Designing Personal Tele-embodiment

Eric Paulos* paulos@cs.berkeley.edu

John Canny* jfc@cs.berkeley.edu

Department of Electrical Engineering and Computer Science
University of California, Berkeley

Abstract

At the intersection of tele-robotics, computer networking, and human social interaction we have chosen to explore an area we identify as personal tele-embodiment. At the core of this research is an emphasis on the individual person rather than the intricate complexities of the machine. While the mechanical elements of our system are essential to its overall functionality, our research is driven solely by the study and understanding of the social and psychological aspects of extended human-human interactions rather than the latest techno-gadgetry. In this paper we emphasize the importance of the human component and describe the development of one such simple, inexpensive, internet-controlled, untethered tele-robot or PRoP *(Personal Roving Presence) that provides several fundamental elements of personal tele-embodiment.*

1 Introduction

Only a few decades ago computers were being praised solely on their ability to tackle complex mathematical problems with little discussion of future applications beyond their then use as sophisticated military and research laboratory calculating engines. Clearly, the computers of today have evolved and assimilated themselves into the daily lives of countless people in ways that were never imagined. Similarly, robotics research over the last few decades has witnessed a myriad of reveling contributions to science and society. While giving proper praise to these contributions, we propose an augmentation to current robotics research that may result in the extension of robotics into the lives of ordinary people in a manner similar to the transition of computers from laboratories to personal homes and bodies [14].

1.1 Human Centered Robotics

Our research ideology is in the spirit of the recently identified area of "human centered robotics" and our approach to problems often share many themes with work in this field. Our conjecture is that by observing humans in their everyday lives, away from mechanisms and automation, we can learn valuable insights into the social and psychological aspects of their existence and interactions. These studies will in turn motivate the formulation of useful, and hopefully successful, new applications for robotics researchers to address. We expect to discover new applications that have traditionally fallen outside of what is viewed as the robotics field of study.

In this paper we make no authoritative claims as to the correct method of this approach nor do we propose a general solution to this problem, as there are likely many. Instead we concentrate on the design of one such system whose goal is to enable personal telepresence. Our belief is that from a brief discussion of the human centered design choices that encompass this project, an emerging human centered theme will dominate this paper.

1.2 Personal Tele-embodiment

Our intention is to provide telepresence[1] to ordinary people in an intuitive and personal manner. In keeping with our research paradigm, we focus not on the mechanical elements of the system but on the choice and implementation of specific skills that empower humans to explore and interact at a distance. We do however include some discussion of the mechanical and robotic components in the design.

Succinctly, we are interested in identifying and distilling a small number of human behavioral traits or skills that are inherent to human communication, understanding, and interaction. We will attempt to implement these traits on intuitive human-interfaced, networked, mechanical systems. The ultimate goal is to provide a reasonable degree of personal telepresence that allows humans to communicate and interact in a useful manner with remote people and places in ways beyond those available with current systems.

Our claim is that such systems can be built now, at minimal cost, and provide powerful new metaphors in mediated human-human communication. Since this area has many near-term applications we expect that researchers will be able to explore a wide variety of techniques for personal tele-embodiment.

*This research was financially supported by BMDO grant number N00014-96-1-1200 administered by ONR.

[1] More specifically we are referring to personal tele-embodiment, tele-robotics, or tele-action. This is to avoid the ambiguity caused by the term *telepresence* which has grown in recent years to describe not only systems involving distant real spaces (*i.e.* tele-robotics) but also distant virtual spaces or VR.

2 Previous and Related Work

Methods of achieving telepresence[2] are not new with one of the first electrically controlled mechanical teleoperational systems being developed by Goertz [8] in 1954. Since then a variety of applications for tele-operated robotics have been explored by numerous researchers. Space does not permit us to include a survey and hence we defer the reader to Sheridan [18]. Most of these system were designed for a single specific task and are quite complex. They also typically require special purpose dedicated hardware and a highly trained operator to control and interact with the mechanism in the remote environment. In our system we strived to constrain its development so that it would be accessible to a wide audience without additional, expensive, or extraordinary hardware. In essence, telepresence for the masses.

