

**Personal Tele-Embodiment**

by

Eric John Paulos

B.S. (University of California, Berkeley) 1991

M.S. (University of California, Berkeley) 1999

A dissertation submitted in partial satisfaction of the  
requirements for the degree of  
Doctor of Philosophy

in

Computer Science

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor John Francis Canny, Chair

Professor James Anthony Landay

Professor Gerald Mendelsohn

Fall 2001

The dissertation of Eric John Paulos is approved:

---

Chair

Date

---

Date

---

Date

University of California at Berkeley

Fall 2001

# **Personal Tele-Embodiment**

Copyright Fall 2001

by

Eric John Paulos

## Abstract

Personal Tele-Embodiment

by

Eric John Paulos

Doctor of Philosophy in Computer Science

University of California at Berkeley

Professor John Francis Canny, Chair

Humans live and interact within the real world but our current online world neglects this. This dissertation explores research into Personal Roving Presence (PRoP) systems that provide a physical, independently mobile proxy, controllable over the internet that enables a new form of telepresence: *tele-embodiment*. Leveraging off of its physical presence in the remote space, PRoPs provide important human verbal and non-verbal communication cues. The ultimate goal is a computer mediated communication (CMC) tool for rich, natural human interaction beyond currently available systems. This dissertation examines the history of several such devices and their development, design, interface, and system architecture. We also investigate relevant social issues, evaluations of several user studies, and the role PRoPs will play in our future.



---

Professor John Francis Canny  
Dissertation Committee Chair

For my loving parents,

Martha and Jack Paulos

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Thesis Statement . . . . .	3
1.2 A Solution . . . . .	3
1.3 Beyond Traditional Mediated Communication . . . . .	5
<b>2 Motivation</b>	<b>7</b>
<b>3 Tele-embodiment</b>	<b>11</b>
3.1 Telepresence . . . . .	12
3.1.1 Robotic Telepresence . . . . .	13
3.1.2 Virtual Reality Telepresence . . . . .	13
3.1.3 Collaborative Work Telepresence . . . . .	14
3.2 Telepresence and Immersion . . . . .	14
3.3 Human Centered Robotics . . . . .	16
3.4 Personal Telepresence . . . . .	16
3.5 Symmetry and Telepresence . . . . .	17
3.6 Importance of the Physical Body . . . . .	19
3.7 PRoPs and Tele-Embodiment . . . . .	19
<b>4 Foundational Work</b>	<b>23</b>
4.1 Mechanical Gaze . . . . .	24
4.1.1 Introduction to Mechanical Gaze . . . . .	24
4.1.2 Motivation . . . . .	26
4.1.3 Background . . . . .	26
4.1.4 Goals . . . . .	27
4.1.5 Why Use Live Images? . . . . .	28
4.1.6 Design Overview . . . . .	30
4.1.7 History . . . . .	31

4.1.8	Hardware . . . . .	32
4.1.9	Robot Interface and Control . . . . .	33
4.1.10	System Utilities . . . . .	37
4.1.11	Navigational Tools . . . . .	41
4.1.12	Usage Summary . . . . .	43
4.2	Legal Tender . . . . .	44
4.2.1	Motivation . . . . .	44
4.2.2	Telepistemology . . . . .	45
4.2.3	Introduction to Legal Tender . . . . .	46
4.2.4	Operation . . . . .	48
4.2.5	Results . . . . .	50
<b>5</b>	<b>Related Work</b>	<b>53</b>
5.1	Historical Telepresence Systems . . . . .	53
5.2	Telepresence and the World Wide Web . . . . .	54
5.2.1	Fixed Cameras . . . . .	54
5.2.2	Movable Cameras . . . . .	56
5.2.3	Robots on the Web . . . . .	56
5.2.4	Mobile Robots on the Web . . . . .	58
5.3	Blimp Related Work . . . . .	60
5.4	PRoP Related Work . . . . .	61
5.5	Computer Mediated Communication . . . . .	64
5.5.1	Tele-Tactile Experiences . . . . .	67
5.5.2	The Medium Is the Message . . . . .	67
5.5.3	Media Richness . . . . .	69
5.5.4	Social Presence . . . . .	70
5.6	Social Psychology . . . . .	71
5.6.1	Trust and Persuasion Online . . . . .	71
5.6.2	The Media Equation . . . . .	72
5.6.3	Non-Verbal Communication . . . . .	73
5.7	Human-Centered Computing . . . . .	73
<b>6</b>	<b>Space Browsers</b>	<b>75</b>
6.1	Introduction . . . . .	76
6.2	Motivation . . . . .	78
6.3	Goals . . . . .	81
6.3.1	Realism . . . . .	82
6.3.2	Globally Accessible . . . . .	82
6.3.3	Inexpensive . . . . .	83
6.4	Space Browsers . . . . .	83
6.4.1	Advantages . . . . .	86
6.4.2	Disadvantages . . . . .	89
6.5	Results . . . . .	92
6.6	The Future of Online Blimps . . . . .	94
6.7	Discussion . . . . .	98

