Realizing Hinokio: Candidate Requirements for Physical Avatar Systems

Laurel D. Riek
The MITRE Corporation
7515 Colshire Drive
McLean, VA USA
laurel@mitre.org

ABSTRACT

This paper presents a set of candidate requirements and survey questions for physical avatar[1] systems as derived from the literature. These requirements will be applied to analyze a fictional, yet well-envisioned, physical avatar system depicted in the film *Hinokio*. It is hoped that these requirements and survey questions can be used by other researchers as a guide when performing formal engineering tradeoff analysis during the design phase of new physical avatar systems, or during evaluation of existing systems.

Categories and Subject Descriptors

I.2.9 [**Artificial Intelligence**]: Robotics—*Operator interfaces*; H.4.3 [**Information Systems Applications**]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*

General Terms

Design, Human Factors

Keywords

Collaboration, Human-Robot Interaction, Physical Avatars, Requirements, Tele-embodiment

1. INTRODUCTION AND RELATED WORK

In today's highly globalized and mobile world, people are frequently expected to collaborate with team members across great distance. Much technology has been developed to help address this, such as video teleconferencing, smart team rooms, and shared whiteboards [12]. Unfortunately, most of these tools are insufficient in providing all the communication modalities present in face-to-face communication,such as gesturing in shared space and other nonverbal cues [5].

Several hybrid solutions have been proposed that incorporate video into a shared virtual space such as Augmented Reality, Shared Reality, and Virtual Rooms [16, 4, 6]. Further, 3D virtual spaces known as Immersive Environments allow users to manipulate objects and collaborate across distance in a completely untethered way (i.e., no head-mounted displays needed) [1]. However, none of these solutions provide sufficient workspace awareness [8], because they restrict users to only using the elements available to them in the virtual space, and confine users to only meeting in specified locations.

In Robotics many researchers have recognized the need for increased mobility and "real world" interaction when performing human-human distance collaboration. Hence, several physical avatar systems with two-way video have been developed to address this need. The Personal Roving Presence system developed at UC Berkeley is a teleoperated mobile robot that provides a video depiction of the face of its remote operator and allows for primitive gesture [18]. Researchers at University of Chicago developed the AccessBot, which is a system that uses a wheeled, life-sized display screen depicting the entire upper torso of a remote collaborator, providing a strong virtual presence for disabled meeting participants [15]. InTouch Health developed the RP-7 (Remote Presence Robotic System) which is a mobile robot that displays the face of a remotely located physician, used to remotely examine patients in UCLAs Intensive Care Unit [26] The BiReality System, developed at HP Labs, is a life-sized, mutually immersive teleoperated robot surrogate that features a 360-degree surround projection display cube [14]. Finally, there have been several efforts looking at androids that resemble humans [11, 24], but they are not usually described as being used for human-human distance collaboration.

It is unclear from the literature how collaboration is affected by the use of such systems because the focus of the research has been on designing and developing the technology, and of the user studies discussed, most are anecdotal. All of these systems claim to provide an improvement in human-human distance collaboration, but exactly *how* they affect collaboration is unknown.

The way humans interact with one another using a physical avatar system is different from how they interact using more traditional collaborative systems. This difference is due to the fact that a physical manifestation of a person

[1] A *physical avatar* system shall be defined as a tele-operated mobile robot that serves as a physical manifestation of a remote user. The robot will display at least a facial physical resemblance to the user, typically via transmitted video.

HRI'07, March 8-11, 2007, Arlington, Virginia, USA.

places a new accuracy burden on the technology. In addition to conveying the correct visual and verbal attributes of a person, physical avatars must also accurately convey non-verbal affect (e.g., gesture and movement) in order to remain true to the user's communicative intent.

2. CANDIDATE REQUIREMENTS

Given the engineering and interaction complexity of physical avatar systems, it can be a daunting task for designers to create systems in that remain true to the users' communicative intent. Therefore, this paper proposes a set of literature-derived candidate requirements to be used as a guide throughout this process. The requirements have been divided into seven areas: Video, Camera, Control, Latency, Gaze and Appearance, Audio, and Gesture. Each area is described below, and a summary of all areas is provided as reference in Table 1.