The exponential growth of the WWW over the past several years has resulted in a plethora of remote controlled mechanical devices which can be accessed via the WWW. Goldberg [9] developed a 3 DOF (Degree Of Freedom) telerobotic system where users were able to explore a remote world with buried objects and, more interestingly, alter it by blowing bursts of compressed air into its sand filled world. Soon afterwards, we developed Mechanical Gaze [16], a tele-robotic system where uses could control a camera's viewpoint and image resolution to observe various museum artifacts placed within the robot's workspace. By 1995, Goldberg had developed another telerobotic system called the TeleGarden [10] in which WWW users are able to observe, plant, and nurture life within a living remote garden.

Others have also argued for a human centered approach to robotics. As Asada pointed out in his 1997 ICRA workshop entitled "Human Centered Robotics," there is a overwhelming need to direct robotics research towards the needs of ordinary people such as human health care, home medicine, and enhanced communication between people. Similar views were expressed by many of the other panel members [1]. This is also a growing body of research into ubiquitous telepresence [4] and human centered robotic designs such a Peshkin's Cobots [3].

Social and psychological aspects of extended human-human interactions motivate the design of our PRoPs and we have identified a wide range of research in this area. Shared spaces and human interaction with video walls such as the VideoWhiteboard [20] designed at Xerox PARC and later Ishii's ClearBoard [11] are fundamental to designing usable PRoPs. We are also interested in the use of video in tele-connecting individuals which has been nicely explored by Kraut and Fish [12; 7] and others [6]. We have also been motivated by Steuer's [19] discussion of the dimensions of telepresence.

[2] "To convey the idea of these remote-control tools, scientists often use the words *teleoperators* or *telefactors*. I prefer to call them *telepresences*, a name suggested by my futurist friend Pat Gunkel." [15]

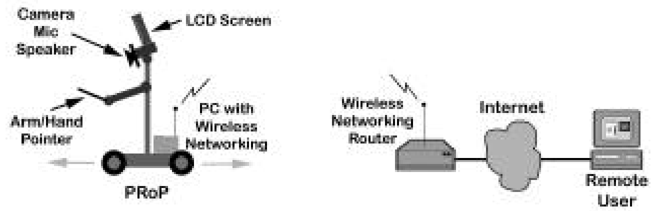

Figure 1: System overview of a typical PRoP hardware configuration.

3 PRoP: Personal Roving Presence

A PRoP is simple, inexpensive, internet-controlled, untethered tele-robot that provides the sensation of tele-embodiment in a remote real space. The first PRoPs were simple helium-filled blimp airborne tele-robots called *space browsers* [17]. However, in this paper we have chosen to focus on more recently developed terrestrial four-wheeled PRoPs.

3.1 Basic Layout

Terrestrial PRoPs, sometimes referred to as *surface cruisers* or *carts*, are designed from simple, inexpensive remote-control vehicles with modifications to slow them to human walking pace and a 1.5 meter vertical pole to provide a realistic human vantage for the camera. On-board the PRoP is a color video camera, microphone, speaker, color LCD screen, a few simple custom electronics, and various drive and servo motors. The basic layout for the system is shown in Figure 1. Unlike the blimps, these PRoPs can travel outdoors, require less maintenance, and provide much longer battery life. They also carry a complete PC on-board with wireless networking hardware attached. Furthermore, we leverage off of wireless communication infrastructures already in existence, greatly extending the inhabitable world of PRoPs. A recently designed PRoP is shown in Figure 2.

3.2 User Control

A user, anywhere on the internet, can use a simple Java applet running within a Java-enabled browser to control the PRoP. As they guide the PRoP forward, backwards and left, right it delivers, via wireless communications, live video and audio to the remote operator's computer through standard free tele-conferencing software that runs on standard personal computers. The remote operator observes the real world from the vantage of the PRoP while listening to the sounds and conversations within close proximity to

