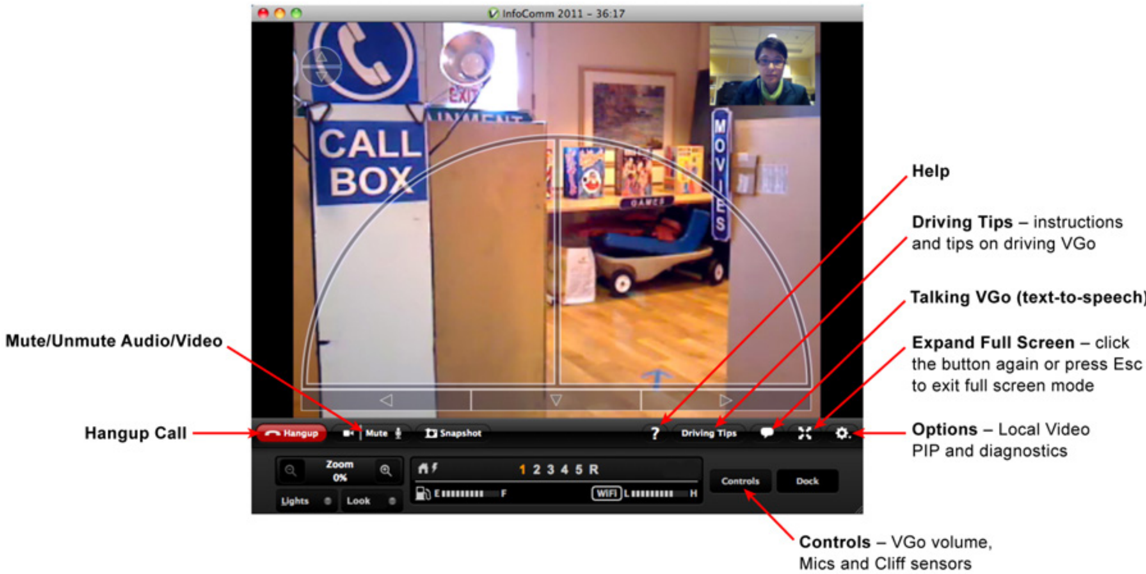


Figure A-2: VGo App user interface showing call related functionality and settings adjustments. Screenshot as of Aug. 2012.

an active call, the robot operator has the ability to independently end the call by pressing the red "Hangup" button. The robot operator can also toggle the mute button to suspend his or her audio only or audio and video together [VGo Communications, 2012b, c]. When the button is toggled to mute, the label is highlighted in orange and the robot's screen is covered with a semi-transparent black overlay with white text saying "Call muted" [VGo Communications, 2012c]. In addition to audio and video communication, the robot driver can type into a chat window and the robot will "speak" the text; the chat window appears when the "Talking VGo" button with a word bubble icon is toggled on.

Pressing the "Control" button launches a menu in which the robot operator can adjust the remote robot's speaker volume. Also, the robot operator can select between two microphone settings [VGo Communications, 2012c]. The first option is to use all four of the robot's microphone for omni-directional sound; the second option uses them only for the forward facing microphone, which is recommended for noisy environments. Finally, the robot operator can enable or disable the robot's ability to prevent itself from driving off an edge (e.g., driving down a flight of stairs). When this option is enabled, a sensor status message will appear at the top of the video screen, as

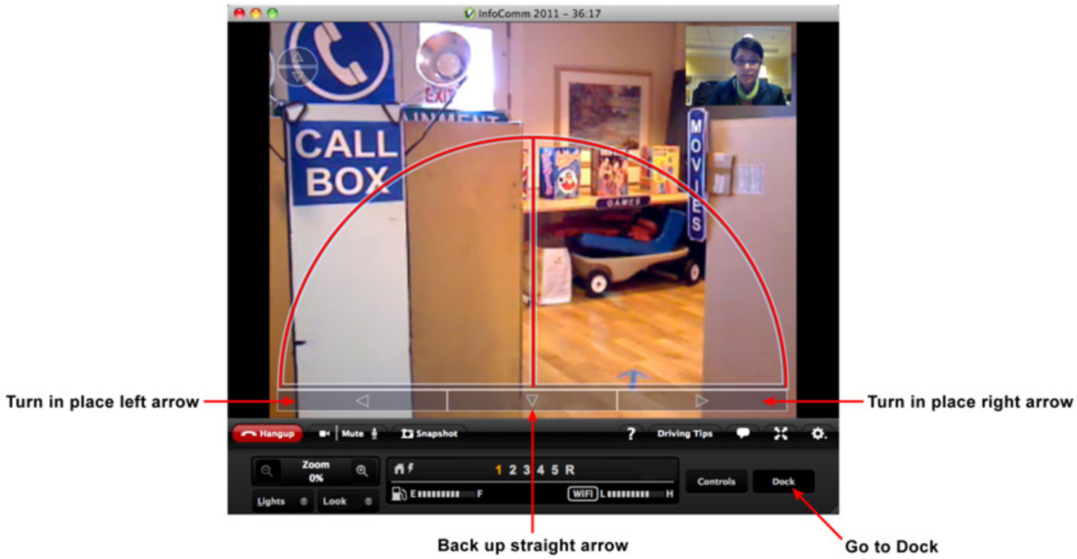


Figure A-3: VGo App user interface showing on-screen navigation methods. Screenshot as of Aug. 2012.

previously described.

The VGo robot can be driven via teleoperation using several methods. First, it can be driven using arrow keys to move the robot forward, back, left, or right. The robot's velocity is scaled over time by the the length of key press; that is, when pressing the "up" arrow, the robot will move forward slowly at first and will increase its velocity the longer the "up" arrow remains pressed [VGo Communications, 2012c]. The robot can also be driven using a proportional joystick widget; this functionality is in beta. Finally, the robot driver can use a mouse to indicate a "Click and Go" velocity based on the the angle and magnitude of the distance from the center point. The overlay, highlighted in red in Figure A-3, only appears when a mouse cursor moves across the video screen; it is otherwise hidden. The overlay also has buttons to turn the robot left and right in place, and a button for the robot to back up straight.

The VGo robot has the ability to autonomously return to its docking station. The "Dock" button will turn green when robot is within range of the docking station, which is approximately 10 feet (3 m) [VGo Communications, 2012c]. The button flashes green while the robot attempts to dock and will automatically re-dock if unsuccessful [VGo Communications, 2012b]. If successful, the dock icon (a house) should appear

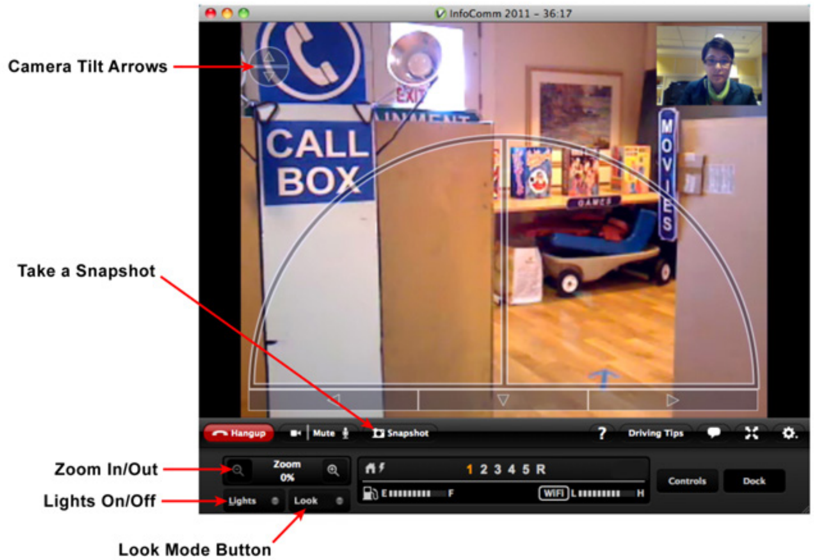


Figure A-4: VGo App user interface showing camera controls. Screenshot as of Aug. 2012.

above the fuel gauge, and the charging icon (a lightning bolt) should appear as the battery charges, as previously noted in Figure A-1.

The VGo robot features a forward facing camera which can be tilted 180 degrees up and down by scrolling a mouse wheel or clicking the "Camera Tilt Arrows," shown in the top left of the video screen in Figure A-4 [VGo Communications, 2012c]. The camera does not have an independent pan mechanism. However, when "Look Mode" is enabled, the robot base will pan and the camera will tilt automatically.[1] The velocity is set from the angle and magnitude a left mouse press and hold originating from the center of the video screen. The array of white LEDs surrounding the camera can be toggled using the "Lights" button. For the "Look" and "Lights" buttons, a small green circular "light" is shown within the button when its functionality is enabled. The camera can be digitally zoomed to 5x without distortion, and the zoom level is reflected in the text below the "Zoom" label [VGo Communications, 2012b]. Finally, the "Snapshot" button with the camera icon will save the robot's current view as a `jpg` to the robot operator's computer [VGo Communications, 2012c].

The camera can also auto-tilt when driving the robot [VGo Communications,

[1]Note that driving is disabled when "Look Mode" is enabled.

2012c]. When driving at slower velocities, the camera will tilt down and provide a view of the area directly around the base of the robot. The camera will tilt up as the velocity increases to show the forward facing view. The camera does not automatically tilt when driving backward.

A.2 VGo Remote

The VGo remote control activates user interface functions on the robot's head: answering and hanging up a call, tilting the VGo's camera up and down, turning the robot's volume up and down, muting the volume, and taking a picture. A USB IguanaWorks IR transceiver is used to emulate the VGo's IR remote control using the Linux Infrared Remote Control (LIRC) package [Bartelmus, 2011]. From our `lirc_vgo_remote` node, we provide rosservices for each of remote control functions: `/vgo/CallAnswer`, `/vgo/CallHangUp`, `/vgo/CameraUp`, `vgo/CameraDown`, `/vgo/CameraScreenshot`, `/vgo/VolumeDown`, `/vgo/VolumeMute`, and `/vgo/VolumeUp`. We emulate a button press by executing `irsend` calls at the command line and specifying the corresponding hex code; see Figure A-5.

```
# vgoremote.conf
# ------------------------------------------------------------------
#
# this config file was automatically generated
# using lirc -0.8.6(emulation) on Tue Dec 28 22:28:13 2010
# devices being controlled by this remote: VGo robot
#

begin remote

  name  VGo
  bits           32
  flags SPACE_ENC|CONST_LENGTH
  eps            30
  aeps          100

  header       9033  4418
  one           574  1622
  zero          574   490
  ptrail        576
  gap        106072
  toggle_bit_mask 0x0

      begin codes
          menu                       0x768900FF
          mute                       0x768940BF
          robotForwardLeft           0x7689609F
          robotForward               0x7689E01F
          robotForwardRight          0x768910EF
          robotLeft                  0x7689906F
          ok                         0x768950AF
          robotRight                 0x7689D02F
          back                       0x768930CF
          robotBackward              0x7689B04F
          cameraTilt+                0x7689F00F
          cameraTilt-                0x768948B7
          vol+                       0x76898877
          vol-                       0x768928D7
          answer                     0x7689A857
          hangUp                     0x7689E817
          snapshot                   0x7689D827
          swapView                   0x7689B847
          drive                      0x76897887
          dock                       0x7689F807
          help                       0x768904FB
      end codes

end remote
```

Figure A-5: Hex code listing of VGo's IR remote control buttons.

A.3 Telepresence Robot Design Guidelines

Tables A.1 and A.2 summarize principles, requirements, and design guidelines from four HRI research groups investigating mobile telepresence robots for social interaction, including our own:

- MITRE (see [Riek, 2007]),
- the ExCITE project (Enabling SoCial Interaction Through Embodiment; see [Cohen et al., 2011a]),
- University of Massachusetts Lowell (see [Desai et al., 2011; Tsui et al., 2011a, b]), and
- Willow Garage (see *Texai design principles*, [Willow Garage, 2011a]).