<b>7</b>	<b>PRoPs</b>	<b>101</b>
7.1	Introduction . . . . .	102
	7.1.1 PRoP Design . . . . .	104
	7.1.2 Outline . . . . .	105
7.2	What's In a Name? . . . . .	105
7.3	History: Getting Here from There . . . . .	107
7.4	Early Prototypes . . . . .	108
7.5	PRoP 0 . . . . .	109
	7.5.1 Hardware . . . . .	109
	7.5.2 Software . . . . .	112
	7.5.3 Interface . . . . .	113
	7.5.4 Disadvantages . . . . .	113
7.6	PRoP 1 . . . . .	116
	7.6.1 Hardware . . . . .	116
	7.6.2 Software . . . . .	119
	7.6.3 Interface . . . . .	119
	7.6.4 Disadvantages . . . . .	120
7.7	PRoP 2 . . . . .	122
	7.7.1 Hardware . . . . .	123
	7.7.2 Software . . . . .	125
	7.7.3 Interface . . . . .	127
7.8	Compression: Making It Fit . . . . .	130
	7.8.1 H.323 . . . . .	131
	7.8.2 H.261 verses H.263 . . . . .	132
	7.8.3 Latency . . . . .	133
	7.8.4 PRoP Compression . . . . .	135
7.9	Designing Tele-embodiment: PRoP Architecture . . . . .	135
	7.9.1 Hardware . . . . .	136
	7.9.2 Physical Body . . . . .	138
	7.9.3 Software . . . . .	139
7.10	Elements of Tele-Embodiment . . . . .	140
	7.10.1 Two-way Audio . . . . .	140
	7.10.2 Two-way Video . . . . .	142
	7.10.3 Mobility . . . . .	145
	7.10.4 Directed Gaze . . . . .	147
	7.10.5 Deictic Gesturing . . . . .	147
	7.10.6 Reflexivity . . . . .	149
	7.10.7 Physical Appearance and Viewpoint . . . . .	151
	7.10.8 Browsing and Exploring . . . . .	153
	7.10.9 Hanging Out . . . . .	154
7.11	Discussion . . . . .	155

<b>8</b>	<b>Control, Navigation, Interface</b>	<b>159</b>
8.1	Let's Get Moving . . . . .	160
8.2	Joystick . . . . .	160
8.3	Keyboard Shortcuts . . . . .	161
8.4	Sonar Feedback . . . . .	161
<b>9</b>	<b>Experiments and Evaluations</b>	<b>163</b>
9.1	Experiment 1: Usability and Acceptance . . . . .	164
9.1.1	Users . . . . .	164
9.1.2	Setup . . . . .	165
9.1.3	Method . . . . .	166
9.1.4	Results and Discussion . . . . .	168
9.2	Experiment 2: Network Effects . . . . .	171
9.2.1	Users . . . . .	171
9.2.2	Setup . . . . .	171
9.2.3	Method . . . . .	172
9.2.4	Results and Discussion . . . . .	173
9.3	Experiment 3: Airport Traveler's Survey . . . . .	176
9.3.1	Users . . . . .	177
9.3.2	Setup . . . . .	177
9.3.3	Method . . . . .	177
9.3.4	Results and Discussion . . . . .	179
9.4	Experiment 4: Student Group Study . . . . .	182
9.4.1	Users . . . . .	183
9.4.2	Setup . . . . .	183
9.4.3	Method . . . . .	183
9.4.4	Results and Discussion . . . . .	184
9.5	Experiment 5: Tele-Lecture . . . . .	185
9.5.1	Users . . . . .	185
9.5.2	Setup . . . . .	186
9.5.3	Method . . . . .	186
9.5.4	Results and Discussion . . . . .	186
9.6	Summary . . . . .	187
<b>10</b>	<b>Social Implications</b>	<b>189</b>
10.1	Human Acceptance and Interaction . . . . .	189
10.2	Privacy and Security . . . . .	192
10.3	Authenticity . . . . .	192
10.4	Responsibility . . . . .	194
10.5	Intimacy, Trust, and Persuasion . . . . .	196
<b>11</b>	<b>Future Work</b>	<b>199</b>
11.1	Physical Design Improvements . . . . .	200
11.2	Force Feedback . . . . .	202
11.3	Haptic Integration . . . . .	202

11.4 Interface and Navigation Evolution . . . . .	203
11.5 Go There Now: Point and Click Interfaces . . . . .	204
11.5.1 Point-and-Click: Visual Servo . . . . .	204
11.5.2 Point-and-Click: Direct Calculation . . . . .	205
11.6 Navigation Automation . . . . .	207
11.7 Smile: Creating A Tele-visit Visual Scrapbook . . . . .	208
11.8 Panel PC Interface . . . . .	210
11.9 Getting to Know You Experiments . . . . .	211
11.10 Application Exploration . . . . .	212
<b>12 Conclusion</b>	<b>219</b>
<b>Bibliography</b>	<b>223</b>
<b>A PRoP Hardware Inventory</b>	<b>241</b>
<b>B PRoP Software Installation</b>	<b>243</b>
<b>C PRoP User Setup</b>	<b>247</b>
<b>D Create and Use Java Certificates</b>	<b>251</b>
<b>E PRoP Power Wiring Diagram</b>	<b>255</b>
<b>F PRoP Base Wiring Diagram</b>	<b>257</b>
<b>G PRoP Body Wiring Diagram</b>	<b>259</b>
<b>H PRoP Hand Wiring Diagram</b>	<b>261</b>
<b>I PRoP Hand Driver Circuit</b>	<b>263</b>