We will assume two things when specifying these candidate requirements. First, for the purposes of simplicity, we assume there will be only one robot-local collaborator (RLC), and one robot-remote collaborator (RRC) present in the collaboration. Second, we will assume a physical avatar system that has features similar to those previously described in the literature: two-way audio, video of the face of the robot-remote collaborator (RRC) displayed on the avatar, video of the area around the avatar transmitted to the RRC, and physical mobility of the avatar via RRC directed commands.

2.1 Video

When transmitting video the system shall:

- *Prevent image distortion*
- *Prevent motion artifacts*
- *Preserve color*
- *Provide visual continuity during times of lag*

Accurate and timely video transmission will help the RLC and RRC feel more like they are communicating face-to-face. Therefore, it is extremely important that people and objects appear as realistic as possible by preventing motion artifacts, preserving color, and preventing image distortion. This requirement is motivated by the result Jouppi et. al. showed in [13].

With regards to latency, video lag will be inevitable, particularly on bandwidth-limited networks. Consequently, a means for visual continuity should be implemented to ensure minimal disruption. Leigh et. al. took measures to overcome this problem when creating the AccessBot [15].

2.2 Camera

The system's camera shall:

- *Provide views to the RRC that closely mimic being physically present, such as wide angle or 360 degrees.*

A wide-angle or 360-degree view of the world will provide greater situational awareness to the RRC. There is a great body of literature in general to support this requirement, but in particular for physical avatar systems it is supported by [13, 18].

2.3 Control

For RRC-issued control the system shall:

- *Permit full mobility*
- *Permit full pan/tilt/zoom camera control*
- *Permit height control*

Given the physical avatar is representing the RRC, it should allow that person all the same mobility and visual field freedoms they would enjoy were they collaborating with colleagues in person.

The RRC should have the ability to adjust their height to "stand" or "sit" as necessary in order to have a more realistic interaction with the RLC. Height disparity between the physical avatar and the RLC was so significant in the first BiReality System that Jouppi et. al. completely redesigned their robot to allow the RLC full height control [14].

2.4 Latency

When the RRC sends teleoperation commands the system shall:

- *Minimize bandwidth latency to be less than 125 ms.*

Given the goal is to mimic in-person communication as much as possible, any gesture, movement action, or camera view change should occur very soon after the RRC transmits the command. Hannaford and Sheridan did some of the foundation work on tolerable bandwidth latency for users operating mobile robots, and found a maximum tolerability limit of 125 ms per command issued [9, 22].

2.5 Gaze and Appearance

When representing the RRC the system shall:

- *Preserve gaze*
- *Portray clear facial appearance and expression*

When portraying a human face it is important to clearly depict expression and gaze, as these are critical aspects for effective communication. This is supported by a wide body of literature, including [2, 10, 20, 23].

2.6 Audio

When transmitting audio the system shall:

- *Provide background-noise detection to the RRC*

A great deal of communication cues can be garnered from background noise in the environment. Paulos et. al. described an unexpected result of providing quality audio in their physical avatar system - RRCs were able to gauge the mood of a room based on perceived subtle background noises around the robot [19].

2.7 Gesture

If the RRC requires the ability to gesture the system shall:

- *Provide at minimum a two degree-of-freedom mechanism for deictic gesture*
- *Ensure the RRC and RLC adequately share perspectives.*

Pointing is one of the most fundamental aspects of human communication. It readily allows for language disambiguation and shared perspective. The requirement that a two degree-of-freedom mechanism for deictic gesture is supported by Brooks and Paulos [3, 18]. However, one should be cautious when designing tele-gesture mechanisms because the greater the degrees of freedom the harder it will be for RRCs to control.

Ensuring that perspective is shared adequately between the RRC and RLC is motivated by Galinsky and Trafton [7, 25]. While speech can also be used to resolve ambiguities when sharing perspective, using gesture to do so more closely resembles in-person collaboration.

3. GENDANKEN EXPERIMENT

At the time this paper was written, no end-to-end physical avatar system was available to the author on which to evaluate our candidate requirements. Instead, it was decided to analyze a fictional, yet well-envisioned, physical avatar system from the Japanese film *Hinokio*. The film is about about a shy, 12-year-old boy named Satoru who is physically disabled and does not wish to attend school. His father, a Roboticist, builds him a bipedal, humanoid, remote-controlled robot named Hinokio (See Figure 1). Using an immersive environment, Satoru controls Hinokio from his bedroom and sends the robot to school in his place.