Table A.1: HRI recommendations for telepresence robots from [Tsui and Yanco, 2013]. (Part 1 of 2)

		[Willow Garage, 2012]	[Riek, 2007]	[Cohen et al., 2011]	[Desai, et al. 2011; Tsui et al., 2011a,c]
Scope		Principles of use	Physical and functional requirements	Requirements and design considerations	Design guidelines
Human roles supported		Pilot	Pilot	Pilot, local, bystander	Pilot, local, bystander
(1) Video	(General)	Reciprocity of Vision (if I see you, you must see me)		"The visitor's [pilot] environment should be immersive so that the pilot would have a first person experience of the destination including full sensory stimulation focusing on immersive vision, audio, and haptics." *(3,9)*	
	(Characteristics)		Prevent image distortion Prevent motion artifacts Preserve color Provide visual continuity during times of lag	Graphics frame-rate Low latency between pilot head movement and display update	Video profile for use while stationary Video profile for use while moving Video resolution sufficient to read ADA compliant text Graceful video degradation by increased compression
(2) Video camera	(Number)		One	Not specified	Two
	(Purpose)		General		Forward facing camera (dedicated camera for conversation) Downward facing camera (dedicated camera for navigation)
	(Field of view)		Camera provides views that closely mimic being physically present (e.g., wide angle or 360 degrees)	Large field of view	Wide field of fiew
	(Resolution)				High resolution camera
	(Articulation)		Permit full pan/tilt/zoom camera control	Not specified	Articulated head or torso (indirect) *(8)*
(3) Audio	(General)			"The visitor's [pilot] environment should be immersive so that the pilot would have a first person experience of the destination including full sensory stimulation focusing on immersive vision, audio, and haptics." *(1,9)* Local should be able to hear the pilot from his/her representation's location and orientation Capture pilot's actions (gestures, speech, interactions)	Audio quality of at least a landline telephone
	(Characteristics)		Provide background-noise detection	Spatial hearing Sound localization	Continuous audio (no cut out) Noise filtering Echo cancellation
	(Volume control)				Volume control by the robot pilot Volume control by the person in the remote environment Automatic volume adjustment based on ambient noise level and distance of remote interactant from robot Appropriate audible occupancy awareness
(4) Wireless communication support	(General)				Access point switching
(5) Mobility	(General)		Permit full mobility	Pilot should be able to control his/her representation at the destination and the view should change according to the position [of the robot]	Walking speed of person
	(Latency)		Minimize bandwidth latency (of teleoperation commands) to be less than 125 ms		Stop transmitting robot commands when unable to transmit video *(13)*
	(Autonomy)			Representation should have autonomous mobility Go there Follow user Identify obstacles, avoid collisions, and report this event to the pilot *(13)*	Assisted navigation *(13)* Follow person (human speed navigation) Go to destination (human speed navigation) Ability to switch between autonomous behaviors
(6) Embodiment	(General)			Local should feel that the pilot representation represents a real person Pilot should have a robotic representation represents all human physical and emotional states Occupy a physical space Appear in real size	
	(Height)		Permit height control	Representation should be able to sit Representation should be able to stand	Height for interacting at eye-level with remote locals standing Height for interacting at eye-level with remote locals sitting
(7) Gesture / Manipulation	(General)		Provide at minimum a two degree-of-freedom mechanism for deictic gesture Ensure shared perspective	Representation should display body language (head movement, eye gaze, facial expression, arm movement, hand movement, finger movement) *(8)* Support gesture and deixis Shake hands with locals at the destination Pick up and move simple objects Perform simple operations in the environment (open a door, press elevator button)	

Table A.2: HRI recommendations for telepresence robots from [Tsui and Yanco, 2013]. (Part 2 of 2)

		[Willow Garage, 2012]	[Riek, 2007]	[Cohen et al., 2011]	[Desai, et al. 2011; Tsui et al., 2011a,c]
(8) Gaze	**(General)**			Pilot should be able to transfer his/her expressions in an unconscious manner	Articulated head or torso *(2)*
	(Facial expression)		Portray clear facial appearance and expression	Representation should display body language (head movement, eye gaze, facial expression, arm movement, hand movement, finger movement) *(7)*	
	(Eye contact)		Preserve gaze	Representation should display body language (head movement, eye gaze, facial expression, arm movement, hand movement, finger movement) *(7)* Pilot should be able to make eye contact with locals at the destination Locals should recognize representation's gaze direction, gestures, and facial expressions.	
(9) Pilot interface	**(General)**			"The visitor's [pilot] environment should be immersive so that the pilot would have a first person experience of the destination including full sensory stimulation focusing on immersive vision, audio, and haptics." *(1,3)* Pilot should be immersed in the destination without feeling headaches or nausea Pilot environment should be safe to use by the pilot	
	(Characteristics)			Isolated environment Capture pilot's actions (gestures, speech, interactions) Move head and see corresponding view of the destination Ability to stop transmission of pilot's actions Ability for pilot to consciously switch between transmitting mode and non-transmitting User invitation and linking to a representation Status of representation (not in use) *(11)* Search for user's events	Platform independent user interface Web-based user interface Provide relevant and accurate sensor information Provide feedback that reduces latency Integrated map of remote environment with "robot is here" indicator
(10) Ease of use	**(Use status)**	Transparency of technology (the pilot is the focus)		Pilot should be able to focus on the communication and other tasks, not on operating the system. System delays should be imperceptible to the pilot Locals should not require equipment or training in order to perceive or interact with the pilot	
(11) Awareness	**(Use status)**			Status of representation (not in use) *(9)*	Visual indicators for when robot is occupied Visual indicators for other robot states
	(Pilot identification)			Locals should know who the pilot is (recognition/identification) Locals should be aware the representation is being operated or inactive Identification and authentication of pilot for access *(12)*	Pilot identification beyond picture and voice
(12) Privacy	**(General)**	Ensuring private communication (no onboard recording of audio/video)		Session can be recorded and saved by pilot and local for later viewing Stop session recording ("off the record") Knowledge that the session is recorded and saved Identification and authentication of pilot for access *(11)* Locals should know when a session is recorded	
(13) Safety	**(General)**			Representation should be safe for interaction with locals	Stop transmitting robot commands when unable to transmit video *(5)*
	(Mobility)			Avoid collisions and any possible harm to the destination or locals or bystanders there Identify obstacles, avoid collisions, and report this event to the pilot *(5)*	Assisted navigation *(5)*
(14) Proxemics	**(General)**	Respect the social norms (the Texai is the pilot in the remote location)			
(15) Virtual content	**(General)**			Pilot should be able to share and present the digital content Pilot and locals should be able to manipulate the digital content	

A.4 Margo's COTS Components

As per Requirement 6 in Table 3.2, additional components had to be commercial off-the-shelf (COTS) components and well supported by the robotics community. Table A.3 details the COTS components utilized in Margo.

Table A.3: Margo's commercial off the shelf (COTS) components

	Function	Selected Components	Manufacturer Information
Base	Robot	VGo Communications' VGo	vgocom.com
	VGo head/base serial communication	FTDI USB-RS422 adapters	ftdichip.com/Support/Documents/DataSheets/Cables/DS_USB_RS422_PCB.pdf
	VGo remote control	IguanaWorks' USB IR transceiver	iguanaworks.net/products/usb-ir-transceiver/
Computing	Computer	fitPC-2	fit-pc.com/web/fit-pc/fit-pc2-specifications/
	Heatsink	fit-PC2 heatsink	fit-pc.com/web/fit-pc/accessories/
	Cooling fan	Evercool Fan EC8010LL05E	evercoolusa.com/?p=1422
	Display	PhidgetTextLCD Adapter LCD Screen	phidgets.com/products.php?product_id=1204 phidgets.com/products.php?category=15&product_id=3650
	USB hub	CyberPower 4-port 2.0 hub	cyberpowersystems.com/products/accessories/usb_hubs/cph420p.html
Sensing	IMU	Microstrain 3DM-GX3-45	microstrain.com/3dm-gx3-25.aspx
	Laser	Hokuyo UGH-08	hokuyo-aut.jp/02sensor/07scanner/download/data/UHG-08LX_spec.pdf
	Rear distance sensing	Sharp 2Y0A02 infrared (IR) sensors Phidget IR distance adapter	phidgets.com/products.php?category=2&product_id=3522_0 phidgets.com/products.php?category=2&product_id=1101
	Bystander interaction	Microsoft Kinect	microsoftstore.com/store/msstore/en_US/pd/Kinect-for-Xbox-360/productID.216507400
	Downward facing camera	Logitech C910	logitech.com/en-us/support/webcams/hd-pro-webcam-c910
	Sensor I/O board	Phidget InterfaceKit 1019	phidgets.com/products.php?product_id=1019_1
Power System	12V battery	CSB EVH12150F2 sealed lead acid battery 12V 15Ah	csb-battery.com/upfiles/dow01320126580.pdf
	12V power supply for laser and Kinect	Minibox DCDC-USB	mini-box.com/DCDC-USB
	9V power supply for fitPC and Phidget InterfaceKit 1019	25W 3A step down adjustable switching regulator DE-SWADJ 3	dimensionengineering.com/datasheets/DE-SWADJ3.pdf
	5V power supply for CyberPower hub, cooling fan	25W 3A step down adjustable switching regulator DE-SWADJ 3	dimensionengineering.com/datasheets/DE-SWADJ3.pdf

Appendix B

Study 1: Scenario Development for Focus Group

Key Questions

Our research focuses on the human-robot interaction between a "personal transportation robot," its operator, and the bystanders physically located in the environment with the robot. We began with several key questions regarding how this collaboration looks such that the robot operator reaches his or her intended destination:

- How would you want to give instructions to the robot? How would you want to **change the robot's current direction**, e.g., to stop in a colleague's office on your way to the mail room? How would you expect the robot to be able to **resume its task** after you had finished talking to your colleague?
- How should the robot ask for help? How often? Are there times when the robot shouldn't bother you? **Should the robot interrupt you**?
- If you ask someone nearby for directions, the robot can also listen to what the person tells you. If the robot doesn't understand part of the directions, **should the robot ask you to ask for a clarification or should it ask the person itself**? How should the robot summarize its interpretation.

Our team collaboratively brainstormed additional questions to investigate specific aspects of the interaction and user preferences.

- Would it help if the commands you give to the robot were standardized (a set of specific words that you would use all the time), or could vary from situation to situation?
- Would you prefer to use a natural speaking style with the robot, or would single word commands be preferable? Would you like to be able to select the words used to control the robot from the words that make most sense to you? (personalized vocabulary)
- Would you like to be able to override the robot using a joystick? How should the robot react when you override it? Should it learn how you would like to move the wheelchair by observing you and then repeating what you did?
- If you had told the robot you needed to be at a specific place at a specific time, but you stopped to do something and the time was approaching after which the robot judged you could no longer make the deadline, should the robot announce that fact?
- If you have a schedule of places to be at certain times, would you like to tell the robot the entire schedule at once, or give it a new place as each time is approaching?
- If you gave the robot a list of places to go and the order was not important, would you prefer the robot to choose the most efficient order, or use the order in which they were listed? If the former, would you prefer that behavior to occur when you specified that the order was not important, or when you did not specify that the order was important?
- What things can the robot do to make you trust it more? Less? Are there things the robot should not do?

- Would you prefer to use headphones to keep the robot's "voice" audible only to you? Would you prefer to use a carefully positioned microphone near your mouth to minimize the amount of "conversation" with the robot that others could hear?
- Does it matter to you whether the robot's "voice" is male or female?
- Do you have prior experience with dictation software? Please describe.

Resulting Scenarios

We then framed these questions in terms of a series of scenarios for "personal transportation robots." We describe both our telepresence robot, Margo, and a robot wheelchair as two forms of "personal transportation robots." The telepresence robot is a physical entity in a remote space which provides an embodiment for a person's virtual self; the robot and the robot operator are separated. The robot wheelchair is collocated with its operator and both are physically present together.

The premise of the telepresence scenario is as follows:

> Imagine you are touring a new [day / long term / assisted living / rehabilitation] care center. You went to this center last year to visit with your friend. You currently are in the hospital, recovering from a major surgery, and are considering moving to this center for aftercare.
>
> The center has 10 telepresence robots that clients' friends and family can use to visit them and participate in their social activities, etc. The telepresence robots are located in the community living room, next to the community kitchen and several private dorm rooms.
>
> You are operating a telepresence robot and touring the center's campus: e.g., visiting private dorm rooms; therapy pool, gymnasium, the nurse station, community kitchen, social activities, etc.

1) Things I want to see today. Unordered list: Community garden space. Group cooking class at 1:00pm. Visit with your friend. Gymnasium (open

from 3:00pm to 6:00pm for general use). Watch a movie. Cafeteria (food options, coffee).

- How would you tell or show your robot what you wanted to do?
- How should the robot tell or show you that it understands?
- When would you tell or show the robot each event?
- Where would you expect the robot to go first? Next? (and so on)

2) Imagine your friend gives you a message that he or she has a visitor coming at 1:30 pm and would like you to join them for a walk on the Dutton Brooks Wetland Trail.[1] You will meet your friend and his or her visitor at the reception desk.

- How would you tell or show your robot about this new event?
- How should the robot tell or show you that it understands this?
- Should the robot tell or show you about the rest of your day? If so, how?
- Should the robot remind you that the gymnasium opens at 3:00pm? If so, when would you expect a reminder?

3) You see a staff member in the hallway. She tells you that, today there will be a group basketball game in the gymnasium, starting 30 minutes into open hours. She asks if you would like to watch basketball game. You say yes.

- Would you want the robot to also listen to your conversation with the staff member? Or should you tell or show the robot what you learned? How would you do this?
- If you wanted to cancel your physical therapy appointment, how would you do this? (remove event). How would you tell or show your robot about the basketball game? (add event)

[1] Dutton Brooks Wetland Trail is an accessible trail at Crotched Mountain.

- If you wanted to change your physical therapy appointment, how would you do this? (update event)

4) The Dutton Brooks Wetland Trail is brand new and you haven't been there yet.

- How would you find out how to get to the start of the trail?
- If you ask someone nearby for directions, would you want the robot to also listen? Or should you tell or show the robot what you learned?
- Describe how to get to the Dutton Brooks Wetland Trail from here.

5) It is 1:15pm and you meet your friend and his or her guest at the reception desk. On the way to the start of the Dutton Brooks Wetland Trail, you are told that your telepresence robot needs a special outdoor connector (e.g. mobile hotspot, 4G, cell phone) to go outside and the receptionist has these.