# List of Figures

1.1	A Personal Roving Presence (PRoP) and its various components as a one individual uses it to interact with another remote person. . . . .	4
4.1	Mechanical Gaze system architecture with Intelledex robot hardware. . . .	34
4.2	The web browser interface for Mechanical Gaze during an exhibit featuring various geologic rock formations. In this case Benitoite crystals. The pan, zoom, and location status controls are on the right while the comments and higher level navigation are below the image. . . . .	38
4.3	Two different closeup views of the roll and pitch tools. . . . .	41
4.4	Another view of the web browser interface for Mechanical Gaze during an exhibit featuring live gecko lizards. This page demonstrates the additional roll and pitch controls. Also, this image demonstrates an earlier unsuccessful interface tool – a small flag icon that raised up and down a mast used to control the height/zooming. This was later replaced by the more useful thermometer interface described in section 4.1.11 and shown in Figure 4.2. . . . .	42
4.5	Initial wide angle view of purported US 100 dollar bills in early phases of Legal Tender experiment. The black square represents an area exclusively reserved for the registered user currently operating Legal Tender. . . . .	47
4.6	A typical portion of a US 100 dollar bill presented to a remote registered user for exclusive access. . . . .	49
4.7	A list of available experiments that were available to registered user of Legal Tender. . . . .	49
4.8	A typical portion of a US 100 dollar bill presented after a puncture test. . . .	50
4.9	Wide angle view of purported US 100 dollar bills after a variety of experiments by numerous users over the course of several weeks. . . . .	51
4.10	Before (left) and after (right) images resulting from a user’s request to perform a thermal test on one of the purported US 100 dollar bills. . . . .	51
5.1	The Trojan Room Coffee Machine image as it appeared in 1994. This $64 \times 64$ pixel greyscale image was the first “webcam” in operation on the web. . . .	55



6.1	A blimp tele-robotic Space Browser traversing within a building. A small table and chair are shown for scale reference. . . . .	77
6.2	One of the first images of early space browsing blimp experimentation in the lab in 1994. . . . .	81
6.3	Schematic of basic space browser configuration . . . . .	84
6.4	An medium-sized blue polyurethane space browsing blimp interacting with a person. . . . .	85
6.5	The Java user interface to the space browser along with live video and audio feed from blimp as it appears on the remote pilot's computer. . . . .	87
6.6	Closeup view of the Java applet user interface to the space browser. . . . .	88
6.7	One of the smallest space browsing blimps designed at just under a meter in length. Using Mylar greatly decreased the blimp size but radically increased its fragility. Impressive blimps with short lives. . . . .	93
6.8	A pair of space browsing blimps in operation. The small silver mylar blimp and the much larger blue polyurethane version in the atrium of a building. . . . .	94
6.9	The <i>Eyeball</i> blimp flying within the Design Center at Ars Electronica in Linz, Austria in 1997. . . . .	95
6.10	Another view of the <i>Eyeball</i> blimp flying within the Design Center at Ars Electronica in Linz, Austria in 1997. . . . .	96
7.1	A graph (not to scale) illustrating the placement of PRoPs in Computer Mediated Communication (CMC) compared to various media in terms of their support for Social Presence and Media Richness. This is not an accurate graph and is only intended to provide a flavor of the role of PRoPs within the CMC landscape. . . . .	103
7.2	One of the earliest PRoP prototype systems. A modified remote control car base with attached hardware (circa 1996). . . . .	108
7.3	PRoP 0 with camera head, video, LCD screen, controllable "arm/hand" pointer, microphone, speaker, and drive-able base. . . . .	110
7.4	A view of the hand/arm pointer, eyeball camera head, and display screen on PRoP 0. . . . .	111
7.5	Closeup view of pan/tilt eyeball camera mounted atop PRoP 0. . . . .	112
7.6	PRoP 0 interface with Java applet (left) and NetMeeting application (right). Notice the highlighting of the head ( <i>i.e.</i> , eyeball) in the Java applet indicating head pan/tilt control mode selection. . . . .	114
7.7	PRoP 0 in action interacting with the author (left) and several others (right). . . . .	115
7.8	PRoP 1 . . . . .	118
7.9	An interaction with PRoP 1. A person and remote PRoP users are collaborating about some passages from a book sitting on the table. . . . .	121
7.10	PRoP 2 and its various components as it is used by a remotely embodied user to interact with a locally embodied participant. . . . .	124
7.11	An encounter with PRoP 2. Notice the locally embodied participant focusing his gaze cue onto the screen rather than the camera. This is a common issue with any form of video-conferencing. . . . .	128

