Towards Measuring the Quality of Interaction: Communication through Telepresence Robots

Katherine M. Tsui, Munjal Desai, and Holly A. Yanco
Department of Computer Science
University of Massachusetts Lowell
One University Avenue, Lowell MA
{ktsui, mdesai, holly}@cs.uml.edu

ABSTRACT

Personal video conferencing is now a common occurrence in long distance interpersonal relationships. Telepresence robots additionally provide mobility to video conferencing, and people can converse without being restricted to a single vantage point. The metrics to explicitly quantify person to person interaction through a telepresence robot do not yet exist. In this paper, we discuss technical requirements needed to support such a communication. We also look at the fields of human-computer interaction (HCI), computer supported cooperative work (CSCW), communications, and psychology for quantitative and qualitative performance measures which are independent of interpersonal relationships and communication task.

Categories and Subject Descriptors

I.2 [**Robotics**]; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Measurement

Keywords

Human-robot interaction, human-computer interaction, embodied video-mediated communication

1. INTRODUCTION

Both video conferencing and telepresence robots are recent technologies. 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 and Skypelaunched in 2003 and 2006,respectively [2,55]. 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 [56]. A number of telepresence robot platforms have emerged in the last decade: InTouch Health's RP-7 in 2003, RoboDynamics' TiLR in 2005,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
PerMIS'12 March 20-22, 2012, College Park, MD, USA
Copyright 2012 ACM 1-4503-1126-7-3/22/12 ...$10.00.

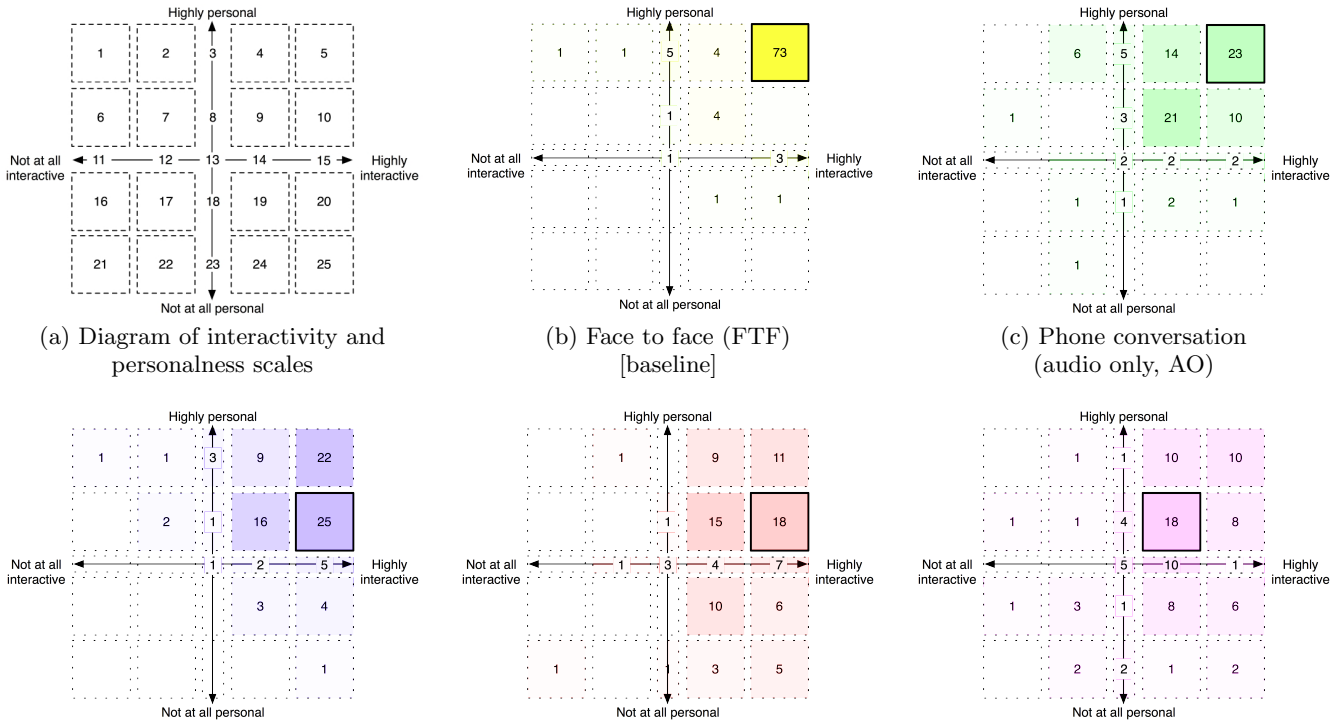
Figure 1: Hugo (an augmented VGo Communication's VGo telepresence robot) is being driven remotely and being used to walk alongside a colleague, actively participating in a mobile conversation. The driver can be seen on Hugo's screen.