- How should the robot tell or show you about needing this outdoor connector?
- How would you go back to the reception desk? How else could this be done with the robot?
- How would you expect the robot to be able to resume its task after you got the outdoor connector?

6) You are back at the start of the Dutton Brooks Wetland Trail, but you cannot find your friend and his or her visitor.

- How would you instruct the robot? What would you do?

7) You find your friends halfway down the Dutton Brooks Wetland Trail. They are taking pictures. You join them and continue down the trail, talking with them.

- What should the robot do?

- How would you instruct the robot if you wanted to take a picture of a deer?
- How would you instruct the robot if you wanted to be in (pose for) a group picture with your friends?

8) The Dutton Brooks Wetland Trail is 1 mile long and you and your friends visited all of it. The battery is getting low and needs to be recharged. You have a battery swap station in at the reception desk at the main entrance, and also one near the community kitchen.

- How should the robot get your attention?
- How should the robot tell or show you this? What should the robot do?

9) You return from the trail and find that the entrance you came out of is blocked. The door is having its motion sensor and automatic opening system replaced. You can't use this entrance.

- What should the robot do? How would you instruct the robot?
- How should the robot ask for help? Should it talk to you? How often?
- In what specific situations would you want the robot to ask you for help? Are there times when the robot shouldn't bother you?

Appendix C

Study 2: Scavenger Hunt

C.1 Proctor Script

After the consent process, an experimenter read the following instructional script to participants in the robot agent condition:

> "For this experiment today, we are going to present that you are hosting a <INSERT THEME> party and will be choosing food and entertainment for your guests. [*Experimenter handed a theme cue card to the participant; Halloween, Christmas, circus, or robot*] You'll also pick out a costume for yourself. You will be taking pictures of everything you need for the <INSERT THEME> party at "KAS Party Central."
>
> "KAS Party Central" has shopping assistants and party planners. Your robot shopping assistant, Margo, will show you around the store. You can use your voice to talk to her and tell her where to go and what you are looking for. Margo will be showing you the store via webcam. The webcam also has 2-way audio, so you can hear what's going on at the store, and the robot can hear anything you say to it. One thing to keep in mind is that there is a few second delay between when you talk and when the robot hears you.
>
> Today's on-site party planner is Kate. She has over ten years of experience

including weddings, family reunions, and children's birthday parties. If you want a recommendation or help making a choice, visit the party planning station or find a call box.

For your <INSERT THEME> party, you will need to choose: food and drinks, a movie based on the recommendation of a party planner, a game for your party guests to play, and a costume for yourself.

Take a picture of each of your choices. When you are done, go to the party planning station and finalize your choices with a party planner. Then "KAS Party Central" will pack and ship your party-in-a-box to you, including complimentary party decorations and favors for your guests.

Here is the shopping list for you. [*Experimenter handed the checklist to the participant and read each item: 2 types of snacks, 1 drink, 1 movie (as recommended by a party planner), 1 party game, 1 costume.*]

And here is a directory of the store. [*Experimenter handed a store directory (Figure C-1) to the participant and showed the remote shopper's starting location and orientation, denoted by the star labelled "START." Experimenter also showed the location of the party planning station and the call box on the map.*] Margo is starting here, facing this wall. This is the call box, where you go it you need help from the party planner. Drive onto the call box and wait; the party planner will come over to help you. This is the party planner station and checkout, where you can also go for help and at the end to finalize your purchases.

Your robot shopping assistant, Margo, also has the store directory.

You will have 15 minutes to shop and 5 minutes to finalize your choices with the party planner. I will let you know when you have 10, 5, and 3 minutes remaining to shop. Do you have any questions before we begin? [*Experiment answered questions.*]

Ok. Let's go shopping. Margo, can you hear me? [*Experiment waited for acknowledgement from wizard.*]

> From now on, please talk to your robot shopping assistant, and not to me. You can now use your voice and talk to your robot shopping assistant and direct her around the store. Go ahead and start talking to Margo."

The instructions for the human-agent condition had minimal modifications. Lines 4–6 in paragraph 2 instead read as:

> "<NAME OF HUMAN ASSISTANT> will be showing you the store via webcam. The webcam also has 2-way audio, so you can hear what's going on at the store, and <HE or SHE> can hear anything you say to <HIM OR HER>. One thing to keep in mind is that there is a few second delay between when you talk and when the shopping assistant hears you."

We provided the subsequent clarification: "<NAME OF HUMAN ASSISTANT> is a person and will be pushing a webcam around on a tripod. We are working on building the robot." Finally, references to the shopping assistant omitted the word "robot."

C.2 Post-experiment Interview

The following post-experiment interview questions were verbally proctored to participants. "*THE REMOTE SHOPPER*" was specified as "the robot" for participants in the robot agent condition and "Kelsey" in the human agent condition.

1. What was your strategy for finding the items on the checklist? [**Alternative**: What order did you go shopping for the items on the checklist? Why? Can you tell me why you took pictures in the order that you did?]

2. What was your strategy for providing verbal instructions/directing the robot? [**Alternative**: How did you talk to *THE REMOTE SHOPPER*? Like it was an adult? Like it was a child? Like it was a pet? Did you give *THE REMOTE SHOPPER* commands, ask questions, say "please" and "thank you?" Why?]

3. Was your strategy the same at the end as it was at the beginning? If not, how did this change over time? [**Alternative**: Did the way you talked to *THE*

REMOTE SHOPPER change from when you first started to when you checked out with the party planner Kate? If so, how and why?]

4. What expectations did you have for *THE REMOTE SHOPPER*? What did you expect *THE REMOTE SHOPPER* to be able to do already? Did *THE REMOTE SHOPPER* meet those expectations? Did *THE REMOTE SHOPPER* exceed those expectations? [**Alternative**: What did you think *THE REMOTE SHOPPER* would be able to do? Was *THE REMOTE SHOPPER* able to do those things? Did *THE REMOTE SHOPPER* do more than all of those things?]

5. Was there anything that *THE REMOTE SHOPPER* did that made you trust the system more? Is there anything *THE REMOTE SHOPPER* should do in the future to make you trust the system more? [**Alternative**: Did *THE REMOTE SHOPPER* do anything that made you think "I can trust *THE REMOTE SHOPPER*?" If so, what?]

6. Was there anything that *THE REMOTE SHOPPER* did that made you feel less trust *THE REMOTE SHOPPER* less? Is there anything *THE REMOTE SHOPPER* should avoid doing in the future that would make you trust *THE REMOTE SHOPPER* less? [**Alternative**: Did *THE REMOTE SHOPPER* do anything that made you think "I should not trust *THE REMOTE SHOPPER*?" If so, what?]

7. Do you have any experience with voice recognition technology? Please describe.

    (a) (*Robot agent condition only*) How did it compare to today's experience?

8. Do you have any pets? Do you talk to your pets? If so, how?

Additional questions were asked by the experimenter based on the events of a participant's session and the responses to the above questions.

C.3 KAS Party Central

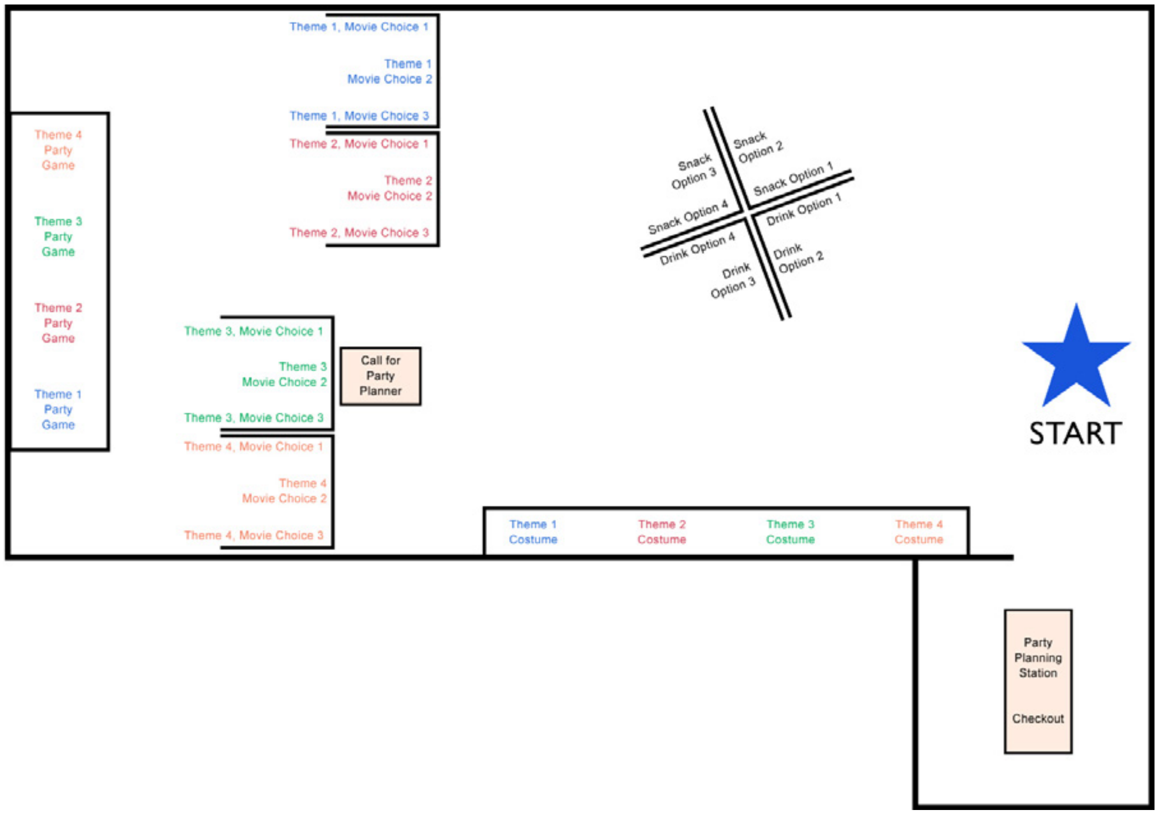
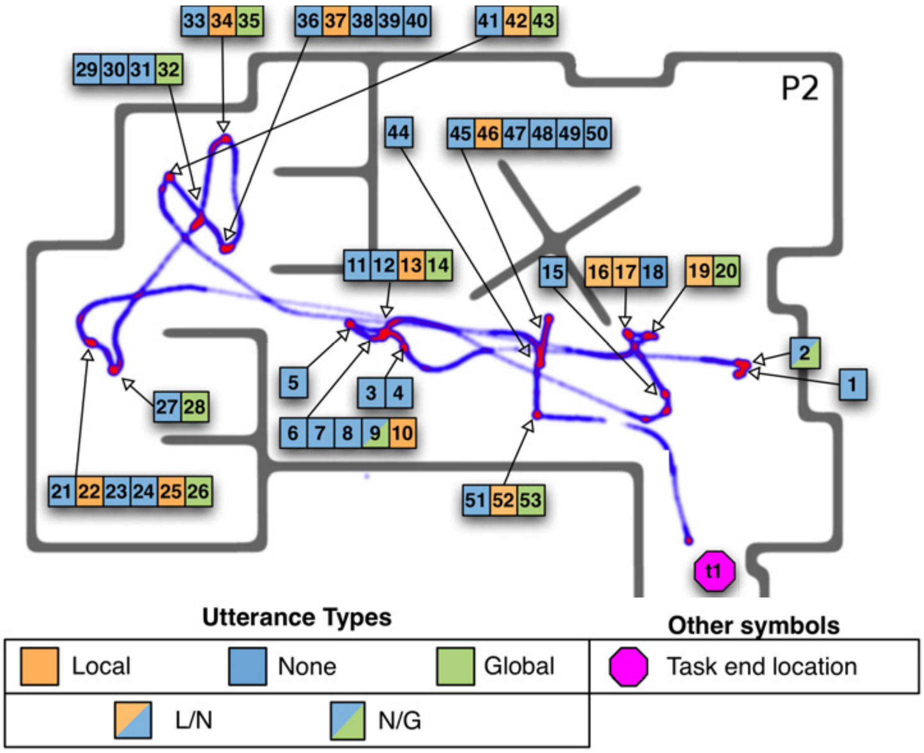


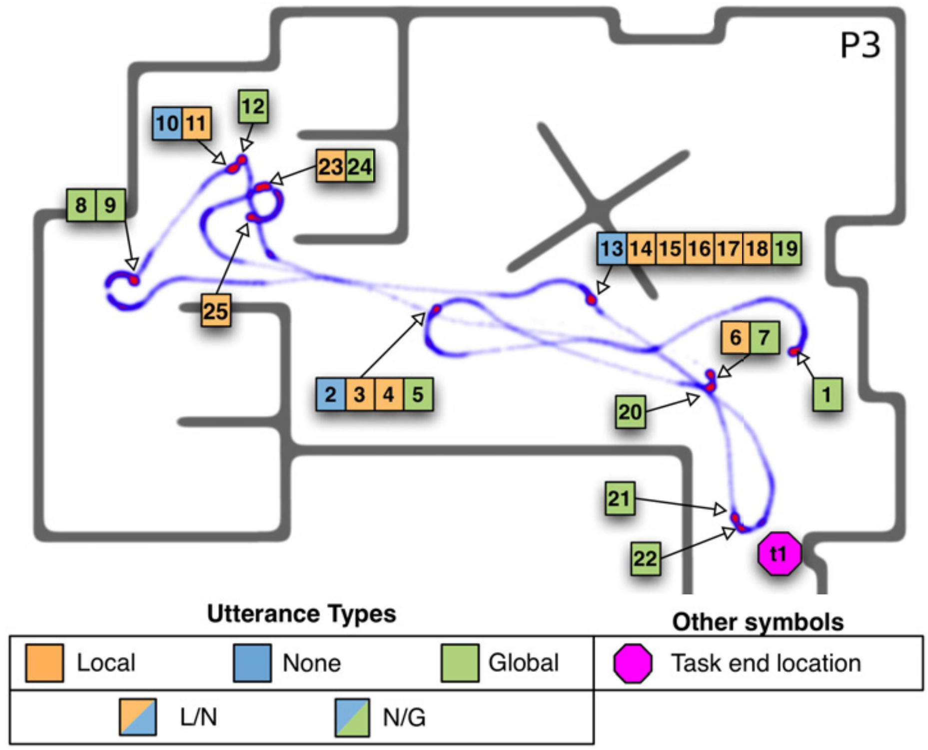
Figure C-1: Placement of specific themed items, snack options, and drink options in "KAS Party Central." The drink options were milk (option 1), fruit punch (2), orange soda (3), and water (4). The snack options were pretzels (option 1), apples (2), cupcakes (3), and cookies (4). We created four party themes (i.e., circus (theme 1), Christmas (2), Halloween (3), robot (4)). There was one choice for each theme's costume and party game, and three movie choices per theme.