7.12	From a remote location a primary user interacts and discusses the contents of a poster with a secondary user through P <sub>RoP</sub> 2 . . . . .	130
7.13	Block diagram of the H.323 System. . . . .	132
7.14	Using a P <sub>RoP</sub> to collaborate in resolving the details of an aircraft jet engine problem (envisionment). . . . .	137
7.15	System overview of the basic P <sub>RoP</sub> hardware configuration . . . . .	139
9.1	Initial setup of P <sub>RoP</sub> 1 for experiment 1 in a large room in the basement of a public library. This room is typically used for classes, public meetings, and workshops . . . . .	166
9.2	User number 2 during Experiment 1. This individuals is maneuvering the P <sub>RoP</sub> to complete the first task. . . . .	167
9.3	P <sub>RoP</sub> 1 during Experiment 2 as operated outside the computer science building at UC Berkeley. The remote user is controlling the P <sub>RoP</sub> from Denmark in this image. . . . .	174
9.4	Close-up view of P <sub>RoP</sub> 1 during Experiment 2 while under the control of a remote user in Denmark. You can see his video image on the screen while the inset image is the view he is receiving.” . . . . .	175
9.5	Images from the color leaflet used to explain the P <sub>RoP</sub> concept to people encountered at the San Francisco International Airport. . . . .	178
11.1	Two views of the prototype point and click interface. Before selection (left) and after completed motion (right). . . . .	206
11.2	The prototype tele-visit visual map construction tool. The path is shown with “interesting landmarks” as images. Higher quality images are displayed in the box on the left with an associated date/time stamp and comment. . . . .	209
11.3	Panel PC interface to a P <sub>RoP</sub> . . . . .	210
11.4	Remote inspection and troubleshooting over an aircraft repair task (envisionment). . . . .	213
11.5	Discussing a problem with the propeller operation on an aircraft (envisionment). . . . .	214
11.6	Troubleshooting another problem with the aircraft (envisionment). . . . .	215
11.7	An inspector discusses repair procedures (envisionment). . . . .	216
11.8	Training a repairman to perform a critical repair using a P <sub>RoP</sub> (envisionment). . . . .	217
11.9	A P <sub>RoP</sub> user has their Tarot cards read on Telegraph Avenue in Berkeley as third generation wireless becomes available greatly increasing the range of P <sub>RoPs</sub> (envisionment). Photo: Peter Menzel . . . . .	218
C.1	Java security warning dialog to grant permission to applet . . . . .	248

# List of Tables

3.1	Definition of Telepresence . . . . .	12
3.2	Definition of Remotely Embodied Participant . . . . .	18
3.3	Definition of Naturally (or Locally) Embodied Participant . . . . .	18
3.4	Definition of Tele-Embodiment . . . . .	21
3.5	Definition of Personal Telepresence . . . . .	22
4.1	Mutilation of national bank obligations from United States Federal Code, Title 18, Section 333. . . . .	49
7.1	PRoP 2 keyboard shortcut mappings for control signals. . . . .	129
9.1	Profile of participants in Experiment 1: Usability and Acceptance. Levels used for computer experience are Beginner, Intermediate, Advanced, and Expert . . . . .	165
10.1	First Law of Personal Tele-embodiment. . . . .	196

## Acknowledgements

Along the extensive, bewildering pilgrimage toward one's final dissertation signature, a vast collection of personal journeys are undertaken. Each important in gaining wisdom, forming character, and maintaining perspective. If you are fortunate, there will be colleagues, friends, and loved ones that will be there to encourage, critique, and advise you as you attempt to conquer these challenges. I would like to express my most sincere gratitude to a few of these extraordinary individuals who have made all the difference to me.

The tolerance of my officemates who listened patiently as my mind wandered across ideas and schemes throughout the years cannot be adequately repaid. It is their insightful feedback and thoughts that helped guide this project during critical moments. It is also with them that I believe I have consumed the most coffee and shared sleepless nights: Aaron Wallack, Ed Nicolson, Brian Mirtich, Ioannis Emiris, Dinesh Manocha, Ming Lin, Dan Reznik, Francesca Barrientos, Yan Zhuang, Jean-Paul Tennant, and Danyel Fisher.

With most of my time spent away from my office and within the Cory and Soda Hall Laboratories, I deep sense of gratitude is also due to the friendly faces I encountered in those habitats: Allison Woodruff, Paul Debevec, Chris Bregler, Eric Fraser, Albert Goto (especially for his late night visits with cold elixirs in hand), Philip Buonadonna, Scott Klemmer, Francisco Ruiz (the evening janitor with his countless words of sincere wisdom), Kalle Cook, and Mark Newman.

To those who aided in the transitions I frequently undertook from the world of graduate school life to the "real world" – thank you. You understood, you listened, you

made me laugh, you made me think, you made me understand (not necessarily in that order): David Pescovitz, Ed Barlow, Erik Hobijn, Zane Vella, RTMARK, Doug Rushkoff, Doug Sery, Andrew Gallagher, Steve Speer, Kelly Sparks, and Wanda Web.