HeadThere's Giraffe (now Giraff Technologies AB) in 2006, Willow Garage's Texai (now Suitable Technologies) in 2009, Anybots' QB and VGo Communications' VGo in 2010, and Gostai's Jazz and 9th Sense's TELO in 2011. This mobile video conferencing technology is currently out of the price range for many personal consumers as the platforms range from $6,000 USD for a VGo robot [69] to $5,000 monthly rental fees for an RP-7 [31]. However, we anticipate that in the near future the telepresence robot will become a common household electronic device, much like the personal computer [65].

We believe that telepresence robots can be used to recreate the closeness a remote person would have if he or she were physically present with his or her family and friends better than a telephone or video chat conversation. Hassenzahl 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. [21]

Telepresence robots provide a remote person with a physical avatar in addition to two-way video and audio (Figure 1). For some people, the robot may still be used exclusively as a conversation tool. Other people may want to use telepresence robots to check on their family, while still others may



(a) Diagram of interactivity and personalness scales

(b) Face to face (FTF) [baseline]

(c) Phone conversation (audio only, AO)

(d) Video conferencing (video mediated communication, VMC)

(e) Telepresence robot (TPR)

(f) Instant messaging/chat (IM)

Figure 2: (a) Participants were asked to categorize communication technologies. Original diagram by Jake Knapp of Google; modified to include region enumeration. (b-f) Frequency counts are shown inside each category and the mode is marked by a solid black outline (n=96).

simply want to be present in a space to feel more included in an activity.

Researchers have investigated the efficacy in which people can use telepresence robots to navigate in remote locations (e.g., [40,59,60,62]), the interfaces to do so (e.g., [40,58,61]), and how the robots should be designed (e.g., [8, 11, 12]). Telepresence robots have great potential to provide utility in workplaces (e.g., [37,62]), in schools (e.g., [53]), in homes (e.g., [9]), and for excursions to museums, sporting events, and the theater (e.g., [5]), for example. However, the quality of a person to person interaction through a telepresence robot has not yet been explicitly quantified. In this paper, 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 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). For each means of communication, MTurk Workers were asked where they would place it in Figure 2a with respect to the communication's personalness 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 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 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 2a) for phone conversations (71%), video conferencing (75%), telepresence