C.4 Annotated Heatmaps



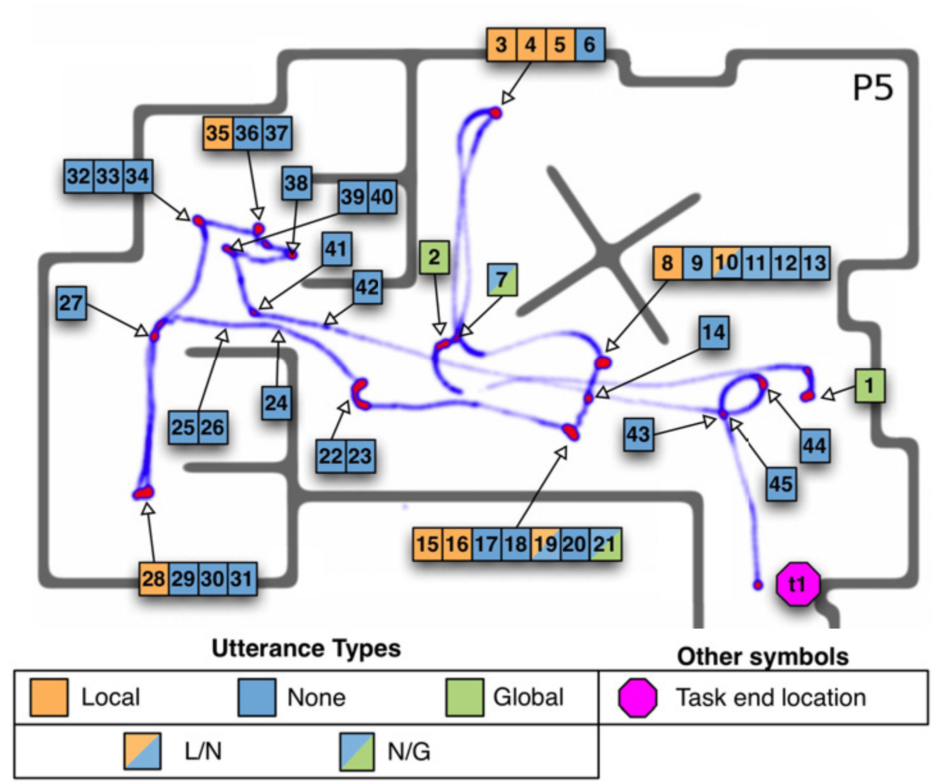
1	N	0:01	Margo, turn left 180 degrees.
2	L/N	0:18	Margo, move forward 10 feet and go to grocery.
3	N	0:44	Stop.
4	N	0:47	Margo, turn left.
5	N	0:55	Stop.
6	N	0:58	Margo, turn right two feet.
7	N	1:07	Stop.
8	N	1.09	Margo, turn left and move forward.
9	N/G	1:22	Margo, turn left and go into grocery.
10	L	1:52	Margo, choose cupcakes.
11	N	2:01	Margo, turn left.
12	N	2:05	Stop.
13	L	2:09	Margo, choose muffins.
14	G	2:17	Margo, go to drinks.
15	N	2:55	Move forward.
16	L	3:06	Margo, choose milk.
17	L	3:16	Margo, look at drinks.
18	N	3:23	Left.
19	L	3:30	Choose G2 drink.
20	G	3:37	Wonderful. Margo, go to movies.
21	N	4:33	Stop.
22	L	4:35	Margo, choose "Young Frankenstein."
23	N	4:44	Margo, move right.
24	N	4:47	Stop.
25	L	4:50	Margo, choose "Young Frankenstein" again.
26	G	4:58	Margo, go to other movies.
27	N	5:08	Stop.
28	G	5:13	Margo, go to other movies.
29	N	5:42	Margo, stop.
30	N	5:43	Move right.
31	N	5:47	Margo, stop.
32	G	5:53	Go to other movies, Margo.
33	N	6:13	Margo, stop.
34	L	6:15	Margo, choose "Dumbo."
35	G	6:22	Thank you, Margo. Margo, go to games.
36	N	6:44	Margo, stop.
37	L	6:45	Margo, choose "Rock 'em, Sock 'em Robots."
38	N	7:02	Margo, go right.
39	N	7:05	Margo, stop.
40	L	7:07	Margo, get closer to robots.
41	N	7:19	Margo, stop.
42	L	7:23	Choose robots.
43	G	7:30	Wonderful. Margo, go to costumes.
	–	7:45	She's doing a pretty good job, don't you think?
44	N	8:15	Margo, back up.
45	N	8:26	Margo, stop.
46	L	829	Margo, choose skeleton.
47	N	8:38	Margo, rise up.
48	N	8:45	Margo, tilt back.
49	N	8:51	Stop.
50	N	8:54	Margo, move closer.
51	N	9:07	Margo, stop.
52	L	9:10	Choose skeleton face.
53	G	9:15	Excellent. Margo, go to party planner station.
t1	–	9:50	***End of Task 1***

Figure C-2: Heatmap of the robot's trajectory labelled with P2's timestamped utterances (robot agent condition). P2 gave three times as many commands that did not require environmental knowledge (n=33) than global ones (n=11). P2 noted that he "spoke concisely and clearly."



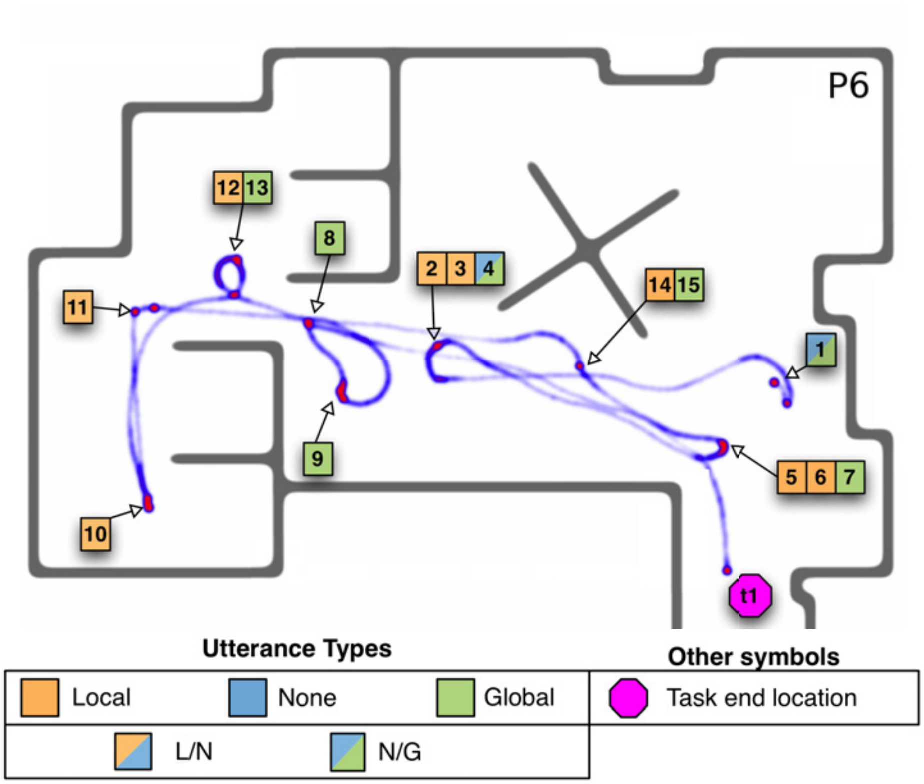
1	G	0:28	Kelsey, I need to go to the grocery and get two snacks.
2	N	1:13	Kelsey, you can stay right there. I'll look at that.
3	L	1:22	Two types of snacks. I'll take some of the cupcakes.
4	L	1:38	I'll take what is underneath the cupcakes.
5	G	2:02	OK. I'll go with, I'll take one type of drink please.
6	L	2:47	Well, I guess I'll get some water, I guess.
7	G	3:00	OK, one movie. I need to get one movie.
8	G	4:15	Is there any more choices for movies?
9	G	4:54	I only see two choices and it doesn't give me... It doesn't really show me the movies. I don't know what the heck that is. There are some more choices there. You can stop at the other choices, please.
10	N	5:32	Stop.
11	L	5:34	I'll take "The Greatest Show on Earth."
12	G	5:47	OK, I need to go to costumes. I need to go to costumes now.
13	N	6:36	OK, you can stop it.
14	L	6:54	Do you have any, do you have any T-shirts for Christmas? I'd like a shirt.
15	L	7:06	That's not... OK, well. Yeah, I found something. I'll take the hat, the Christmas hat and the shirt.
16	L	7:24	Now, I need to a wig and a mask. I need to find a wig.
17	L	7:59	I still...I need to find a wig and a mask.
18	L	9:12	That's a mask, but that's not a Chris... Well, that's the only thing I could think of. OK, I'll take that, that mask, that's right there.
19	G	9:32	I don't know. It's the only thing [inaudible]. Nothing else. OK. Go to the party planner station, please.
20	G	9:48	I need to go to the party planner station.
t1	–	10:04	***End of Task 1***
21	G	12:31	I want you to point me toward whatever Christmas movies there are.
22	G	12:45	You need to bring me back to Christmas movies.
23	L	13:49	Yeah. Yeah, I will take that one.
24	G	15:05	I need one party game.
25	L	15:44	A rock and...I'll that the "Rock n' Rock," if that's what it says, I guess. I'll take the last one.

Figure C-3: Heatmap of the robot's trajectory labelled with P3's timestamped utterances (human agent condition). P3 revisited sections of the store with the party planner after the completion of Task 1 (utterances 21–25). P3 primarily used declarative language.