And for the many explosive, excruciatingly loud times at “the other laboratory” where robotics, machines, electronics, energy, and power in general take on a whole new meaning. A necessary balance for a healthy graduate career: Mark Pauline and a truly insane and perfectly wonderful collection of misfits: Greg Leyh, Karen Marcello, Michael Fogarty, Kevin Binkert (for excessive use of his machine shop), Flynn Mauthe, Todd Blair, Violet Blue, Chris Bohren, Ralf Burgert, Frank Hausman, Michael Dingle, Jon Sarriugarte (for taking me to shoot guns when it was really necessary), Sara Filley, Warren Flynn, GX Jupiter-Larsen, John Law, Christian Ristow, Michael Shiloh, Brian and Amy, Kimric Smythe, Sergio Becerrill, Kal Spelletich, Scott Mitchell, Riko Knight, Liisa Pine, Diana DiFrancesco, and Tim Wadlington.

The constant encouraging words, feedback, and discussion from the many supporters of the Experimental Interaction Unit that invigorated my spirit during many moments of frustration.

To all those (employees and customers alike) that tolerated my ridiculously frequent visits to Farley’s, the Bay View Boat Club, and the Uptown throughout the years. Most importantly to all of the diligent, understanding employees at Al Lasher’s Electronics in Berkeley who provided me with numerous necessary electronic parts to keep my research moving forward and saved countless demos from almost certain failure. Also to Jim Slater and his team at the now defunct Nomadic Research for several eleventh hour PRoP resus-

citations.

I am also grateful to my colleague and friend, Ken Goldberg. His experience and wisdom have contributed in invaluable ways to this work. He continues to be an inspiration.

Bruce MacDonald, my father's dear best friend, for understanding me and helping my family to comprehend the realities and insanities of graduate life.

Finally, my closest friend and the one who has stood by my side during my most difficult moments: Jill Miller. I could not ask for a more understanding, patient, supportive, and loving woman by my side. I'm looking forward to expressing my gratitude to her for many years to come. Love you Jill.

Most importantly, it is to my beloved parents that I dedicate this work. They instilled in me from an early age the importance of believing in myself and the knowledge that anything is possible. These credos have been a constant source of motivation and inspiration without which none of this would have been possible. Their lifetime of unconditional love and support made this moment possible. My most heartfelt thanks to Martha and Jack Paulos.

# CHAPTER 1

## INTRODUCTION

*We intend to sing the love of danger, the habit of energy, and the strength of daring.*  
– *The Futurist Manifesto, 20 February 1909*

The current state of online culture presents us with a dilemma.<sup>1</sup> We are physical beings who experience the world through our bodies. The notion of a separation between abstract mind and physical body has been battered and eventually buried by Western philosophers since Kant [Kant, 1998, Carpenter, 1998]. In its place came new ideas, important among them phenomenology [Hegel, 1979, Merleau-Ponty, 1992], an articulating of perception and action as processes involving mind, body, and world.

In the East, Zen has acknowledged the importance of the body and action-in-the-world from the beginning. But when we access the on-line world, even a three-dimensional virtual world,<sup>2</sup> it is the “mind” that enters, not the body. Although this modeled virtual

---

<sup>1</sup>Portions of this chapter have appeared previously by the author in *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet* [Paulos and Canny, 2000].

<sup>2</sup>An online virtual community where individual users assume mobile, expressive identities within a three-dimensionally modeled world. These worlds are inhabited by modeled 3D objects along with other users represented by various 3D forms. The users of these worlds are networked together to simulate a unified

“body” may be augmented with an exotic 3D form, such a form is an avatar<sup>3</sup> in name only. The body stays outside. It is seen as a transducer, moving text or audio data in through keyboard, mouse, or microphone, and catching data from a monitor and speakers.

Realism is often described and measured in terms of digital fidelity, the number of pixels or the number of color bits. Motion may be described with the number of degrees of freedom, the virtual body becoming an abstract mobile entity. If we build virtual proxies that “look” realistic enough, shouldn’t the virtual experience be equivalent, or possibly better than the real? The biggest danger and most likely outcome is that we will succeed but the resulting experience will still be second-rate. From an epistemological point of view, we may be convinced by the sight and sound of the virtual world, but we will not be satisfied by our interactions with it. The experience of being in the world is much more than observing it.

The problem is that the view of “body-as-transducer” ignores the role of the body in motor-intentional acts. The dominant approach to Computer Mediated Communication (CMC) has broken the interaction into communication channels such as video, audio, haptics, *etc.* Notions of quality, reliability, latency are applied to these channels, mostly in a context-independent way. They are then adapted to the body’s (the body-as-transducer) perceptual performance.

But two human beings in the same room interact on a wholly different level. The

---

consistent view of the virtual 3D world and its inhabitants.

<sup>3</sup>Literally, an avatar is the incarnation of a Hindu deity (as Vishnu). However, in Computer Mediated Communication (CMC) jargon it often refers to a alternate representative of a person online typically in a chat room or virtual 3D world. We use the term reluctantly here only to draw a parallel between the more commonly understood virtual world avatar and a real physical version such as the PRoP. However, we will refrain for the duration of the dissertation from using this term to avoid confusion with its more relevant spiritual and religious connotation.