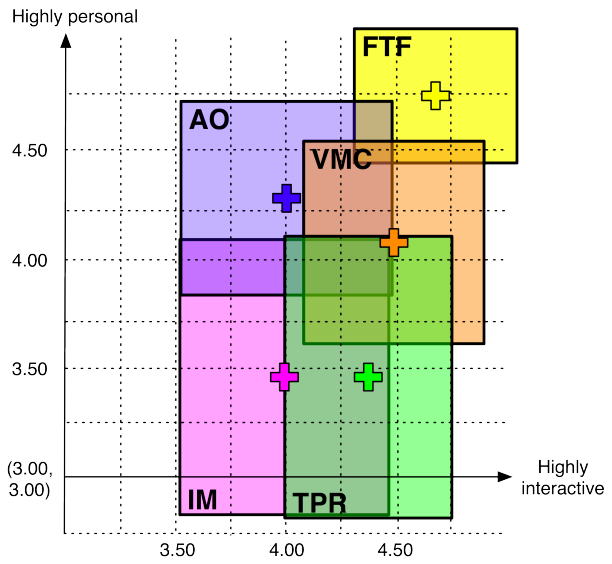


Figure 3: Averages and standard deviations for face to face (FTF), phone call (audio only, AO), video conferencing (video mediated communication, VMC), telepresence robot (TPR), and instant messaging/chat (IM). Plus signs denote averages in the form (interactivity $\bar{I}$, personalness $\bar{P}$), and rectangles denote $\pm$ 1 SD.

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 3 shows the averages and standard deviations for the communication methods. We conducted unpaired t-tests for all of the communication method permutations 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 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.

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 [50]. In terms of the speech fidelity, the audio quality must be comparable at least to that of a landline phone [12]. The ITU-T G.711 Recommendation was initially designed for the Public Switched Telephone Network with 64kbps bandwidth in 1972 [30]. 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 [29]. Rosenberg notes that as the audio fidelity increases, the length of a conversation also increases [50]. 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 is one item of a series of subjective rating questions measuring the quality of speech listed in ITU-T Recommendation P.805 (see Table 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 [27]. MOS can also be derived from simulation tests such as the Perceptual Evaluation of Speech Quality (PESQ) [25].

Speech intelligibility is measured on a 5-point scale the like MOS scale [57]. 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 [7].

Table 1: Subjective evaluation of conversational quality from ITU-T Recommendation P.805 [27]

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

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 [57]. The Speech Transmission Index (STI) predicts nonsensical speech accounting for the speech and noise range, bandwidth, and physical characteristics of the environment [23]. 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 [57]. Barnett and Knight proposed a common intelligibility scale where CIS = 1 + log(STI) [4]. The Speech Intelligibility Index (SII) is similar to STI and also predicts syllabic phonemes in speech [1]. The SII value also ranges between 0 and 1 where values less than 0.45 are poor and above 0.75 are good [57].

Speech intelligibility can also be quantified in terms of the number of echoes, feedback occurrences, and cutouts (e.g., [20, 41]). We designed a study, detailed in [12], 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 1).

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 [67]. 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 [12], to compare the video 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 2).

Table 2: Video characteristics rating questions for comparing QB and VGo telepresence robots and EVO phone used in Desai et al. [12].

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

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) [12]. 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 be subjectively tested, including sample questions regarding an image's color, contrast, borders, movement continuity between frames, flicker, and smearing/blurring [24]. 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 test-

Table 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	[16]		[16]	
Number of words				
• in total	[45]		[19, 45]	
• per participant	[45]	[42]	[42, 45]	
Rate of words over time / percentage dialogue			[19]	[54]
Duration of conversation	[16]		[16, 19]	[54, 66]
Number and/or duration of silences	[32, 52, 64]	[32, 42, 52]	[17, 32, 42, 52, 64]	
Number of overlaps		[42]	[42]	
• simultaneous starts	[45, 52, 64]	[52]	[17, 45, 52, 64]	
• floor holding/disfluencies (e.g., "um," "er")	[45, 52]	[52]	[45, 52]	
• sentence completion	[45, 52]	[52]	[45, 52]	
• interruptions	[45, 52]	[52, 64]	[17, 19, 45, 52, 64]	
Number of explicit handovers (e.g., question, name of next speaker)	[32, 45, 52]	[32, 52]	[32, 45, 52]	
Number of turns (attempts to gain the floor to speak)	[32, 33, 45, 52]	[32, 42, 52, 64]	[19, 32, 42, 45, 52, 64, 68]	[54]
Duration of turn / words per turn	[32, 45, 52, 64]	[32, 52]	[19, 32, 45, 52, 64]	
Distribution of turns	[45, 52]	[52]	[45, 52]	
Number of backchannels				
• verbal (e.g., "mm," "uh huh," "okay")	[32, 45]	[32]	[19, 32, 45]	
• head nod	[32]	[32]	[32]	
• gaze	[33, 48]		[68]	[54, 66]
Number of gestures (i.e., kinetic, spatial, point, other)	[6, 32]	[32]	[32]	

ing of the quality of television pictures [26]. 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) [28].