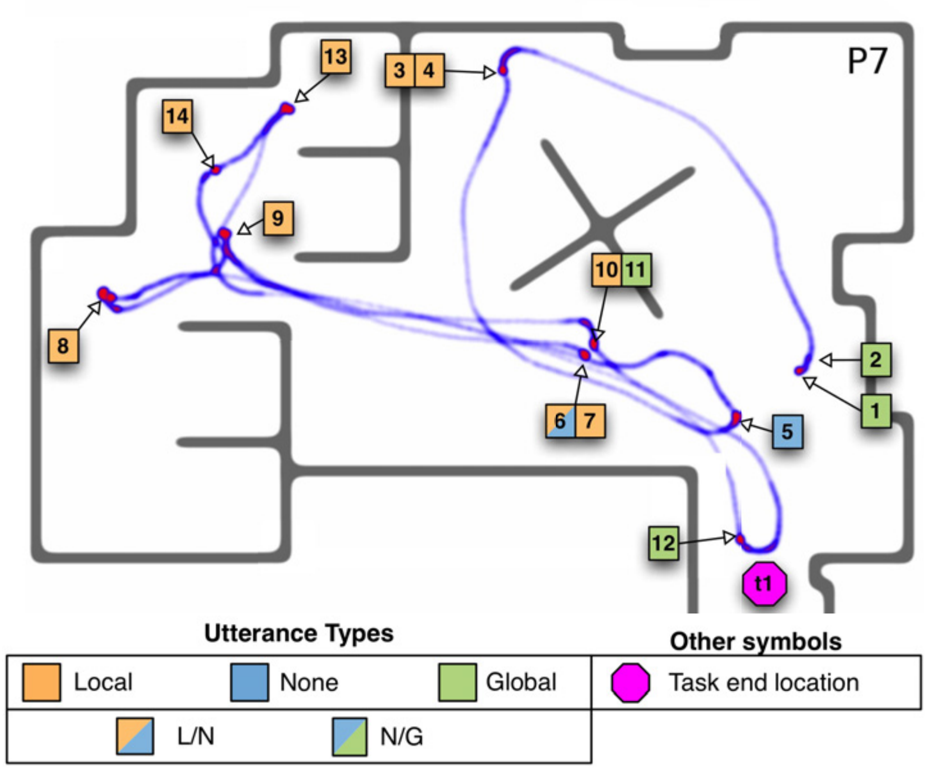
1	G	0:20	I want to go to grocery, please.
2	G	1:06	OK, go to snacks.
3	L	1:47	It will be Oreos, please.
4	L	1:58	Pretzels and...
5	L	2:07	Pretzels.
6	N	2:30	Go forward.
7	N/G	2:45	Left. Costumes.
8	L	3:11	The black shirt.
9	N	3:17	Turn a bit to the right.
10	L/N	3:28	Blue shirt. Turn that, right, left.
11	N	3:46	Turn left.
12	N	3:51	Left again.
13	N	3:59	Back up a little bit. Back up.
14	N	4:18	Back up more.
15	L	4:33	Get some orange, orange soda.
16	L	4:38	Orange soda.
17	N	4:51	Turn left.
18	N	4:52	Come back to right a little bit?
19	L/N	5:01	To the right more. Bottom right.
20	N	5:26	Left.
21	N/G	5:39	Forward. Call box.
22	N	6:29	OK. Turn left. Turn left again.
23	N	6:45	Forward, into movies.
24	N	6:54	Left.
25	N	6:59	Forward.
26	N	7:03	Turn left.
27	N	7:17	Forward.
28	L	8:35	Take a picture of "Robots."
29	N	8:42	Turn right, then left.
30	N	8:45	Turn left.
31	N	8:49	Go forward. Forward, then turn.
32	N	9:10	Turn left.
33	N	9:22	Back up, please. Back up a little bit, please.
34	N	9:35	Back up a little bit, please.
35	L	9:48	Rock 'Em Sock 'Em Robots.
36	N	9:58	Turn around, the other way.
37	N	10:10	Forward. [inaudible]
38	N	10:15	Back up.
39	N	10:24	Go to the right.
40	N	10:28	Forward.
41	N	10:33	Left.
42	N	10:45	Forward.
43	N	10:59	Left.
44	N	11:07	Right, turn around.
45	N	11:22	Forward
t1	–	11:36	***End of Task 1***

Figure C-4: Heatmap of the robot's trajectory labelled with P5's timestamped utterances (human agent condition). P5 gave more than four times as many commands that did not require environmental knowledge (n=40) than global ones (n=9). He "initially talked to her [Kelsey] like a human" but changed his verbal instruction style to give FBLR directions, which he said felt that he had more control.



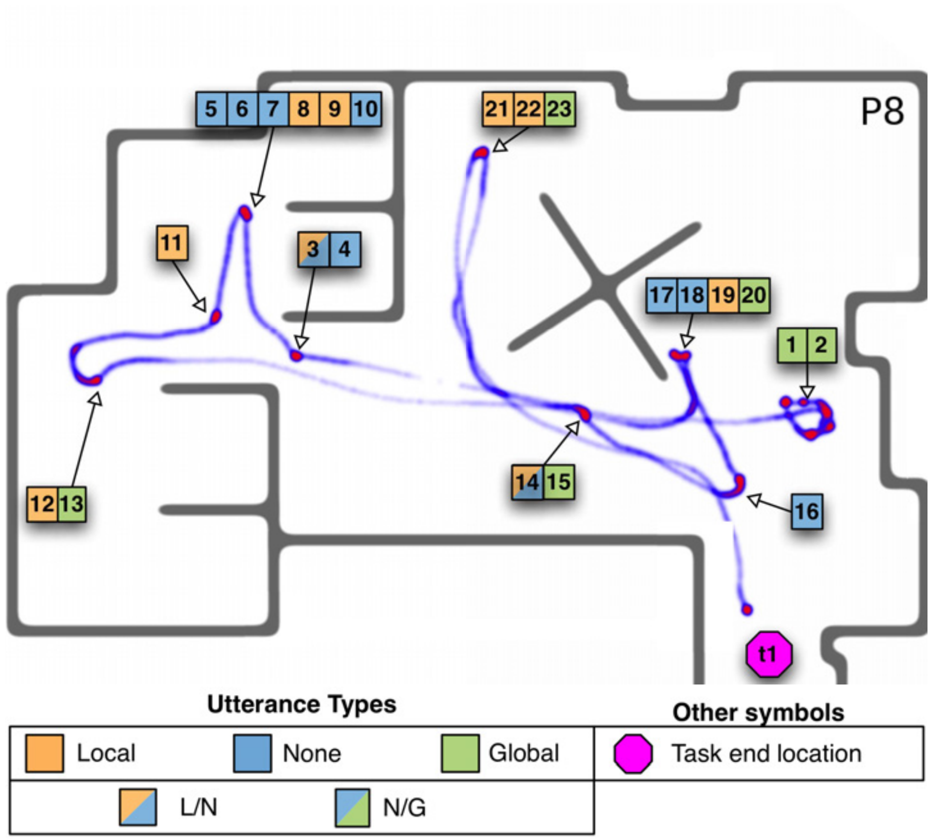
1	N/G	0:11	OK. Can you go left towards the grocery section?
2	L	1:04	OK. Would you be able to get the cupcakes that are located the top right of the area in front of you?
3	L	1:35	Also the ones below? Would you take a picture of that?
4	N/G	1:46	OK. Now would you go to the area that's to the right that has the drink selection?
5	L	2:29	OK. Would you please take a picture of the water to the left at the bottom?
6	L	2:45	And also the milk that's on the right side of the bottom?
7	G	3:03	OK. Now would you go to the entertainment section, which is located to the left of the grocery section?
8	G	4:25	If you go to the call box, which is...
9	G	5:50	Would you go back to the movie section?
10	L	7:19	OK. Thank you. Margo, would you please take a picture of the "Star Wars" poster?
11	G	8:10	Margo, is there a section in the entertainment category that has games?
12	L	9:03	OK. If you would take a picture of the circus game.
13	G	9:23	And would you be able to go to the area in entertainment, to the right of entertainment that has the costume selection?
14	L	10:27	Would you please take a picture of the left shirt that is gray?
15	G	10:48	Now, if you would go to the checkout area, to the right of the costumes.
t1	–	11:28	***End of Task 1***

Figure C-5: Heatmap of the robot's trajectory labelled with P6's timestamped utterances (robot agent condition). P6 primarily used interrogative language with compound and complex sentence structure.



1	G	0:03	I need two types of snacks.
2	G	0:19	One type of drink. One movie. One party game, and one costume.
3	L	0:59	OK. Think I want... Oreos.
4	L	1:16	And... pretzels.
5	N	2:14	OK. Can you move forward a little bit?
6	L/N	2:26	That's good. I would like milk.
7	L	2:46	... and juice.
8	L	3:52	I would like “WALL-E.”
9	L	4:38	I'll pick “Which Witch.”
10	L	5:33	I'll do the black shirt with the skeleton.
11	G	5:50	Now it says to go to the party planning station and finalize choices.
t1	–	6:21	***End of Task 1***
12	G	7:44	Go to movies
13	L	9:10	OK. I think I'll go with that one.
14	L	10:18	We can try the circus one.

Figure C-6: Heatmap of the robot's trajectory labelled with P7's timestamped utterances (human agent condition). P7 revisited sections of the store with the party planner after the completion of Task 1 (utterances 12–14). P7 gave two commands requiring global environmental knowledge at the beginning of Task 1, then specified only the items she wanted for her party using commands requiring local environmental knowledge.



1	G	0:09	I would like you to go to the entertainment store.
2	G	2:42	I would like you to go to the entertainment store, please.
3	L/N	3:35	Thank you. I want one movie. What the heck is the circus thing? I cannot tell what Santa is. Now, can you tell her to move right.
4	N	4:07	Margo, would you move right, please?
5	N	4:42	Margo, would you turn left?
6	N	4:51	Right there.
7	N	5:11	Margo, I have a question for you.
8	L	5:16	Could you pick up Rock 'Em Sock 'Em Robots?
9	L	5:36	Would you take a picture of the circus?
10	N	5:53	Would you move left, please?
11	G	6:14	Would you go to movies?
12	L	6:54	Would you take a picture of "The Greatest Shown on Earth?"
13	G	7:10	Would you please go to costumes?
14	L/N	7:55	I can't even tell what half of these things are. Would you take a picture of the blue shirt?
15	G	8:14	Would you please go to type of drink?
16	N	9:02	Could you move forward?
17	N	9:19	Stop.
18	N	9:20	Could you turn left?
19	L	9:36	Would you take a picture of the Gatorade2?
20	G	9:46	Would you go to types of snacks?
21	L	10:47	Would you take a picture of Chips Ahoy?
22	L	10:58	Would you take a picture of Oreos?
23	G	11:09	Would you go to the Party Planner Station?
t1	–	12:00	***End of Task 1***

Figure C-7: Heatmap of the robot's trajectory labelled with P8's timestamped utterances (robot agent condition). P8 primarily used interrogative language.

Appendix D

UI Guidelines

In Chapter 5, we drew from HCI interface design guidelines (i.e., Nielsen's usability heuristics [Nielsen, 1994a]), guidelines for accessible web design [Kurniawan and Zaphiris, 2005; W3C, 2008] and accessible consumer products [Vanderheiden and Vanderheiden, 1992], including our previous work [Tsui et al., 2008, 2009, 2010]. We developed a more complete set of heuristics for assistive robotics by pulling applicable heuristics from the literature and using the Model Human Processor [Card et al., 1983] as a theory-based framework; this set was documented in Tsui et al. [2009] and can be found in Tables D.1 and D.2.

Table D.1: UI heuristics relevant to assistive robotics; from [Tsui et al., 2009]. (Part 1 of 2)

Heuristics	Source
Provide appropriate amounts of information for decision-making, judgment, and prediction	[Card et al., 1983]
◦ Show what the system is doing and what state it is in	[Nielsen, 1994b]
– Is there an alert of degraded mode?	
– Is there feedback during long operations?	
– Is it obvious when the robot is expecting input from the user?	
◦ Provide option awareness to enable decision makers to know what courses of action are available, what their likelihoods of success are, and what their relative costs are	[Drury et al., 2009]
◦ Provide sufficient historical information to understand trends and make predictions	[Endsley, 1988]
Use existing long-term and working memory	[Card et al., 1983]
◦ Minimize process length	[Vanderheiden and Vanderheiden, 1992], [Tsui et al., 2008]
– Does the task require prolonged retention of information to complete?	
– Are there steps that are unnecessary for the user to complete; could the system automate some of the steps?	
– The focus is on keeping the process within working memory.	
◦ Provide consistency and standards	[Nielsen, 1994b]
– Is the meaning of any action, icon, or menu option overloaded? Not only overloaded within the application, but between the application and something external that is a de-facto standard.	
– Users should not have to wonder whether different words, situations, or actions mean the same thing.	
– Follow platform conventions.	
◦ Exploit previous knowledge in the world if reasonable	[Norman, 2002], [Nielsen, 1994b]
◦ Provide knowledge in the interface so that people do not have to remember it	[Norman, 2002], [Nielsen, 1994b]
– Can interface operation be tied to knowledge that people have already learned?	
– Note: it may not be reasonable to exploit real-world knowledge if that leads to the interface being too inefficient for the users' needs	
Reduce motor processing time	[Card et al., 1983]
◦ Accommodate the ability to choose among access devices	[Vanderheiden and Vanderheiden, 1992], [Tsui et al., 2008]
– Are there multiple methods of robot control/communication suitable for people of varying disabilities?	
◦ Support shortcuts	[Nielsen, 1994b]
– Are frequently used operations easily accessed?	
– Are reasonable defaults given that provide the most commonly expected value?	
– Leverage information from other sources (e.g., previously entered information; auto zip code for address, city, and state).	
Reduce mental processing cycles	[Card et al., 1983]
◦ Use simple language	[Nielsen Norman Group Report, 2001]
– Is there any technical jargon?	
– Can the language be simplified and still get the point across?	
◦ Avoid having the user make mental translations	
– Employ direct manipulation to avoid cognitive mapping	
– Fuse data so the user does not have to make mental mappings (e.g., distance and video).	
Support flexibility to match differing expectations	[Nielsen, 1994b]
◦ Provide multiple ways to access a function/complete a task	
– Can this task be done more than one way?	
– Does each method map to a separate conceptual model of the task?	
◦ Provide user control and freedom of actions	[Nielsen, 1994b]
– Can actions be undone or repeated?	
– Can current tasks be canceled?	
– "Users should be free to select and sequence tasks (when appropriate), rather than have the system do this for them. Users often choose system functions(e.g., wrong menu option) by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Users should make their own decisions (with clear information) regarding the costs of exiting current work. The system should support undo and redo."	[Nielsen, 1994b]
◦ Be consistent with how the human brain processes information	[Miller, 1956], [Miller et al., 1971], [Zellner, 1988]
– Is information presented in the appropriate order?	
– Is basic information presented before more detailed information?	
– Is pre-attentive processing used?	[Treisman and Gelade, 1980]
◦ Enable interface customization and retention of user's preferences	[W3C, 1994]
– Support adjustment of prompting levels.	[Tsui et al., 2008]