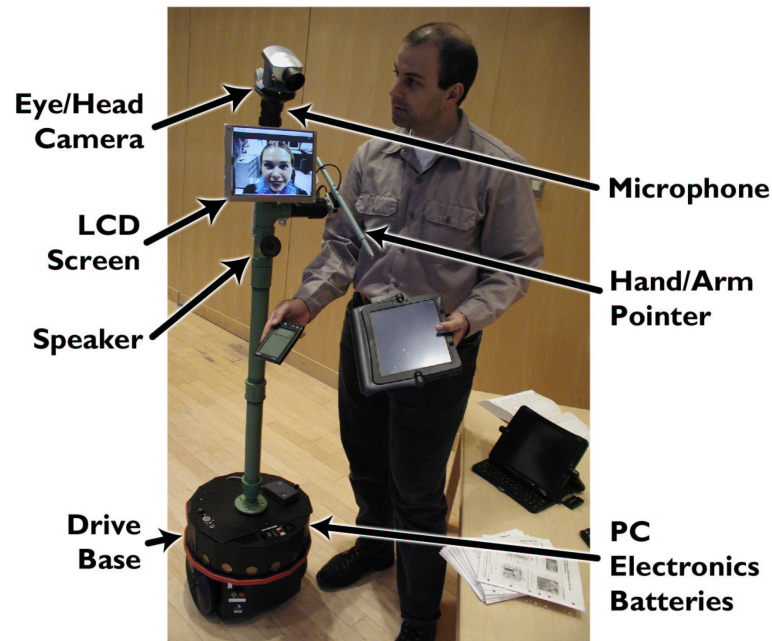
eyes are not just transducers but cues to attention, turn-taking, and sometimes deception. The hands complement speech with gesture in both conscious and unconscious ways. Dialogue is not a process of turn-taking speech, but a continuous and intimate coupling of speaker and listener. Much of dialogue is non-verbal and subconscious. We believe that CMC must be approached through an understanding of these behaviors, all of which involve mind and body together, and which use the body and its senses in many different ways. This theme is so important that we coined the term *tele-embodiment* to emphasize the tele-body. Tele-embodiment will be discussed in further detail in Chapter 3.

## 1.1 Thesis Statement

Overall, our claim consists of three components. First, the notion of *presence* can be efficiently constructed as a union of sufficiently realistic reproducible human cues and capabilities. Secondly, independently controlled spatial mobility is an essential element of this remote presence. Finally, a significantly wide range of common human communications, interactions, and activities are better captured, experienced, and expressed between distant locations using simple, novel, internet-controlled, untethered tele-robots that act as a physical proxy for people than by currently deficient computer mediated communication tools. It will be the goal of this dissertation to address these statements.

## 1.2 A Solution

The best solution to designing these rich tele-interaction devices remains an open problem. But we are initiating a solution path as the central theme of this dissertation. We



**Figure 1.1:** A Personal Roving Presence (PRoP) and its various components as a one individual uses it to interact with another remote person.

have already constructed numerous simple, inexpensive, internet-controlled, unteathered tele-robots to act as a physical proxy for people. These systems support what we refer to as new form of telepresence<sup>4</sup> called *personal tele-embodiment* (See pages 12, 21 and 22 for more formal definitions of these terms).

These Personal Roving Presence devices or PRoPs (see Figure 1.1) are not built to be anthropomorphic in form but to approach anthropomorphism of function. That is, they should support at least gaze, proxemics (body location), gesture, and dialogue. They are “body-like” because human-interaction is an intensely body-centered activity. They exist not in a virtual world but in the physical world. So they interact directly with people or

<sup>4</sup>“To convey the idea of these remote-control tools, scientists often use the words *teleoperators* or *telefactories*. I prefer to call them *telepresences*, a name suggested by my futurist friend Pat Gunkel.” as quoted by Marvin Minsky in 1980 when discussing the early usage of the term telepresence [Minsky, 1980].

groups of people rather than another virtual world proxy.

### 1.3 Beyond Traditional Mediated Communication

By operating in the real world, PRoPs expose the differences between natural human interaction and CMC. A PRoP is an individual presence and represents a unique remote participant. Unlike a videoconferencing system, it is not a “window” to somewhere else. The social capabilities of PRoPs contrast with those of live participants. We can explore what skills they have and which are lacking, depending on the context. And the contexts that we can study are broader than traditional teleconferencing, thanks to skills like mobility, proxemics, and deictic gesturing.

PRoPs need not be realistic portraits of humans because our motor-intentional behaviors are flexible. Our PRoPs are cubist statues, with rearrangements of face and arms, and separation of eyes from gaze. The arrangements are dictated by function and engineering constraints. The constraints on a personal social tele-robot are far from complete at this point, so we expect the design to be in a fluid state for some time.