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., [64]). 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 [62]. 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 team cohesion statements [39]. 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 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., [46]). Fels et al. [15] 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. [34] 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 [18] can be used to enumerate qualitative data such as observer notes (e.g., [66]) and interviews about the participants' experiences (e.g., [13, 35, 37]). Fish et al. [16] looked at the conversational content from face to face and video-mediated interactions. In the PEBBLES project, Fels et al. [14] 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 with-

Table 4: Select items from Witmer and Singer's Presence Questionnaire [70]

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

out 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 [51] for an overview). Witmer and Singer developed the Presence Questionnaire (PQ) to measure personal and social presence in virtual environments [49, 70]. The PQ items are rated on a 7-point semantic differential scale. Four subscales have been derived using factor analysis: involvement (α=0.89),[1] 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 4.

Yarosh and Markopoulos developed the Affective Benefits and Cost of Communication Technologies (ABCCT) to study communication technologies for personal use [71]. 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 [22, 63]). 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} [71]. There are four benefits subscales: emotional expressiveness, engagement and playfulness, presence in 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 [71].

5. APPLICATION OF COMMUNICATION MEASURES

We will conduct a pilot study (n=3) in which people with special needs will operate an augmented VGo telepresence robot Hugo in their families' homes [61]. These participants are students and clients of the Crotched Mountain Rehabilitation Center (CMRC) community; for clarity, we will refer to them as "the participants at CMRC." Our goal is to establish if our target population finds benefit from socially engaging with their families through the telepresence robot as compared to video conferencing. We anticipate that the initial sessions may be subject to a novelty effect from the technologies; in our previous research, we have observed this novelty effect cease within 15 minutes of using a telepresence robot. The person being visited by the participant at CMRC (herein known as "the remote person") will interact with the telepresence robot for two sessions, and the VGo video conferencing software on a laptop for two sessions.

Neither video nor audio of the communication transmitted or received through our telepresence robot will be recorded during our studies. It is important for our participants to understand that our telepresence robot will not record audio or video and thereby ensuring their privacy. The lack of audio and video recording prevents analysis of many of the quantitative communication measures in Table 3. However, we will note the duration of the conversation and the level of conversational success as in Fels et al. 2001 [14]. We will ask both the participant at CMRC and the remote person to recall topics of conversation immediately following the end of the communication as in Kiesler et al. 2008 [34].

After the second use of each technology, we will administer the quality of speech rating questions listed in ITU-T Recommendation P.805 (Table 1), the Presence Questionnaire [70], and the ABCCT questionnaire [71] both to the participant at CMRC and his/her family. Following the completion of all four sessions (two with the robot and two with the laptop), we will conduct interviews based on the events that occurred during the sessions to gauge if the participant at CMRC and his/her family found the telepresence robot and the video conferencing software to be useful.

We will then conduct a longitudinal follow-on study in which participants at CMRC will be loaned our telepresence robot for up to one month each. They will be able to use the telepresence robot whenever they want. Like the pilot study, no audio or video will be recorded given the nature of this study. We will additionally note the frequency of the telepresence robot's use, the duration of the conversations, and the audio and video statistics of each session.

6. CONCLUSIONS AND FUTURE WORK

We have discussed potential quantitative and qualitative performance measures needed to assess the communication portion of the interaction, which are independent of interpersonal relationships and communication task. Further, we have described how the questions from the ITU-T Recommendation P.805, the Presence Questionnaire, and the ABCCT questionnaire will be use in studying the differ-

[1]Cronbach's alpha measures the internal consistency of related questions and α>0.7 is considered reliable [10, 44].

ences between telepresence robots and video conferencing. Interaction through a telepresence robot also includes the concept of presence inherent to the ability of independently moving about a remote space. Researchers have investigated the Temple Presence Inventory [38] and the Oneness Questionnaire [3] to measure presence achieved through robotic telepresence interactions [36, 43, 47]. We believe that items from these scales and the Presence Questionnaire can be added to explicit communication measurements to provide a means to assess the quality of a person to person interaction through a telepresence robot.