Table D.2: UI heuristics relevant to assistive robotics; from [Tsui et al., 2009]. (Part 2 of 2)

Heuristics – continued...	Source
Aid in perception	[Card et al., 1983]
◦ Provide aesthetic and minimalist design	[Nielsen, 1994b]
– Is the interface easy to understand?	
– Is the layout as minimal as possible?	
– "Interfaces should not contain information which is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes their relative visibility."	[Nielsen, 1994b]
◦ Present content appropriately	[Vanderheiden and Vanderheiden, 1992], [W3C, 1994]
– Does content presentation leverage a person's sensory ability?	
– Is it high-contrast with a large font for low-vision users?	[Nielsen Norman Group Report, 2001], [Vanderheiden and Vanderheiden, 1992]
– Is there volume adjustment?	[Vanderheiden and Vanderheiden, 1992]
– Are there color combinations that affect the color-blind?	[Vanderheiden and Vanderheiden, 1992]
Ensure safety	[Vanderheiden and Vanderheiden, 1992], [Heinzmann and Zelinsky, 2003]
◦ Ensure robot does not have a physical form that can induce injury	
– Are there any sharp edges that a person might get cut on or cause bruising?	
– Are there joints that could pinch?	
– Any surfaces that are too hot to touch?	
◦ Ensure robot does not have behaviors that can induce injury	[W3C, 1994], [U.S. Department of Justice, 1994]
– Can the robot harm a person unintentionally?	
– Are warning signs placed around robots that are unaware of their surroundings?	
– If the robot has an audio output, is it too loud?	[U.S. Department of Justice, 1994]
– Ensure interface components do not flash more than 3 times per second.	[W3C, 1994]
◦ Provide fail-safe mechanisms	[Tejima and Stefanov, 2005]
– Is there an E-stop which can override the robot's actions?	
– If a component of the robot malfunctions, does it do so in a way that could harm someone either actively or passively?	
Prevent errors	[Nielsen, 1994b]
◦ Provide context-sensitive help when asked	[W3C, 1994], [Nielsen, 1994b]
– Is help always available?	
– Does that help relate directly to the current task or action?	
– Does the system have an expectation about what the user is trying to do?	
◦ Prevent capture errors	[Norman, 2002]
– Are there any sequences that are similar that lead to two different states, particularly ones that are similar in the initial states?	
◦ Prevent description errors	[Norman, 2002]
– Are there more than one objects that look the same?	
– Is it possible to perform the right action on the wrong object?	
◦ Prevent mode errors	[Norman, 2002]
– Do not change modes unexpectedly.	
– Minimize the use of modes.	
– Minimize the number of mode changes.	
Maximize the user's trust	
◦ Ensure robot performs in a predictable manner	[Bartneck, 2002]
– Is the robot's appearance consistent with it's actions?	
– Is the state of the robot displayed at the right time in relation to its actions?	
– Is the display state "intensity" appropriate?	
– Does the robot perform as you expect?	
◦ Ensure robot performs in accordance with polite social etiquette	
– Does the robot behave politely with respect to the user's cultural or generational expectation?	
◦ Provide feedback and interaction that matches technical abilities	[Bartneck and Okada, 2001]
– Does the robot look more sophisticated than it really is?	
– Does the robot's interaction make you feel it is more capable than it is?	
◦ Reduce anxiety	
– Does interaction with the robot cause anxiety?	
– Does the robot look "too real" (uncanny valley)?	[Fong et al., 2003], [DiSalvo et al., 2002]

Appendix E

Study 3: Interview Questions

The interview questions below were verbally proctored. Follow-up questions were asked as needed to obtain more relevant information.

Session 1: Demographic and technology experience

SECTION 1: In-person outings (experience since injury)

1. When was the last time you went on an outing to an art gallery or museum?
2. Where was it?
3. Did you go by yourself? **Yes** **No**
   If not, were you with family? Friends?
4. What did you see?
5. How would you describe your overall experience that day? [Prompt: what kind of words/adjectives would you use to describe your overall experience of that day (e.g., fun, enjoyable, boring)?]
6. How well did the gallery's/museum's accessibility suit your needs?

7. Do you know if there any arrangements you or another person needed to make before going (e.g., transportation to/from, schedule staff guidance)?

   Yes **No**

   If so, what were the arrangements?

SECTION 2: Video games (all prior experience)

1. Do you play or have you played any avatar/character based video games (e.g., World of Warcraft, Minecraft, Sims)? **Yes** **No**

   IF NO, SKIP TO VIDEO GAME QUESTION 11

2. If yes, what game are you currently playing, or what was the most recent game you played?

3. How long since you've last played? When was the last time you played that video game?

4. How many hours do you typically play avatar based video games in a two week period?

5. Describe your current/most recently played avatar based video game.

6. Do/did you play with other human players, online or in person? **Yes** **No**

7. If yes, what is the objective of the game?

8. What are your physical controllers/access methods to play the game? (e.g., controller, touchscreen)?

9. Is the game in a first person view, third person view, or both? Describe. Which do you prefer?

10. How do you move your avatar/character from one place to another/to a new location?

11. Have you ever played an online videogame? **Yes** **No**

12. If yes, did you experience high delay while playing? How strongly does delay affect your experience?

SECTION 3: Vehicles (question #4 experience after injury; all other questions all experience)

1. Have you ever had a state issued driver's license? **Yes** **No**
   IF NO, SKIP TO CAR GPS SECTION

2. Do you currently drive, or have you ever driven an automobile (non-commercial car, truck, van, motorcycle)? **Yes** **No**
   IF NO, SKIP TO CAR GPS SECTION

3. What was your most recent vehicle (make, model, year)?

4. Were there any modification done to your vehicle to facilitate your driving? If so, please describe (e.g., hand shifters). **EXPERIENCE AFTER INJURY ONLY**

5. Was it manual or automatic transmission?

6. Did it have cruise control? **Yes** **No**
   Did you use this feature? **Yes** **No**

7. Did it have a rear backup camera? **Yes** **No**
   Did you use this feature? **Yes** **No**

8. Did it have self parallel parking? **Yes** **No**
   Did you use this feature? **Yes** **No**

SECTION 4: Car GPS Navigation (all prior experience)

1. Have you used a car navigation system/GPS? **Yes** **No**

2. What device do you use, or have you used? onStar? A dedicated device (e.g., Garmin, Magellen)? Smartphone app (e.g., Google Navigation, Waze)?

3. Describe its capabilities; did it have:

   (a) voice guidance (text to speech output) **Yes** **No**

   (b) turn by turn navigation **Yes** **No**

   (c) route highlighting (showing the planned route) **Yes** **No**

   (d) photo of the destination **Yes** **No**

   (e) first person point of view
   (icon facing the same direction as what the vehicle is actually traveling)?
   Yes **No**

   (f) third person point of view with North at the top of the screen?
   Yes **No**

4. How frequently do you use it? **(times per day, week, month, year)**

Session 1: Interview following initial training

User Interface

1. What was easy to do with the user interface?

2. What was hard to do with the user interface? Why?

3. What would you change about the user interface? Why? How would you change it?

Session 1: Interview following the first in-robot gallery visitation

Gallery

1. Describe your **most favorite** exhibit. *[Prompt: What did it do? What color was it? Did it make any sounds? Did it move?]*
    (a) Why is it your favorite? What made it stand out to you from all the other exhibits? *[Prompt: Is it funny? Pretty?]*
    (b) How did interacting with it make you feel? *[Prompt: Does it make you laugh/smile?]*
2. Describe your **least favorite** exhibit. *[Prompt: What did it do? What color was it? Did it make any sounds? Did it move?]*
    (a) Why is it your least favorite? Why did you like it the least?
    (b) How did interacting with it make you feel?

Navigation

1. What was your strategy for exploring the art gallery and visiting the exhibits? *[Alternative phrasing. <u>Option 1</u>: Did you have a plan for exploring the art gallery and visiting the exhibits? If so, what was it? <u>Option 2</u>: Which exhibit did you visit first? What made you choose to go to this one first? Which exhibit did you visit next? What made you choose to go to that one next?]*
2. What was your strategy for directing the telepresence robot?
3. Was your strategy the same or different at the end of today's session as it was at the beginning?
    If different, how did this change over time?

Expectations

1. What expectations did you have for driving the robot?

   (a) Did the robot meet those expectations? **Yes** **No**

   (b) If so, did the robot exceed those expectations?

   (c) If not, did the robot fall short of those expectations?

2. Was there anything that the robot did that you did not expect/understand or that confused you? Please describe.

3. Did the robot behave in any strange, unexpected, or unacceptable way?

   Yes **No**

   If so, please describe.

4. Was there anything that you wished the robot could have done (but was not able to) in today's session?

User Interface

1. What was easy to do with the user interface?

2. What was hard to do with the user interface? Why?

3. What would you change about the user interface? Why? How would you change it?

Misc

1. How would you describe your experience today using the robot to visit an art gallery to a friend?

2. Do you have any suggestions for other places we could place a robot like the one you drove today?

3. What other things would you want to do with a robot like the one you drove today?

Session 2: Interview following the second in-robot gallery visitation

Gallery

1. Describe your **most favorite** exhibit. *[Prompt: What did it do? What color was it? Did it make any sounds? Did it move?]*

   (a) Why is it your favorite? What made it stand out to you from all the other exhibits? *[Prompt: Is it funny? Pretty?]*

   (b) How did interacting with it make you feel? *[Prompt: Does it make you laugh/smile?]*

   (c) **[If different from session 3]** Why is this exhibit now your new favorite?

2. Describe your **least favorite** exhibit. *[Prompt: What did it do? What color was it? Did it make any sounds? Did it move?]*

   (a) Why is it your least favorite? Why did you like it the least?

   (b) How did interacting with it make you feel?

   (c) **[If different from session 3]** Why is this exhibit now your new favorite?

Conversation with person in the gallery (confederate)

CONFEDERATE'S NAME: ______________

1. What was **[INSERT CONFEDERATE'S NAME]**'s favorite exhibit?
   Why was it his/her favorite?

2. Was there anything normal about having a conversation with the person in the gallery while using the robot? Describe.

3. Was there anything strange about having a conversation with the person in the gallery while using the robot? Describe.

4. How was **[INSERT CONFEDERATE'S NAME]** feeling today? What was his/her mood?

5. Was there any time in today's session that you **[INSERT CONFEDERATE'S NAME]** took too long to respond to something you said? **Yes** **No**
   How did it make you feel? Describe the situation.

6. Was there any time in today's session that you felt **[INSERT 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.

7. Was there any time in today's session that you felt like you had to talk to **[INSERT CONFEDERATE'S NAME]** when you didn't want to?
   Yes **No**
   Describe the situation.

8. Was there any time in today's session where you couldn't understand what **[INSERT CONFEDERATE'S NAME]** was saying? **Yes** **No**
   Why?

Navigation

1. What was your strategy for exploring the art gallery and visiting the exhibits?

2. What was your strategy for directing the telepresence robot?

3. Was your strategy the same or different at the end of today's session as it was at the beginning?
   If different how did this change over time? **Yes** **No**

4. Was your strategy today the same as or different from the last time you drove the telepresence robot in the art gallery? **Yes** **No**
   If different, how did it change?

Expectations

1. What expectations did you have for driving the robot?

   (a) Did the robot meet those expectations? **Yes** **No**

   (b) If so, did the robot exceed those expectations?

   (c) If not, did the robot fall short of those expectations?

2. Was there anything that the robot did that you did not expect/understand or that confused you? Please describe. **Yes** **No**

3. Did the robot behave in any strange, unexpected, or unacceptable way?

   Yes **No**

   If so, please describe.

4. Was there anything that you wished the robot could have done (but was not able to) in today's session?

   If so, please describe. **Yes** **No**

User Interface

1. What was easy to do with the user interface?

2. What was hard to do with the user interface? Why?

3. What would you change about the user interface? Why? How would you change it?

Misc

1. How would you describe your experience today using the robot to visit an art gallery to a friend?

2. Do you have any suggestions for other places we could place a robot like the one you drove today?

3. What other things would you want to do with a robot like the one you drove today?

Session 2: Interview during in-person visitation

In-person visit to the gallery

1. Which exhibits look the same visiting them in person?
2. Which exhibits look mostly the same but slightly different visiting them in person? Describe. Colors? Sounds? Motions?
3. Which exhibits look completely different visiting them in person? Describe. Colors? Sounds? Motions?

Appendix F

Measuring the Quality of an Interaction

Researchers have investigated the efficacy in which people can use telepresence robots to navigate in remote locations (e.g., [Michaud et al., 2010; Tsai et al., 2007; Tsui et al., 2011a, d]), the interfaces to do so (e.g., [Michaud et al., 2010; Takayama et al., 2011; Tsui et al., 2011b]), and how the robots should be designed (e.g., [Cohen et al., 2011b; Deml, 2007; Desai et al., 2011]). Telepresence robots have great potential to provide utility in workplaces (e.g., [Lee and Takayama, 2011; Tsui et al., 2011a]), in schools (e.g., [Sheehy and Green, 2011]), in homes (e.g., [Coradeschi et al., 2011]), and for excursions to museums, sporting events, and the theater (e.g., [Beer and Takayama, 2011]), for example. However, the quality of a person to person interaction through a telepresence robot has not yet been explicitly quantified. In this appendix, we discuss the performance measures needed to assess a communication by leveraging work from the fields of human-computer interaction (HCI), computer supported cooperative work (CSCW), communications, and psychology.[1]

[1] Portions of this appendix were published in [Tsui et al., 2012].