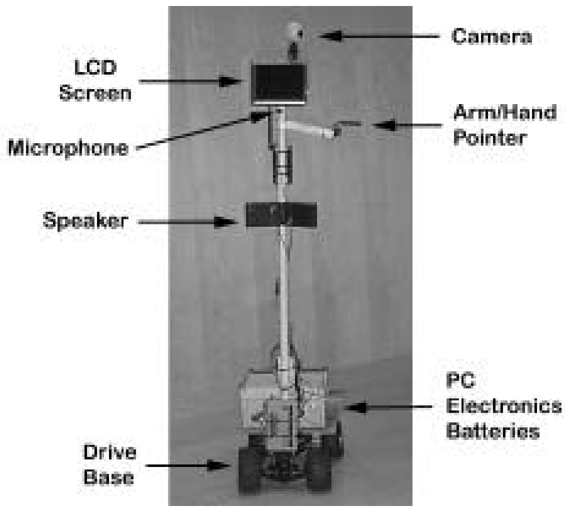

Figure 2: A PRoP with camera head, video LCD screen, controllable "arm/hand" pointer, microphone, speakers, and drive-able base.

it. The user converses with groups and individuals by simply speaking into the microphone connected to their desktop or laptop computer, the sound delivered via the internet and then a wireless link to the PRoP's on-board speaker.

4 Human Centered Design

We should stress that PRoP design choices were based largely on the study and observation of *people* in their daily lives rather than on investigations into elaborate hardware. It is this methodology that we hope to emphasize in this paper and thus we elaborate on its role in this section.

Initially we observed that many frequently occurring activities in our daily lives are not captured or conveyed by modern technology. Despite the advances of our tele-connected world of telephones, pages, cellular phones, and internet communications, we noticed that many subtle, yet extremely important elements of human communication and interaction such as atmosphere, morale, chaos, *etc.* were missing from the experience.

Most people still shop by wandering the shelves, looking for specials, seeing the item they want, and asking about its features. We wander hallways with chance encounters with people and objects playing a significant roll in our daily lives. Our social interactions are variegated, and we spontaneously move from talking to one individual to another, to a group, to another group, *etc.* In all these activities, our senses, our mobility, and our situated physical form play essential roles. However, our current technological communication channels are far too structured to capture these important nuances. In the following subsections we trace the human centered robotics methodology employed in the evolution of PRoPs.

4.1 Aural

Sound is one of the most elementary and obvious methods of human communication. Therefore, the PRoP design includes a two-way, full-duplex audio channel that allow users to engage in remote conversations. One unexpected result of studying people using this audio feature was the importance of background "noise" near the PRoP. The experience of using the PRoP was noticeably more compelling when users were able to gauge the general mood of the remote location by receiving a variety of subtle aural cues such as doors opening, elevators arriving, people approaching, nearby conversations, music playing, automobile traffic, wind blowing, *etc.*

4.2 Visual

Despite the horrific failure of the Picturephone of the 1960's it is clear that recent improvements in speed, resolution, and miniaturization, have made video a viable and useful channel for human communication. Although video may add little useful information to a telephone conversations between people, the visual appearance of a remote location (color, shape, size, occupancy, lighting, *etc.*) is essential to conveying several of the previously discussed intangible communication elements when tele-visiting a remote location.

Again we considered the activities of people and identified the need for at least two levels of video resolution. The system should provide a wide angle view similar to the human eye for navigating and recognizing people (and objects) and also a smaller field of view for reading text on paper, white-boards, doors, and computer screens.

We also noticed that with only one-way video, PRoPs could be mistaken as tele-operated surveillance tools or autonomous reconnaissance drones. Both of these tasks are far from the intended application of PRoPs. We removed this video-asymmetry by adding a small (15 cm diameter) LCD screen with a video feed from the remote user. This two-way video is also an appropriate mechanism for transmitting a richer representation of the remote user through their facial gestures and expressions. When bandwidth is a problem and the screen is used only to display a still image of the remote user, we find it still succeeds in conveying the identity and existence of the remote user.

4.3 Mobility

Ambulation, even within a single building, is a significant portion of an individual's daily routine and thus we included mobility as a vital characteristic of PRoPs. But how sophisticated should the mobility be? We found that simple car-like navigation of a PRoP on the ground was fairly straightforward for a user to understand and control though a relatively simple interface. It also provided enough freedom for users to maneuver within (and outside of) buildings. This was the simple design of our first PRoP.

However, since human interactions occur where humans can travel, PRoPs must be able to reach much of the world accessible to humans. Again, we are not attempting to create an android or anthropomorphic robot so we will not attempt to handle what we call dextrous human motions. In particular we see little need for PRoPs to climb fences, swing from ropes, leap over ditches, repel down cliffs, slide down poles, *etc*.