Building PRoPs requires an understanding of the psychology of interaction, including the importance of of gaze, backchanneling, gesture, posture, and eventually subconscious cues. PRoPs provide a novel experimental platform for studying these phenomena. They provide a vehicle for the dissection of behaviors and the senses that support them. We can turn sensing and action channels on and off so that their effects can be studied. This is not to say that social behaviors decompose this way. In fact, our thesis is that they don't. However, we can discover the importance of various sensing and action channels on higher

behaviors by pulling switches and looking for change at the higher levels.

Ultimately, we hope to use the knowledge gained from P<sub>RO</sub>P<sub>S</sub> to design more satisfying online presences. Electronic interaction is strongly influenced by the medium [McLuhan, 1963, Turkle, 1997, Greenspan et al., 2000, Williams, 1977, Anderson et al., 1997, O'Malley et al., 1996, S. Whittaker and O'Connail, 1997, Reeves and Nass, 1996]. Personality is not a property of the abstract mind, but of the mind-body as experienced through all its motor-intentional modes [Kant, 1998, Hegel, 1979, Merleau-Ponty, 1992]. We find this theme as far back as the writings of Rumi in the thirteenth century in an excerpt from *Story Water* [Rumi, 1997]:

*The body itself is a screen  
to shield and partially reveal  
the light that's blazing  
inside your presence.*

*Water, stories, the body,  
all the things we do, are mediums  
that hide and show what's hidden.*

*Study them,  
and enjoy this being washed  
with a secret we sometimes know,  
and then not.*

If we can understand these modes, we have at least a glimmer of hope of building online tele-embodiment tools that are an acceptable alternative for the physical world. In the real world, we rely on others for most of our knowledge. If we can believe and trust the people we meet online, we can continue to learn and prosper as online beings. Without intimacy and trust, our online existence will remain an impoverished substitute.

## CHAPTER 2

# MOTIVATION

*There are times when arithmetic problems come our way and we might wish that we ourselves owned a computer to do the work for us. Such a situation would have its disadvantages, however. Electronic computers are bulky, expensive, complicated, and can be handled only by people with special training.*

*– An Easy Introduction to the Slide Rule by Isaac Asimov 1965*

Over the past three decades the computer has transformed from computer-as-calculator to computer-as-word-processor. More recently, standard application of Moore's Law<sup>1</sup> and the addition of networking to the computational fabric have been the major impetus for the rapid adoption of computer mediated communication (CMC) channels such as email, chat, and videoconferencing. The result is the next phase of the technological transformation as the evolution continues towards computer-as-medium-for-social-communication.<sup>2</sup>

---

<sup>1</sup>The observation made in 1965 by Gordon Moore, co-founder Intel, that the number of transistor per square inch on integrated circuits had doubled every year since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future. In subsequent years, the pace slowed down a bit, but data density has doubled approximately every 18 months, and this is the current definition of Moore's Law, which Moore himself has blessed.

<sup>2</sup>This movement is marked by the rapid creation and adoption of computational tools designed almost

Not unlike the introduction of the telephone or telegraph, the integration of the computer as a communication tool has already profoundly altered the means of human communication and interaction [Fischer, 1992, Marvin, 1988]. Although computers have empowered individuals with new methods of establishing and maintaining contact, it is clear that numerous essential components of human interaction have been lost compared to “face-to-face” (F2F) encounters. Of course many of these components were intentionally discarded, creating unique communication tools around which new social conventions concerning their acceptability and appropriateness developed. For example, unlike F2F, email is primarily text based and asynchronous.<sup>3</sup> As a result it is typically less interruptive than F2F communication. Socially, the acceptable behavior around the latency between messages, style of message, and duration, are quite different for email communication than F2F. The missing information allows for new models and types of acceptable interactions.

But what exactly have we relinquished during our rapid technological adoption process? And is it even possible to design systems that encapsulate the richness, warmth, and subtleties of face-to-face encounters? What would such a system look like? How will humans interface across this new medium? What will be the new paradigms? Most importantly, do we even need such a system and if so where will it fit into the existing social structure?

This dissertation undertakes a scientific exploration of these important questions. Drawing from literature and recent research in computer science, robotics, and social psychology, this dissertation describes the iteration through several designs and implementations solely to augment social human communication such as email, chat, virtual worlds, instant messaging, and videoconferencing.

---

<sup>3</sup>By asynchronous we mean that the recipient does not have to be there when the message arrives and they can read it and process it at their leisure.

tations of tools for personal telepresence. The conclusion demonstrates applications and usefulness of such devices through scientific evaluations and usability studies.

The work in this dissertation has evolved over several years and numerous projects. Mechanical Gaze (see Section 4.1) was our introduction to internet telepresence. It was designed to allow remote users to browse around a table-top filled with museum artifacts and objects. Quickly it became apparent that it would be more interesting to “get up from the table” and browse the physical space around the remote room. This led to the development of Space Browsers: helium filled, internet tele-operated, human sized blimps (see Chapter 6). Space Browsers met many of the goals of browsing remote physical spaces. However, users found interacting with inhabitants of the remote space far more useful and compelling. Unsited for this task, we abandoned Space Browsers and began to study human communication and interaction in more detail. We spoke with social psychologists and set out to design a system focused on facilitating human communication and interaction at a distance. Thus was born the Personal Roving Presence (PRoP) (see Chapter 7).