Working in a fictional universe, the filmmakers were free to create an end-to-end physical avatar system that was bug-free, bandwidth unlimited, fully mobile, and easily controlled. But the system did have some notable limitations, such as making Hinokio appear like a robot instead of like Satoru. We will briefly examine each requirement in the context of how the physical avatar system was presented in the film.

3.1 Video and Camera

Using his workstation, Satoru has a fully immersive view of the world (See Figure 2). The video he sees is provided by Hinokio's camera, which has a wide-angle lens. Occasionally the view is distorted, particularly when the robot is moving quickly. Given Hinokio is intended to represent a 12-year old (who are usually quite active), image stabilization would be quite useful.

3.2 Control

Hinokio is a bi-pedal humanoid robot with full arm, head, and leg articulation, as well as complete, two-handed manipulation capability. Satoru controls Hinokio's legs using a joystick and head through a roll/pitch/yaw motion capture device. It is unclear how such precise arm and hand manipulation was accomplished; the filmmakers must have realized the difficulty in haptic interface creation for high degree-of-freedom manipulators. Regardless, the manipulation seemed to carry a high learning curve; Satoru accidently punched one of his friends harder than he had intended. Adding force-feedback control to the system or using a different interface modality could help mitigate such problems.

3.3 Latency

At one point in the film Satoru is upset, and decides to forego his daily ritual of plugging-in Hinokio to charge. Hence, the robot "dies" and no longer responds to commands. Fortunately Satoru chose a reasonable place for the robot's demise; it was inside a deserted building. However, were this sort of misuse to happen in a real life situation, one might worry that when the robot loses communication or power it could fall on top of someone. Therefore, builtin safety mechanisms are of upmost importance to mitigate such circumstances of lost connectivity.

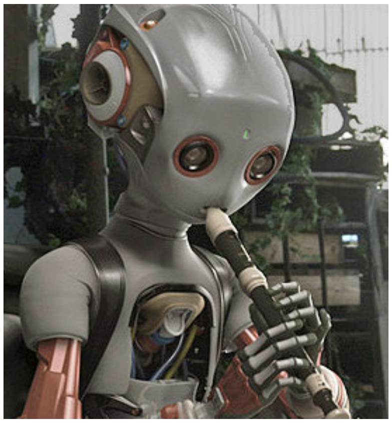

Figure 1: The avatar is quite dexterous, shown here playing a flute. ***Image © 2005 Hinokio Film Venturer***

3.4 Gaze and Appearance

Here the system is lacking: Hinokio does not represent Satoru's facial expressions nor his likeness at all. (However, gaze is preserved due to Hinokio's eyes and head being able to move to view objects). Interestingly, the lack of likeness seems to be sufficient for collaboration. The students interacting with Hinokio eventually begin to anthropomorphize, which is consistent with the literature ([11] and [21]). Though it seems in some cases this anthropomorphizing is inaccurate; the students occasionally attribute personality traits to Hinokio that Satoru does not truly have.

3.5 Audio

Audio is another unusual design decision on part of the filmmakers - Hinokio's "voice" sounds nothing like Satoru's, and is instead is rather mechanical sounding. It is possible this design decision was to preserve privacy, but its most likely effect is a lack of trust among people interacting with the robot. Indeed, when Satoru first attends class as Hinokio and introduces himself he is at once teased by his classmates, probably because they do not realize that Hinokio is actually an avatar. While it may not be technologically feasible to create a physical likeness of the RRC, one should always at least aim for vocal likeness.

3.6 Gesture

Overall, Hinokio adequately conveys Satoru's intended gestures. Occasionally there are times of ambiguity, but they are resolved verbally.

4. DISCUSSION

Presently, no near-term system meets all of the proposed requirements, so it will be necessary for those designing new physical avatar systems to perform engineering tradeoff analysis to determine which requirements are most important for the specific physical avatar being constructed. The survey questions presented in Table 2 can help guide such an analysis. For example, if building a physical avatar system to act as a surgical aid, latency and control would likely be

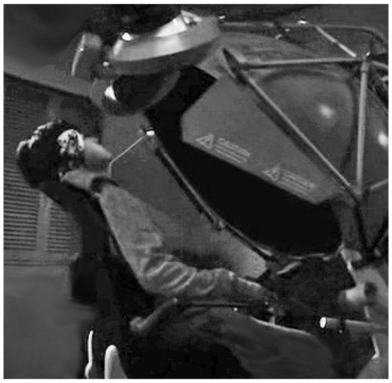

Figure 2: Satoru's interface for control. He is visually immersed in the remote environment when looking at the hemispheric display. Satoru's head has a roll/pitch/yaw motion capture device. ***Image © 2005 Hinokio Film Venturer***

given much greater priority than gesture and appearance. See [17] (and its references) for more detailed instructions on how to perform requirements prioritization.