Our basic philosophy is that PRoPs should be able to access the *majority* of locations most humans inhabit daily. Aiming for simplicity, we feel that PRoPs should be able to perform simple locomotion through fairly benign terrains such as mild inclines, curbs, stairs, and small variations in ground surface (*i.e.* sidewalks, grass, dirt, *etc*.). This includes traveling outdoors and also means that PRoPs must be be untethered (*i.e.* wireless). It is also important to impede the overall speed of the PRoP, typically through various gear reductions, to roughly mimic human walking pace.

4.4 Directed Gaze

We quickly learned that although remote users can see, hear, and move around, navigating still remained a tedious task and did not facilitate the ability to quickly glance around a room to get a sense of its size, occupants, *etc*. This problem was remedied by incorporating a small movable "head" (*i.e.* a camera on a controllable pan-tilt platform) onto the PRoP. Our device is similar to the GestureCam [13] which allows a remote participant in a conversation to have direct control of his or her visual field of view. This relatively simple PRoP "head" provides a vitally important element of human communication, direction of attention or gaze as discussed by several researchers [5; 11]. This allows PRoPs to perform human-like conversational gestures such as turning to face someone in order to see them, address them, or just give attention to them. These actions are also visible to people interacting locally with the PRoP and provide simple gestural cues to let individuals know when they are being addressed or looked at by the remote user.

4.5 Pointing and Simple Gesturing

By watching people interact we realized the importance gestures to of human communication. With our PRoPs remote users immediately found the need to point out a person, object, or direction to the individual in the remote space. Although the movable head could be used as a crude substitute, it lacked the correct visual gestural aesthetic of pointing and was often ambiguous to individuals watching the PRoP. We added a simple 2 DOF pointer so that remote users could point as well as make simple motion patterns. These motion patterns allowed the PRoP user to express additional non-verbal communications gestures such as interest in a conversation, agreement with a speaker, or to gain attention for asking a question in a crowded room.

Adequate pointing does not require a mechanism as complex as a human hand, since it is gross motion and not dexterity that is needed for the social function of gesturing. There has been a significant amount of research into gesture *recognition*. These systems typically aim to identify a human motion, typically made with a mouse, and interpret it as a known gesture. For example, a quick up-down motion of the mouse may be recognized as the "scroll page" gesture. However, we are making a conscious choice to use such symbolic descriptions of gestures only as a last resort. Instead we prefer to use continuous input devices like mice and joysticks to provide direct gestural input from the user to the PRoP. For example, compare typing text to a speech synthesizer, with spoken text transmitted through a speech compression algorithm. The synthesis approach may provide clean-sounding speech at low bandwidth, but all nuance and emotional content is lost. Similarly, music which is generated by computer from an annotated musical score is lifeless compared to music played by a human from that score, even if the recording mechanism is identical (*i.e.* MIDI).

In fact it is not really surprising that through these crude devices and narrow communication channels, that rich and complex communication is possible. Recall that actors transmit their gestures to audience members tens of meters away, dancers and mimes work without speech, and puppeteers work without a human body at all. All of us use the telephone without a visual image of our interlocutor. Our task in gesture transmission is to isolate the key aspects of gesture so as to preserve meaning as closely as possible. Some factors are clearly important, such as time-stamping to preserve synchronization and velocity. Others, such as mapping human degrees of freedom to robot "arm/hand" degrees of freedom are much less so.

4.6 Physical Appearance and Viewpoint

Although not anthropomorphic, we observed that PRoP design is loosely coupled to a few human-like traits which are important visual cues for successful communication and interaction. Clearly, a small ground-based robot conveys a rodent-like perspective of the world. However, a large robot is typically unable to navigate down narrow hallways, pass through doors, and impedes normal human traffic flow in a building. Furthermore, larger more industrial-type mobile robots are also more likely to frighten people, detracting from their use in human communication and interaction.

Since they stand in as a physical proxy for a remote user, it makes sense that PRoPs should be roughly the same size as a human. We attached a 1.5 meter vertical pole at the center of the PRoP to provide a realistic human vantage for the camera. In general we have found that the positioning of various attachments on the PRoP (*i.e* head, pointer, arm, *etc*.) should have some correspondence to the location of an actual human body part that provides the equivalent functionality. Also, all of the communication channels should be from the point of view of the PRoP (*i.e.* from on-board the tele-robot). It does not suffice to simply have a camera someplace in the room where the PRoP is currently located.