7. ACKNOWLEDGMENTS

This research has been funded in part by NSF (IIS-1111125, IIS-0905228, IIS-0546309). We would like to thank Jake Knapp of Google and Elizabeth Craig of North Carolina State University. We also thank Anybots and VGo Communications for loaning us prototype robots. Figure 1 photo by John Fertitta of UMass Lowell.

8. REFERENCES

[1] Amer. Natl. Standards Institute. S3. 5-1997, Methods for the Calculation of the Speech Intelligibility Index, 1997.

[2] Apple. Apple Introduces iChat AV and iSight. Press release, June 2003. `http://www.apple.com/pr/library/2003/jun/23ichat.html`, accessed Sept. 2010.

[3] J. Bailenson and N. Yee. A Longitudinal Study of Task Performance, Head Movements, Subjective Report, Simulator Sickness, and Transformed Social Interaction in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 15(6):699–716, 2006.

[4] P. Barnett and R. Knight. The Common Intelligibility Scale. *Inst. of Acoustics*, 17(7):201–206, 1996.

[5] J. Beer and L. Takayama. Mobile Remote Presence Systems for Older Adults: Acceptance, Benefits, and Concerns. In *Proc. of Intl. Conf. on Human-Robot Interaction*, pp. 19–26. ACM, 2011.

[6] M. Bekker, J. Olson, and G. Olson. Analysis of Gestures in Face-to-Face Design Teams Provides Guidance for How to Use Groupware in Design. In *Proc. of 1st Conf. on Designing Interactive Systems: Processes, Practices, Methods, & Techniques*, pp. 157–166. ACM, 1995.

[7] L. Beranek. Airplane Quieting II: Specification of Acceptable Noise Levels. *Trans. Amer. Soc. Mech. Engrs*, 69:97–100, 1947.

[8] B. Cohen, J. Lanir, R. Stone, and P. Gurevich. Requirements and Design Considerations for a Fully Immersive Robotic Telepresence System. In *Proc. of Human-Robot Interaction Wksp. on Social Robotic Telepresence*, 2011.

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

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

[11] B. Deml. Human Factors Issues on the Design of Telepresence Systems. *Presence: Teleoperators and Virtual Environments*, 16(5):471–487, 2007.

[12] M. Desai, K. M. Tsui, H. A. Yanco, and C. Uhlik. Essential Features of Telepresence Robots. In *Proc. of IEEE Intl. Conf. on Technologies for Practical Robot Applications*. IEEE, 2011.

[13] X. Ding, T. Erickson, W. Kellogg, S. Levy, J. Christensen, J. Sussman, T. Wolf, and W. Bennett. An Empirical Study of the Use of Visually Enhanced VoIP Audio Conferencing: The Case of IEAC. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1019–1028. ACM, 2007.

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

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

[16] R. Fish, R. Kraut, R. Root, and R. Rice. Evaluating Video as a Technology for Informal Communication. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pp. 37–48. ACM, 1992.

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

[18] B. Glaser and A. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Publ., 1977.

[19] D. Grayson and L. Coventry. The Effects of Visual Proxemic Information in Video mediated Communication. *ACM SIGCHI Bulletin*, 30(3):30–39, 1998.

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

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

[22] W. IJsselsteijn, J. Baren, P. Markopoulos, N. Romero, and B. Ruyter. Measuring Affective Benefits and Costs of Mediated Awareness: Development and Validation of the ABC-Questionnaire. *Awareness Systems*, pp. 473–488, 2009.

[23] Intl. Electrotechnical Commission. 60268-16. Objective Rating of Speech Intelligibility by the Speech Transmission Index, Mar. 1998.