When analyzing interaction between humans using an existing physical avatar system or designing a new one, it is important that both sides of the collaboration are given equal consideration. Furthermore, it is likely that the functional and aesthetic requirements for the RRC will differ from those of the RLC. For example, the RLC might require loud speakers and high-gain microphones if the robot is to be situated in a noisy environment, whereas the audio needs of RRC may be fulfilled by an inexpensive, off-the-shelf headset. For physical avatar systems, dual contextual design is of the utmost importance in order to facilitate the best collaborative experience.

5. REFERENCES

[1] Y. Boussemart, F. Rioux, F. Rudzicz, M. Wozniewski, and J. Cooperstock. A Framework for Collaborative 3D Visualization and Manipulation in an Immersive Space using an Untethered Bimanual Gestural Interface. In *Virtual Reality Systems and Techniques*, November 2004.

[2] C. Breazeal. How to Build Robots That Make Friends and Influence People. In *Proceedings of the 1999 IEEE International Conference on Intelligent Robots and Systems*. IEEE, 1999.

[3] A. Brooks and C. Breazeal. Working with Robots and Objects: Revisiting Deictic Reference for Achieving Spatial Common Ground. In *Proceedings of the 2006 ACM Conference on Human Robot Interaction*. ACM, 2006.

[4] J. R. Cooperstock. Interacting in Shared Reality. In *Conference on Human-Computer Interaction*. HCI International, July 2004.

[5] J. T. Costigan. A Comparison of Video, Avatar, and Face-to-Face In Collaborative Virtual Learning Environments. Master's thesis, University of Illinois, 1997.

[6] D. deMoulpied and T. Aiken. Room-Based Multimodal User Interface System. *The MITRE Corporation Technology Transfer Office*, July 2003.

[7] A. Galinsky and G. Ku. The effects of perspective-taking on prejudice: the moderating role of self-evaluation. *Personal Social Psychology Bulletin,*, 30:594–604, 2005.

[8] C. Gutwin and S. Greenberg. The Importance of Awareness for Team Cognition in Distributed Collaboration. In *Team Cognition: Understanding the Factors that Drive Process and Performance*, pages 177–201. APA Press, 2004.

[9] B. Hannford. Ground Experiments Toward Space Teleoperation with Time Delay. In *Teleoperation and Robotics in Space*, chapter 4, pages 87–106. AIAA, 1994.

[10] R. Hassin and Y. Troupe. Facing Faces: Studies on the Cognitive Aspects of Physiogomy. *Journal of Personality and Social Psychology*, 78(5):837–852, 2000.

[11] H. Ishiguro. Interactive Humanoids and Androids as Ideal Interfaces for Humans. In *Proceedings of the 2006 ACM Conference on Intelligent User Interfaces*. ACM, 2006.

[12] H. Ishii, M. Kobayashi, and J. Grudin. Integration of Inter-personal Space and Shared workspace: ClearBoard Design and Experiments. In *Proceedings of the 1999 ACM Conference on Computer Supported Cooperative Work*. ACM, 1999.

[13] N. Jouppi, N. Iyer, S.Thomas, and A. Slayden. BiReality: Mutually-Immersive Telepresence. In *Proceedings of the ACM Conference on MultiMedia*. ACM, 2004.

[14] N. Jouppi and S. Thomas. Telepresence Systems with Automatic Preservation of User Head Height, Local Rotation, and Remote Translation. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, 2005.

[15] J. Leigh, M. Rawlings, J. Girado, G. Dawe, R. Fang, M.-A. Khan, A. Cruz, D. Plepys, D. J. Sandin, and T. A. DeFanti. AccessBot: An Enabling Technology for Telepresence,. In *Proceedings of the The 10th Annual Internet Society Conference*. INET, 2000.