5 Discussion

We have circulated our ideas on this human centered approach to personal tele-embodiment and received several recurring comments and questions which we would like to address.

- **PRoP sounds like just another acronym, where are the new ideas?**

  Certainly, we hesitate to introduce yet another buzzword to the plethora of techno-jargon. However, it seems productive to use a common term to distinguish the growing research in this area. Obviously, methods of achieving telepresence are not new, nor are systems that allow tele-communication. Similarly, techniques and studies of human communication have been examined for centuries. What we feel is new is the merger of these methods and the primary focus on the individual person to guide the design choices of the entire system. We believe that even a small amount of attention to the human element in personal robotic design will reap countless benefits. This paper represents our best attempt to convey this direction of personal robotics research.

- **Isn't this just an extension of video teleconferencing?**

  While standard (and internet-based) video teleconferencing provides an arguably more realistic interface than many other forms of telecommunications, it is more of an enhancement to existing technology rather than a new form of communication. With video teleconferencing we find ourselves fixed, staring almost voyeuristically through the gaze of an immovable camera atop someone's computer monitor. As actions and people pass across the camera's field of view, we are helpless to pan and track them or follow them into another room. The result is a "one-sided" experience where the remote user feels immersed but there is no physical presence at the remote end with which people can interact. In essence we still lack mobility and autonomy. We cannot control what we see or hear. Even if we had cameras in every room and the ability to switch between them, the experience would still lack the spatial continuity of a walk around a building.

  We claim that users desire a more realistic perception of physical remote embodiment. We realized the importance of immersing the PRoP user in the remote space by providing continuity of motion and control of that motion. These elements provide the user the visual cues necessary to stitch together the entire visual experiences into a coherent picture of a building and its occupants and distinguish our work from that of standard video teleconferencing.

- **Isn't this just another form of telepresence?**

  Our approach differs fundamentally from more traditional versions of telepresence which involve an anthropomorphic proxy or android. Instead, PRoPs attempt to achieve certain fundamental human skills *without* a human-like form. More importantly, our research is driven by the study and understanding of the social and psychological aspects of extended human–human interactions rather than the need to create an *exact* re-creation of the remote experience. For example, we have already observed that even with poor video and crude motor controls, a PRoP provides adequate functionality to qualify as a useful tool for tele-visiting.

- **Why introduce the term tele-embodiment?**

  PRoPs allow human beings to project their presence into a real remote space rather than a virtual space, using a robot instead of an avatar. This approach is sometimes called "strong telepresence" or "tele-embodiment" since there is a mobile physical proxy for the human at the end of the connection. The physical tele-robot serves both as an extension of its operator and as a visible, mobile entity with which other people can interact. We coined the term *tele-embodiment* to emphasize the importance of the physical mobile manifestation.

- **I don't want a robot to stand in for me. Modern technology is already creeping into my life too much.**

  We do not believe that we can ever replace true human interactions, nor is it our goal to do so. Instead we are attempting to extend current human communication methods. That is, our intention is to provide the means for individuals to perform visits and interactions that would not otherwise be possible due to monetary, time, or distance constraints. Similarly, it is hoped that visits that now consume hours of traveling time can be tele-conducted in a fraction of the time with little loss of content. We expect this to result in additional free time for individuals to undertake more fulfilling endeavors rather than to be occupied solely with traveling.

- **Sure but robotics has always been concerned with people?**

  True, but much of that concern has been directed mainly towards safety issues when robots are operating near humans. Instead we claim that human centered robotics focuses directly on the tasks and issues that are part of daily human activity first, before the design of the robot. Furthermore, those observations of people directly influence the design decisions of the final system.

  Of course we should stress that it is vital that safety be a primary concern when designing PRoPs. We propose a teleoperational variation on Asimov's first law of robotics[3] which stipulates that at no time should a PRoP ever be capable of injuring a human being, regardless of the action or inaction of the remote teleoperator.

6 Conclusion

Our claim is that PRoPs provide an extremely useful, functional, powerful new tool for supporting human communication and interaction at a distance. They enable a variety of important work and social tele-activities far beyond what we perform currently with our computers and networks.