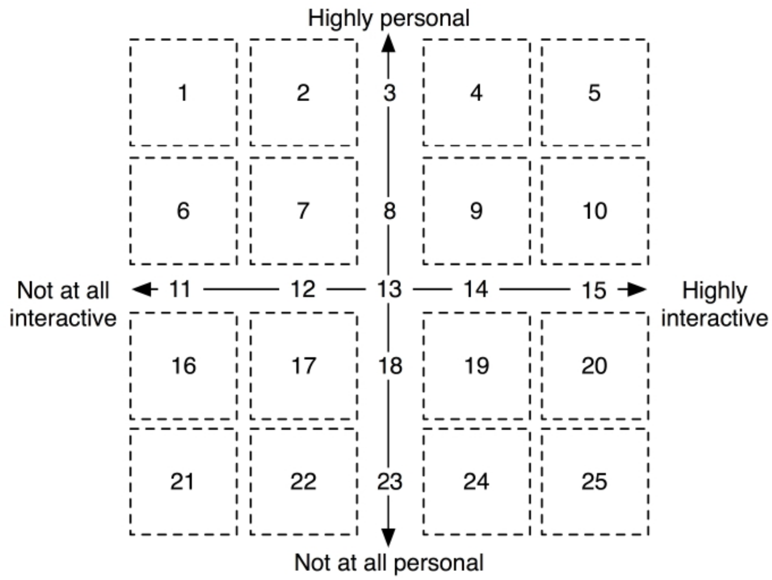
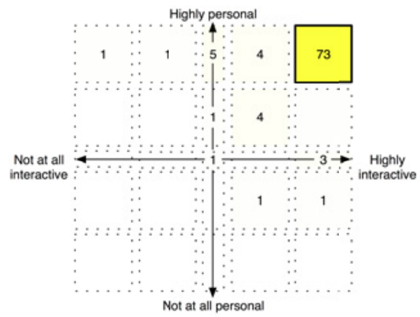


Figure F-1: Diagram of interactivity and personalness scales. Participants were asked to categorize communication technologies. Original diagram by Jake Knapp of Google; modified to include region enumeration.

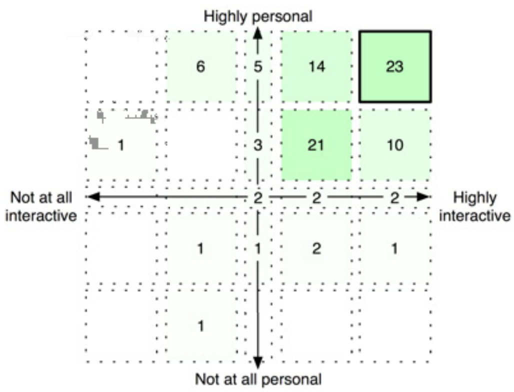
F.1 Comparison of Interaction Mediums

We conducted a survey to investigate how people would categorize several communication technologies with respect to interactivity and personalness. The baseline was "face to face" (FTF) interaction; the technologies included video conferencing, telephone call, telepresence robot, and instant messaging/chat. Each technology has at least one layer of indirection. For example, a phone conversation can be misinterpreted given the lack of facial expression. Text-based instant messaging additionally lacks vocal intonation but includes some level of emotion through emoticons and meta-actions (e.g., smiley face :), *hug*). Video conferencing has audio and facial expressions and gestures seen through a webcam; however, the webcam provides a single vantage point and is subject to adjustment (or lack thereof) by the video conferencing recipient. Telepresence robots also have two-way audio and video, and additionally provide a mobile embodiment to the remote party which allows for independent movement.

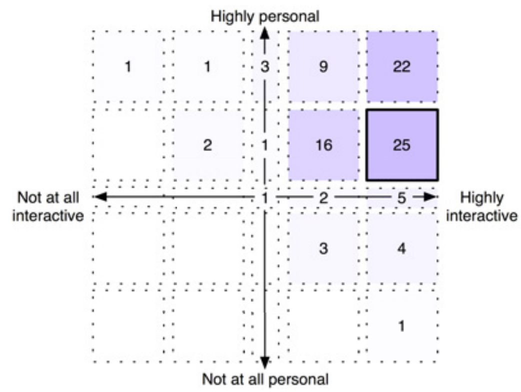
The survey was conducted using Amazon's Mechanical Turk (MTurk) [Amazon, 2009]. For each means of communication, MTurk Workers were asked where they would place it in Figure F-2a with respect to the communication's personalness



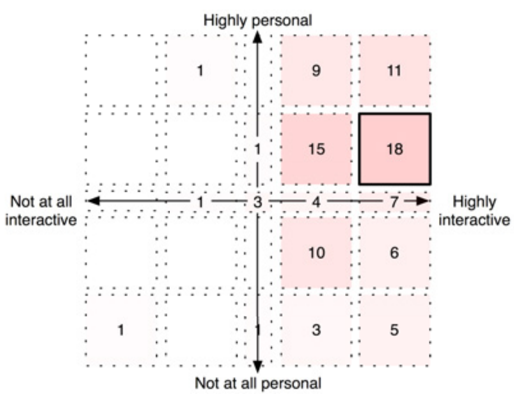
(a) Face to face (FTF) [baseline]



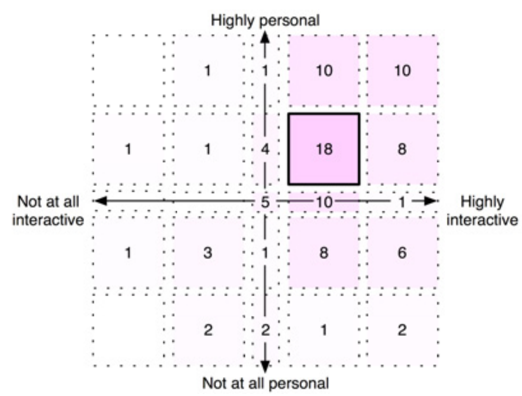
(b) Phone conversation
(audio only, AO)



(c) Video conferencing
(video mediated communication, VMC)



(d) Telepresence robot (TPR)



(e) Instant messaging/chat (IM)

Figure F-2: Frequency counts are shown inside each category and the mode is marked by a solid black outline (n=96).

and interactivity. That is for example, a highly personal and highly interactive communication method would be placed in the top-right quadrant in category 5. Because telepresence robots are an emerging commercial technology, we showed MTurk Workers photos of five examples: VGo, RP-7, QB, Texai, and TiLR. We also provided the following definition: “A telepresence robot can be thought of embodied as video conferencing on wheels: the robot is a representation of you. You can see what is around the robot through its camera and hear through its microphones. People with the robot can hear and see you too.” Ninety-six people participated in the survey and were each paid $1.00.

Figures F-2b-f show the category frequency for each communication method. Face to face interaction was chosen en masse as both highly personal and highly interactive; 76% of the participants (73 of 96) selected category 5 in Figure F-2a. The communication technologies however had less of a consensus. Participants selected categories in the top right quadrant (categories 4, 5, 9, and 10 in Figure F-2a) for phone conversations (71%), video conferencing (75%), telepresence robot (55%), and instant messaging (48%). The communication technologies were rated all as personal and interactive but to varying degrees given that 25 or fewer participants' votes comprised the modes.

We then transformed each communication method's categorical data into continuous data by separating each axis and assigning values. For the interactivity axis, a value of one was assigned to the left-most category (not at all interactive) and five to the right-most (highly interactive). The frequency count for each column was summed and divided by the number of participants (n=96), thus yielding the weight of the value. We multiplied each category value by its calculated weight. Summing these results provided the average value in rational form, which provided insight if a communication method split two categories on a single axis. We similarly calculated the average value along the personalness axis where a value of one was assigned to bottom-most category (not at all personal) and five to the top-most (highly personal).

Figure F-3 shows the averages and standard deviations for the communication methods. We conducted unpaired t-tests for all of the communication method permu-

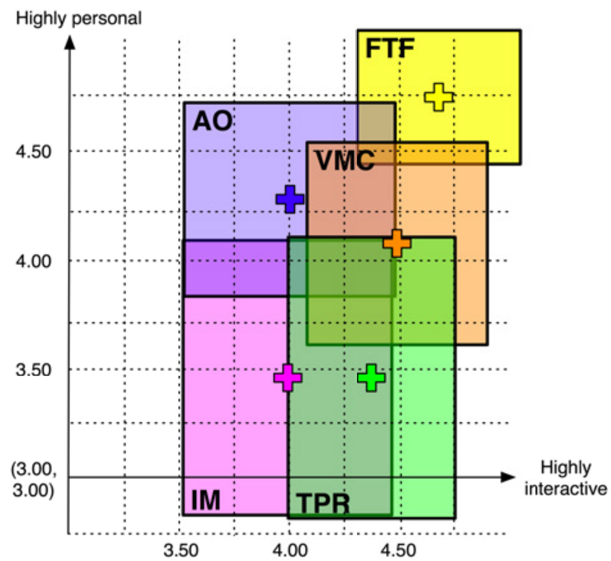


Figure F-3: Averages and standard deviations for face to face (FTF), phone call (audio only, AO), video conferencing (video mediated communication, VMC), telepresence robot (TPR), and instant messaging/chat (IM). Plus signs denote averages in the form (interactivity $\bar{I}$, personalness $\bar{P}$), and rectangles denote $\pm$ 1 SD.

tations with respect to personalness and also with interactivity. The significance value is α=0.005 as we divided the goal 95% confidence value by the ten test permutations. Face to face interaction rated as the most personal and the most interactive form of communication ($\bar{P}_{FTF}$=4.75 (0.61), $\bar{I}_{FTF}$=4.64 (0.74)) We found that the face to face interaction was significantly more personal than all of the communication technologies ($p_{personal}$<0.002). It was significantly more interactive compared to a phone call and instant messaging ($p_{interactive}$<0.001), but not so when compared to video conferencing (p<0.158, t(190)=1.419) or telepresence robots (p<0.010, t(190)=2.586).

Phone calls were also highly personal but less interactive than face to face interactions ($\bar{P}_{AO}$=4.34 (0.88), $\bar{I}_{AO}$=3.94 (0.96)). We found that phone calls were significantly more personal than instant messaging and telepresence robots ($p_{personal}$<0.001), but significantly less interactive than video conferencing (p<0.001, t(190)=3.570) and also telepresence robots (p<0.007, t(190)=2.720) though not significantly. On the other hand, video conferencing was highly interactive but less personal than face to face interactions ($\bar{P}_{VMC}$=4.11 (0.92), $\bar{I}_{VMC}$=4.48 (0.82)). We found that video conferencing was both significantly more personal and more than interactive instant messaging (p<0.001). When compared to telepresence robots, video conferencing was significantly more personal ($p_{personal}$<0.001) but was not significantly different with respect to interactivity (p<0.295, t(190)=1.052).

As shown in Figure F-2e, 92% of the participants rated telepresence robots as interactive despite being given only pictures of telepresence robots and a brief description as to their capabilities. However, there was a lack of consensus as to how personal an interaction using a telepresence robot could be. We hypothesize that this result is because telepresence robots are a new commercial product and while people may know of their existence, they are not yet familiar with them. Therefore, we must look at performance measures that assess the quality of interaction through telepresence robots in pieces: the quality of a communication from a technical standpoint (audio and video), and the quality of a human-human communication through a telepresence robot.

F.2 Audio Signal Measures

The most important component of communicating through a telepresence robot is the conversation itself. Rosenberg notes that audio quality can be measured in terms of being able to understand speech and the fidelity of the speech itself [Rosenberg, 2010]. In terms of the speech fidelity, the audio quality must be comparable at least to that of a landline phone [Desai et al., 2011]. The ITU-T G.711 Recommendation was initially designed for the Public Switched Telephone Network with 64kbps bandwidth in 1972 [Intl. Telecommunication Union, 2011]. G.711's digital counterpart, the ITU-T G.729 Recommendation, was established in 1996 and is popular for voice-over-IP telecommunication given its low bandwidth requirements (8kbps), although at the cost of high compression [Intl. Telecommunication Union, 2009]. Rosenberg notes that as the audio fidelity increases, the length of a conversation also increases [Rosenberg, 2010]. In a study of Skype's SILK codec versus G.729, he reports that users spent 40% longer in calls with the SILK super-wide bandwidth (24kHz) codec.

A codec's speech fidelity is measured by its Mean Opinion Score (MOS), which

Table F.1: Subjective evaluation of conversational quality from ITU-T Recommendation P.805 [2007]

Question	Scale
What is your opinion of the connection you have just been using? [Mean Opinion Score (MOS)]	1=bad quality; 5=excellent quality
How would you assess the sound quality of the other person's voice?	1=severe distortion; 5=no distortion at all, natural
How well did you understand what the other person was telling you?	1=severe loss of understanding; 5=no loss of understanding
What level of effort did you need to understand what the other person was telling you?	1=severe effort required; 5=no special effort required
How would you assess your level of effort to converse back and forth during the conversation?	1=severe effort required; 5=no special effort required
Did you detect (insert distortion of interest here)? If yes, how annoying was it?	yes/no 1=severe annoyance; 5=no annoyance

is one item of a series of subjective rating questions measuring the quality of speech listed in ITU-T Recommendation P.805 (see Table F.1). Telecommunication users may be explicitly asked to rate the quality of their connection on a 5-point semantic differential scale where 1=bad and 5=excellent. MOS can be determined using controlled user studies in which the sound origin, sound destination, and background noise are manipulated [Intl. Telecommunication Union, 2007]. MOS can also be derived from simulation tests such as the Perceptual Evaluation of Speech Quality (PESQ) [Intl. Telecommunication Union, 2001].