[16] P. Liu, N. Georganas, and P. Boulanger. Designing Real-Time Vision Based Augmented Reality Environments for 3D Collaborative Applications. In *Canadian Conference on Electrical and Computer Engineering*. IEEE, May 2002.

[17] N. Mead. Requirements Prioritization Introduction. *Software Engineering Institute, Carnegie Mellon University*, 2006.

[18] E. Paulos and J. Canny. Designing for Personal Tele-embodiment. In *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*. IEEE, 1998.

[19] E. Paulos and J. Canny. PRoP: Personal Roving Presence. In *Proceedings of the 1998 SIGCHI conference on Human Factors in Computing Systems*. ACM, 1998.

[20] A. Powers and S. Kiesler. The Advisor Robot: Tracing Peoples Mental Model from a Robots Physical Attributes. In *Proceedings of the 2006 ACM Conference on Human Robot Interaction*. ACM, 2006.

[21] B. Robins, K. Dautenhahn, R. Bockhorst, and A. Billard. Robots as Assistive Technology - Does Appearance Matter? In *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*. IEEE, 2004.

[22] T. Sheridan. Space Teleoperation through Time Delay: Review and Prognosis. *IEEE Transactions on Robotics and Automation*, 9(5):592–606, 1993.

[23] R. Stiefelhagen and J. Zhu. Head Orientation and Gaze Direction in Meetings. In *Proceedings of the 2002 Conference on Human Factors in Computing Systems*. ACM, 2002.

[24] I. Toshima, H. Uematsu, and T. Harahara. A Steerable Dummy Head That Tracks Three-Dimensional Head Movement: TeleHead. *Acoustical Science and Technology*, 24(5):327–329, 2003.

[25] G. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, and A. Schultz. Enabling Effective Human-Robot Interaction Using Perspective-Taking in Robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 35(4), 2005.

[26] P. Vespa. Robotic Telepresence in The Intensive Care Unit. *Critical Care*, 9(4):319–320, 2005.

6. APPENDIX

Name	**Requirement**
Video	When transmitting video the system shall: • Prevent image distortion • Prevent motion artifacts • Preserve color • Provide visual continuity during times of lag
Camera	The system's camera shall: • Provide views to the RRC that closely mimic being physically present, such as wide angle or 360 degrees
Control	For RRC-issued control the system shall: • Permit full mobility • Permit full pan/tilt/zoom camera control • Permit height control
Latency	When the RRC sends teleoperation commands the system shall: • Minimize bandwidth latency to be less than 125 ms
Appearance and Gaze	When representing the RRC the system shall: • Preserve gaze • Portray clear facial appearance and expression
Audio	When transmitting audio the system shall: • Provide background-noise detection to the RRC
Gesture	If the RRC requires the ability to gesture the system shall: • Provide at minimum a two degree-of-freedom mechanism for deictic gesture • Ensure the RRC and RLC adequately share perspectives.

Table 1: *Requirements.* These are candidate physical and functional requirements for physical avatar systems. The requirements assume that only one RLC and one RRC are participating in the collaboration and that the physical avatar system has features similar to those described in the literature.

Requirement Name	**Survey Questions**
Video	How accurately are the respective collaborators represented? How frustrating is it for the RRC to view a distorted image? Does the system freeze in times of lag?
Camera	Can the RLC and RRC see enough of one another's respective worlds in order to effectively collaborate on shared spatial tasks?
Control	How often does the RRC require help from the RLC when performing tasks? Is the RLC able to "look" the RRC in the eye?
Latency	What happens to the avatar when bandwidth latency is high? How well does the system recover from network-dropped commands?
Appearance and Gaze	Can the RRC turn to face the RLC as easily as if they were in person? Is the RLC able to detect when the RRC is expressing agreement?
Audio	What level of sound can the RRC hear (fingers snapping, foot falls, etc)? What kinds of sounds are important for the collaboration task but are not being "heard"? When the RRC "speaks" through the avatar, does the voice sound identical to being transmitted over a telephone?
Gesture	Is the object or location the RRC points to the correct one? Is the RLC able to interpret the RRC's gestures?

Table 2: *Survey Questions.* These questions are intended to help guide designers and engineers during the tradeoff analysis phase of building new physical avatar systems. Furthermore, if one is evaluating existing physical avatar systems, these questions can be used as a starting point for experimental design.