PRoPs are also an ideal platform for studying computer-mediated human interaction because they operate in existing social spaces and can interact with groups of humans. Despite our limited experience using PRoPs, we have been able to identify several factors that we consider vital to providing the most compelling overall experience for both the remote and local users. This is why our research draws as much on the sociology of group interactions as on sensing and actuation techniques. In fact we need the former to drive our choices for the latter.

Furthermore we believe that robotics as a research area is poised to begin significant contributions into the daily lives of people and society in ways that we are likely to not yet even imagine. We liken this to the movement of the elaborate institutional calculating engines of only a few decades ago into the casual daily interactions we observe between humans and computers today.

We claim that personal tele-embodiment is an example of human centered robotics. Most importantly, we emphasize and demonstrate the importance of conducting robotics research that focuses on the individual person rather than the intricate complexities of the machine and call for research following this methodology.

[3] "A robot may not injure a human being or, through inaction, allow a human being to come to harm." Handbook of Robotics, 56th Edition, 2058 A.D., as quoted in *I, Robot* by Asimov [2]

References

[1] Asada, H. H., Bekey, G., Tarn, T. J., Kazerooni, H., Leifer, L., Picard, R., and Dario, P. Workshop: Human centered robotics. In *IEEE International Conference on Robotics and Automation* (April 1997).

[2] Asimov, I. *I, Robot*. Garden City, N.Y. : Doubleday, 1950.

[3] Colgate, J. E., Wannasuphoprasit, W., and Peshkin, M. A. Cobots: Robots for collaboration with human operators. In *IMECE* (1996).

[4] Doherty, M., Greene, M., Keaton, D., Och, C., Seidl, M., Waite, W., and Zorn, B. Programmable ubiquitous telerobotic devices. In *Proceedings of SPIE Telemanipulator and Telepresence Technologies IV* (October 1997), vol. 3206, pp. 150–157.

[5] Donath, J. The illustrated conversation. In *Multimedia Tools and Applications* (March 1995), vol. 1.

[6] Finn, K., Sellen, A., and Wilbur, S., Eds. *Video-Mediated Communication*. L. Erlbaum Associates, 1997.

[7] Fish, R., Kraut, R., Root, R., and Rice, R. Evaluating video as a technology for informal communication. *Communications of the ACM 36*, 1 (1993), 48–61.

[8] Goertz, R., and Thompson, R. Electronically controlled manipulator. *Nucleonics* (1954).

[9] Goldberg, K., Mascha, M., Gentner, S., Rothenberg, N., Sutter, C., and Wiegley, J. Robot teleoperation via www. In *International Conference on Robotics and Automation* (May 1995), IEEE.

[10] Goldberg, K., Santarromana, J., Bekey, G., Gentner, S., Morris, R., Sutter, C., Wiegley, J., and Berger, E. The telegarden. In *SIGGRAPH* (1995).

[11] Ishii, H., and Kobayashi, M. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *ACM SIGCHI* (1992).

[12] Kraut, R., Fish, R., Root, R., and Chalfonte, B. Informal communication in organizations: Form, function, and technology. In *ACM Conference on Computer-Supported Cooperative Work* (1988).

[13] Kuzuoka, H., Kosuge, T., and Tanaka, M. Gesturecam: a video communication system for sympathetic remote collaboration. In *ACM Conference on Computer-Supported Cooperative Work* (October 1994).

[14] Mann, S. Smart clothing: The wearable computer and wearcam. *Personal Technologies* (March 1997). Volume 1, Issue 1.

[15] Minsky, M. Telepresence. *Omni 2*, 9 (June 1980).

[16] Paulos, E., and Canny, J. Delivering real reality to the world wide web via telerobotics. In *IEEE International Conference on Robotics and Automation* (1996).

[17] Paulos, E., and Canny, J. Ubiquitous tele-embodiment: Applications and implications. *International Journal of Human-Computer Studies/Knowledge Acquisition* (1997). Special Issue on Innovative Applications of the World Wide Web.

[18] Sheridan, T. B. *Telerobotics, Automation and Human Supervisory Control*. MIT Press, 1992.

[19] Steuer, J. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication 4* (Autumn 1992).

[20] Tang, J., and Mineman, S. Videowhiteboard: Video shadows to support remote colaboration. In *ACM SIGCHI* (1991).