[24] Intl. Telecommunication Union. Rec. P.910, Subjective Video Quality Assessment Methods for Multimedia Applications, Sep. 1999.

[25] Intl. Telecommunication Union. 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, Feb. 2001.

[26] Intl. Telecommunication Union. Rec. BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2002.

[27] Intl. Telecommunication Union. Rec. P.805, Subjective Evaluation of Conversational Quality, Apr. 2007.

[28] Intl. Telecommunication Union. Rec. J.247, Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference, Aug. 2008.

[29] Intl. Telecommunication Union. G.711: Pulse Code Modulation (PCM) of Voice Frequencies, Nov. 2009.

[30] Intl. Telecommunication Union. G.729: Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP), Mar. 2011.

[31] InTouch Health. FAQ, 2011. `http://www.intouchhealth.com/ITHFAQs.pdf`, accessed Nov. 2011.

[32] E. Isaacs and J. Tang. What Video Can and Cannot Do for Collaboration: A Case Study. *Multimedia Systems*, 2(2):63–73, 1994.

[33] K. Jokinen, M. Nishida, and S. Yamamoto. Eye-gaze Experiments for Conversation Monitoring. In *Proc. of 3rd Intl. Universal Communication Symp.*, pp. 303–308. ACM, 2009.

[34] S. Kiesler, A. Powers, S. Fussell, and C. Torrey. Anthropomorphic Interactions with a Robot and Robot-like Agent. *Social Cognition*, 26(2):169–181, 2008.

[35] D. Kirk, A. Sellen, and X. Cao. Home Video Communication: Mediating "Closeness". In *Proc. of Conf. on Computer Supported Cooperative Work*, pp. 135–144. ACM, 2010.

[36] A. Kristoffersson, S. Coradeschi, K. Severinson Eklundh, and A. Loutfi. Sense of Presence in a Robotic Telepresence Domain. *Universal Access in Human-Computer Interaction. Users Diversity*, pp. 479–487, 2011.

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

[38] M. Lombard and T. Weinstein. Measuring Presence: The Temple Presence Inventory. In *Proc. of Intl. Wksp. on Presence*, 2009.

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

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

[41] G. Miller and J. Licklider. The Intelligibility of Interrupted Speech. *J. Acoustical Soc. of Amer.*, 1950.

[42] A. Monk and C. Gale. A Look Is Worth a Thousand Words: Full Gaze Awareness in Video-mediated Conversation. *Discourse Processes*, 33(3):257–278, 2002.

[43] D. Nguyen and J. Canny. More than Face-to-Face: Empathy Effects of Video Framing. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pp. 423–432. ACM, 2009.

[44] J. Nunnaly. *Psychometric Theory*. McGraw-Hill, New York, 1978.

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

[46] L. O'Gorman. Latency in Speech Feature Analysis for Telepresence Event Coding. In *Intl. Conf. on Pattern Recognition*, pp. 4464–4467. IEEE, 2010.

[47] B. Okdie, R. Guadagno, F. Bernieri, A. Geers, and A. Mclarney-Vesotski. Getting to Know You: Face-to-Face Versus Online Interactions. *Computers in Human Behavior*, 27(1):153–159, 2011.

[48] K. Otsuka, J. Yamato, Y. Takemae, and H. Murase. Quantifying Interpersonal Influence in Face-to-Face Conversations Based on Visual Attention Patterns. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1175–1180. ACM, 2006.

[49] E. Perse. *Presence Questionnaire*, pp. 276–283. Routledge, Taylor & Francis, 2009.

[50] J. Rosenberg. Quality Matters, Aug. 2010. `http://www.tmcnet.com/ucmag/columns/articles/99344-quality-matters.htm`, accessed Nov. 2011.

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

[52] A. Sellen. Remote Conversations: The Effects of Mediating Talk with Technology. *Human-Computer Interaction*, 10(4):401–444, 1995.

[53] K. Sheehy and A. Green. Beaming Children Where They Cannot Go: Telepresence Robots and Inclusive Education: An Exploratory Study. *Ubiquitous Learning*, 3(1):135–146, 2011.