Speech intelligibility is measured on a 5-point scale the like MOS scale [Steeneken, 2006]. Steeneken notes that speech intelligibility can be predicted using several methods. The Speech Interference Level (SIL) subtracts the average noise level within the 500-4000Hz range from the estimated speech level [Beranek, 1947]. The expected SIL result is a decibel level where values less than 3 are bad, between 3 and 10 are poor, between 10 and 15 are fair, between 15 and 21 are good, and above 21 are excellent [Steeneken, 2006]. The Speech Transmission Index (STI) predicts nonsensical speech accounting for the speech and noise range, bandwidth, and physical characteristics of the environment [Intl. Electrotechnical Commission, 1998]. The STI value ranges between 0 and 1 where values less than 0.30 are bad, between 0.30 and 0.45 are poor, between 0.45 and 0.60 are fair, between 0.60 and 0.75 are good, and above 0.75 are excellent [Steeneken, 2006]. Barnett and Knight proposed a common intelligibility scale where CIS = 1 + log(STI) [Barnett and Knight, 1996]. The Speech Intelligibility Index (SII) is similar to STI and also predicts syllabic phonemes in speech [Amer. Natl. Standards Institute, 1997]. The SII value also ranges between 0 and 1 where values less than 0.45 are poor and above 0.75 are good [Steeneken, 2006].

Speech intelligibility can also be quantified in terms of the number of echoes, feedback occurrences, and cutouts (e.g., [Haas, 1972; Miller and Licklider, 1950]). We designed a study, detailed in [Desai et al., 2011], to investigate the use of telepresence robots in ad-hoc scenarios, specifically moving down a hallway while simultaneously having a conversation. We noted each run in which echo, feedback, and cutout occurred through analysis of the robot driver's screen captured video which included audio. It is also possible to obtain a speech intelligibility measure qualitatively as telecommunications users may explicitly be asked in post-experience surveys; ITU-T Recommendation P.805 contains four questions relating to intelligibility (Table F.1).

F.3 Video Signal Measures

Audio is critical for carrying the content of a communication between two parties. Video can communicate emotion through facial expression and gestures, mutual gaze, and conversational attention [Vertegaal et al., 2001]. Video information is also critical for telepresence robots in navigating a remote location. Due to the mobility afforded by these robots, the information must be transferred wirelessly. Video streams constitute a significant portion of the data transferred and can be adversely affected by the network connection. The quality of a wireless connection is influenced by several factors including bandwidth, latency, and packet loss.

We designed one study, detailed in [Desai et al., 2011], to compare the video

Table F.2: Video characteristics rating questions for comparing QB and VGo telepresence robots and EVO phone used in [Desai et al., 2011].

Item	**Scale**
Overall quality	1=poor, 7=good
Field of view	1=too narrow, 7=too wide
Scale perception	1=could not gauge scale, 7=could gauge scale
Contrast/white balance	1=poor, 7=high
Resolution	1=too low, 7=too high
Color depth	1=low/grayscale, 7=high/true color
Degradation in quality	1=very noticeable, 7=not at all noticeable
Pauses in video	1=few, 7=many
Latency	1=low, 7=high

streams from the QB and VGo telepresence robots against a Sprint EVO Android phone. We placed an eye chart four feet in front of the robot and asked the participants to read the letters from both the phone and the robot's video display. We asked the participants to follow a person (an experimenter) through an area with a hallway, cubicles, and a cafeteria. Following each run, the participants rated the video from the robot and EVO phone with respect to field of view, ability to perceive scale, pauses in video, latency, contrast, resolution, color depth, and quality of degradation on a 7-point semantic differential scale (see Table F.2).

Based on the results and our observations, the guiding principle for video streams for telepresence robots is to have two video profiles: one while the robot is mobile (dynamic video profile), and another profile for when the robot is not moving (stationary video profile) [Desai et al., 2011]. Two profiles are needed because the required video characteristics are mutually exclusive at times. Video is the most important sensor information while controlling a telepresence robot. A dynamic video profile should contain characteristics including low latency, few pauses, graceful video degradation, and scale perception. While the robot is stationary, the video profile should contain characteristics including sharp contrast/white balance, increased resolution, and 8-bit color depth or higher.

ITU-T Recommendation P.910 provides a protocol by which multimedia content can

Table F.3: Quantitative communication performance measures surveyed from HCI, CSCW, communications, and psychology. Communication modes included face to face (FTF), audio only (AO), video-mediated communication (VMC), and embodied VMC (eVMC) including telepresence robots.

Measurement	Study Examples			
	FTF	AO	VMC	eVMC
Frequency of communication over time	(5)		(5)	
Number of words				
• in total	(1)		(1),(2)	
• per participant	(1)	(3)	(3),(1)	
Rate of words over time / percentage dialogue			(2)	(4)
Duration of conversation	(5)		(5),(2)	(6),(4)
Number and/or duration of silences	(7),(8),(9)	(7),(8),(3)	(3),(10),(7),(8),(9)	
Number of overlaps		(3)	(3)	
• simultaneous starts	(7),(1),(9)	(7)	(9),(10),(7),(1)	
• floor holding/disfluencies (e.g., “um,” “er”)	(7),(1)	(7)	(7),(1)	
• sentence completion	(1),(7)	(7)	(7),(1)	
• interruptions	(1),(7)	(7),(9)	(1),(7),(2),(10),(9)	
Number of explicit handovers (e.g., question, name of next speaker)	(1),(8),(7)	(7),(8)	(1),(8),(7)	
Number of turns (attempts to gain the floor to speak)	(11),(7),(1),(8)	(3),(7),(8),(9)	(8),(7),(1),(2),(9),(12),(3)	(4)
Duration of turn / words per turn	(7),(1),(8),(9)	(7),(8)	(8),(7),(1),(2),(9)	
Distribution of turns	(7),(1)	(7)	(7),(1)	
Number of backchannels				
• verbal (e.g., “mm,” “uh huh,” “okay”)	(1),(8)	(8)	(8),(1),(2)	
• head nod	(8)	(8)	(8)	
• gaze	(13),(11)		(12)	(6),(4)
Number of gestures (i.e., kinetic, spatial, point, other)	(8),(14)	(8)	(8)	

Key: (1) [O'Conaill et al., 1993]; (2) [Grayson and Coventry, 1998]; (3) [Monk and Gale, 2002]; (4) [Sirkin et al., 2011]; (5) [Fish et al., 1992]; (6) [Venolia et al., 2010]; (7) [Sellen, 1995]; (8) [Isaacs and Tang, 1994]; (9) [van der Kleij et al., 2009]; (10) [Geelhoed et al., 2009]; (11) [Jokinen et al., 2009]; (12) [Vertegaal et al., 2000]; (13) [Otsuka et al., 2006]; (14) [Bekker et al., 1995] be subjectively tested, including sample questions regarding an image's color, contrast, borders, movement continuity between frames, flicker, and smearing/blurring [Intl. Telecommunication Union, 1999]. Questions are rated on a modified MOS n-point scale where 1=bad and n=excellent. ITU-R Recommendation BT.500 provides a protocol for subjective testing of the quality of television pictures [Intl. Telecommunication Union, 2002]. Questions are rated on either a 5-point MOS scale, a 5-point impairment scale (1=very annoying, 2=annoying, 3=slightly annoying, 4=perceptible but not annoying, and 5=imperceptible), or a 7-point comparison scale (-3=much worse, 0=same, +3=much better). Video signal quality can be measured objectively using simulation tests such as the Perceptual Evaluation of Video Quality (PEVQ) [Intl. Telecommunication Union, 2008].

F.4 Human-Human Communication Measures

A high fidelity video and audio channel given sufficient bandwidth provides the foundation for a human-human communication. One common evaluation technique used by companies investing in new telecommuting or virtual team collaboration technologies is to ask a group of sample users to solve a task collectively. The outcome is measured based on the quality of the solution and the time it took to converge (e.g., [van der Kleij et al., 2009]). Another evaluation technique is to insert the new technology into an existing workflow. Organizational behavior is measured prior to and after the intervention. We used this technique in one of our remote worker studies, detailed in [Tsui et al., 2011a]. We selected six remote participants who had recurring meetings with teammates in Mountain View, CA; the remote participants, located across the United States and Europe, used either a QB or VGo telepresence robot to attend their meetings in place of their normal video conferencing setup. Our pre- and post-experiment questionnaires included 5-point Likert scale [Likert, 1974] team cohesion statements [Michalisin et al., 2004]. These statements, however, would not be appropriate for investigating how telepresence robots affect familial relationships. Our goal is to investigate quantitative and qualitative communication performance measures which are independent of interpersonal relationships and communication task.

Quantitative Measures. Table F.3 summarizes quantitative communication performance measures and provides examples of studies utilizing them. These studies have been drawn from HCI, CSCW, communications, and psychology and look at different communication methods (i.e., face to face (FTF), audio only (AO), video mediated communication (VMC), and embodied video mediated communication (eVMC)). The frequency counts (e.g., number of words, silences, overlaps, handovers, turns, backchannels, gestures) and lengths (e.g., duration of conversation, silences, turns) may be calculated from a recording into speech patterns and speaker segmentation post-hoc coding. Researchers are also investigating real time methods of processing audio signals (e.g., [O'Gorman, 2010]). Fels et al. [Fels et al., 1999] counted the number of successful, partially successful, and failed communications in the PEBBLES (Providing Education By Bringing Learning Environments to Students) telepresence robot project. Kiesler et al. [Kiesler et al., 2008] included a count for correctly recalling information facts after interacting with a robot or robot-like agent.

Qualitative Measures. Open and axial coding from grounded theory [Glaser and Strauss, 1967] can be used to enumerate qualitative data such as observer notes (e.g., [Venolia et al., 2010]) and interviews about the participants' experiences (e.g., [Ding et al., 2007b; Kirk et al., 2010; Lee and Takayama, 2011]). Fish et al. [Fish et al., 1992] looked at the conversational content from face to face and video-mediated interactions. In the PEBBLES project, Fels et al. [Fels et al., 2001] counted behavioral instances, specifically the communication interaction, concentration, and initiative of the remote participant.

Self report scales can provide a means to measure subjective qualitative data. A human-human communication without a medium (or face to face, FTF) 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 [Rubin et al., 2009] for an overview). Witmer and Singer developed the Presence Questionnaire (PQ) to measure personal and social presence in virtual

Table F.4: Select items from Witmer et al.'s Presence Questionnaire [2005]

Question	Scale
How much did the visual aspects of the environment involve you?	not at all / somewhat / completely
How much did the auditory aspects of the environment involve you?	not at all / somewhat / completely
How completely were you able to actively survey or search the environment using vision?	not at all / somewhat / completely
How well could you identify sounds?	not at all / somewhat / completely
How well could you localize sounds?	not at all / somewhat / completely
How closely were you able to examine objects?	not at all / pretty closely / very closely
How well could you examine objects from multiple viewpoints?	not at all / somewhat / extensively

environments [Perse, 2009; Witmer et al., 2005]. The PQ items are rated on a 7-point semantic differential scale. Four subscales have been derived using factor analysis: involvement (α=0.89),[2] sensory fidelity (α=0.84), adaptation/immersion (α=0.84), and interface quality (α=0.57). The involvement and sensory fidelity subscales contain seven items relating to auditory and visual communication which can be applied to telepresence robots shown in Table F.4.

Yarosh and Markopoulos developed the Affective Benefits and Cost of Communication Technologies (ABCCT) to study communication technologies for personal use [Yarosh and Markopoulos, 2010]. They created a simple language version for native English speakers ages 8-10. The ABCCT-child was derived from interviews of parent-child conversations, discussion with social connectedness experts, and an examination of the adult ABC-Q (Affective Benefits and Costs in Communication Questionnaire [IJsselsteijn et al., 2009; Van Baren et al., 2004]). The ABCCT-child investigates the benefits (α=0.88) and costs (α=0.80) of using a communication technology. The questionnaire has 22 items which are rated on a 5-point scale {never, rarely, sometimes, usually, always} [Yarosh and Markopoulos, 2010]. There are four benefits subscales: emotional expressiveness, engagement and playfulness, presence in

[2]Cronbach's alpha measures the internal consistency of related questions and $\alpha>0.7$ is considered reliable [Cronbach, 1951; Nunnaly, 1978].

absence, and opportunity for social support. Three subscales comprise the costs scale: feeling obligated, unmet expectations, and threat to privacy. Unlike the Presence Questionnaire, the ABCCT questionnaire does not explicitly discuss the quality of auditory and visual communication. Instead, it 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. The ABCCT-child questionnaire items are fully detailed in [Yarosh and Markopoulos, 2010].