[54] D. Sirkin, G. Venolia, J. Tang, G. Robertson, T. Kim, K. Inkpen, M. Sedlins, B. Lee, and M. Sinclair. Motion and Attention in a Kinetic Videoconferencing Proxy. *Human-Computer Interaction*, pp. 162–180, 2011.

[55] Skype. Skype Introduces Video Calling for Macintosh Users. Press release, Sept. 2006. `http://about.skype.com/2006/09/skype_introduces_video_calling.html`, accessed Sept. 2010.

[56] Skype S.A. Amendment No. 2 to Form S-1 Registration Statement. Prospectus, Mar. 2011. `http://www.sec.gov/Archives/edgar/data/1498209/000119312511056174/ds1a.htm#rom83085_3a`, accessed Nov. 2011.

[57] H. Steeneken. Standardisation of Performance Criteria and Assessments Methods for Speech Communication. In *European Conf. on Speech Communication and Technology*, vol. 1, pp. 255–258, 2006.

[58] L. Takayama, E. Marder-Eppstein, H. Harris, and J. Beer. Assisted Driving of a Mobile Remote Presence System: System Design and Controlled User Evaluation. In *IEEE Intl. Conf. on Robotics and Automation*, pp. 1883–1889, 2011.

[59] T. Tsai, Y. Hsu, A. Ma, T. King, and C. Wu. Developing a Telepresence Robot for Interpersonal Communication with the Elderly in a Home Environment. *Telemedicine and e-Health*, 13(4):407–424, 2007.

[60] K. Tsui, M. Desai, H. Yanco, and C. Uhlik. Telepresence Robots Roam the Halls of My Office Building. In *Proc. of Human-Robot Interaction Wksp. on Social Robotic Telepresence*, 2011.

[61] K. Tsui, A. Norton, D. Brooks, H. Yanco, and D. Kontak. Designing Telepresence Robot Systems for Use by People with Special Needs. In *Proc. of Intl. Symp. on Quality of Life Technologies 2011: Intelligent Systems for Better Living, held in conjunction with RESNA 2011 as part of FICCDAT*, 2011.

[62] K. M. Tsui, M. Desai, H. A. Yanco, and C. Uhlik. Exploring Use Cases for Telepresence Robots. In *Proc. of Intl. Conf. on Human-Robot Interaction*. ACM, 2011.

[63] J. Van Baren, W. IJsselsteijn, P. Markopoulos, N. Romero, and B. de Ruyter. Measuring Affective Benefits and Costs of Awareness Systems Supporting Intimate Social Networks. In *CTIT Wksp. Proc. Series*, vol. 2, pp. 13–19, 2004.

[64] R. van der Kleij, J. Maarten Schraagen, P. Werkhoven, and C. De Dreu. How Conversations Change over Time in Face-to-Face and Video-mediated Communication. *Small Group Research*, 40(4):355–381, 2009.

[65] V. Venkatesh and S. Brown. A Longitudinal Investigation of Personal Computers in Homes: Adoption Determinants and Emerging Challenges. *MIS Quarterly*, pp. 71–102, 2001.

[66] G. Venolia, J. Tang, R. Cervantes, S. Bly, G. Robertson, B. Lee, and K. Inkpen. Embodied Social Proxy: Mediating Interpersonal Connection in Hub-and-Satellite Teams. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1049–1058. ACM, 2010.

[67] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt. Eye Gaze Patterns in Conversations: There Is More to Conversational Agents than Meets the Eyes. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*, pp. 301–308. ACM, 2001.

[68] R. Vertegaal, G. van der Veer, and H. Vons. Effects of Gaze on Multiparty Mediated Communication. In *Graphics Interface*, pp. 95–102, 2000.

[69] VGo Communications. VGo Communications. Webpage, 2010. `http://vgocom.com`, accessed Oct. 2010.

[70] B. Witmer, C. Jerome, and M. Singer. The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators & Virtual Environments*, 14(3):298–312, 2005.

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