## CHAPTER 3

# TELE-EMBODIMENT

*I think everybody should be a machine. I think everybody should like everybody.*  
– *Andy Warhol*

Only a few decades ago computers were being praised solely for 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 fascinating contributions to science and society. Decades of robotics laboratory research into autonomy, computer vision, sensing, navigation, planning, mechanics, and design are finally propelling the first true emergence of personal and home robotics. With this movement, robotics is taking on a new social form and role. This dissertation addresses issues directly related to this augmentation of social functions to current robotics research. These *nuovo*-robotic or in fact anti-robotic extensions place

entirely new technological tools into the lives of ordinary people. They are anti-robotic because they are not designed to function as an android or anthropomorphic human. Nor are they designed to mimic robots as portrayed by Hollywood and perceived by a large portion of popular culture and society. Instead the focus is on social form and function over mechanized abilities. This epoch can easily be likened to the the transition of computers from laboratories to personal homes to environment [Weiser, 1991], to the human body [Mann, 1997].

### 3.1 Telepresence

Telepresence is a term used in several different communities [Steuer, 1992]. In each arena it manifests a different meaning. This section attempts to outline the major themes of telepresence in each milieu. Before we begin we will clarify this ambiguity in terminology and define telepresence in Table 3.1.

<p><b>Telepresence:</b> A user interface through which an operator receives sufficient information about a physical dynamically controlled mechanism called the teleoperator and the task environment, displayed in a sufficiently natural way, that the operator feels physically presence at the remote site. This can be a matter of degree. Naturally, an operator, upon reflection, knows where he or she really is. Nevertheless, the illusion of telepresence can be compelling if the proper technology is used for the task. The important distinction is that the remote location is real and the user has control of some physical system within that environment.</p>
---

**Table 3.1:** Definition of Telepresence

### 3.1.1 Robotic Telepresence

Methods for achieving telepresence (sometimes also called teleoperation) are not new to the field of robotics. One of the first electrically controlled mechanical teleoperational systems called the ECM-1 was developed by Goertz [Goertz and Thompson, 1954] for handling hazardous nuclear materials in the laboratory. Since then a variety of applications for tele-operated robotics have been explored [Sheridan, 1992] (also see Chapter 5 for further discussion of related internet tele-operated robotics projects).

However, most of these systems are designed for a single specific task and are quite complex. They also typically require expensive special purpose dedicated hardware and a highly trained operator to control and interact with the mechanism in the remote environment. By design, PRoPs strive to constrain their development so that they will be accessible to a wide audience without additional, expensive, or extraordinary hardware. In essence, telepresence for the masses. More importantly, unlike typical telepresence systems employed in remote inspection or hazardous exploration tasks, the primary application of personal tele-embodiment systems like the PRoP is to facilitate human communication and interaction.

### 3.1.2 Virtual Reality Telepresence

Telepresence is often used in conjunction with discussions surrounding the experience of a user interacting within a computer generated synthetic virtual world. This form of telepresence or virtual presence as it is also called occurs when a person's experience of sensory information is generated only by and within a computer. The resulting experience

compels the user with a feeling of being presence in an environment other than the one the person is actually in. This is not the definition or usage of telepresence involved in the discussion of PRoPs.

### **3.1.3 Collaborative Work Telepresence**

The computer supported collaborative community often uses the term telepresence in its description of shared interactive work environments, many of which incorporate some form of videoconferencing. The terminology usage is often meant to highlight the concept of shared workspace or environments. The idea is that if local and remote users are able to work on projects in a manner similar to all parties being in the same location, there is a notion of shared workspace or remote presence of others. However, the usage of telepresence here does not typically control any physical dynamic system on the other end. The sharing is usually confined to shared files, desktops, application workspaces, *etc.* A nationwide project, the National Tele-Immersion Initiative, has been ongoing for several years now into designing systems to support many of these tasks [Lanier, 2001]. While our work is closely aligned with the computer supported collaborative work community, this is not the definition or usage of telepresence involved in our discussion of PRoPs.

## **3.2 Telepresence and Immersion**

These definitions are intended to capture the essence of telepresence. However, it is clear that even these definitions are open to a wide range of interpretations. For example, certainly what one would call “traditional” forms of telepresence qualify. That is, those

with simple master slave setups with live or slightly delayed remote camera feeds delivered back to the operator. These forms of telepresence often require a “sense of presence” or immersion in order to complete a task such as remote inspection, handing of objects, or remote repair operations. But this definition does not always require a video image. How about a telephone? Certainly it is an interface between two people. Switch to a speaker-phone and the link between two spaces becomes apparent. A telephone user may feel a sense of presence in a remote location. But this is a function of the task and context. As the meeting breaks up and the room empties, the lone speaker-phone user will quickly lose her sense of immersion. Likewise, the information must flow back to the remote user.

An expert driving a remote control car around a room may be able to express herself and make her presence and individuality known to the room’s occupants. But the driver of the remote control car will fail to feel any sense of presence in the remote location. Once again, if we add a camera and wireless video link we begin to enter the grey area of what is telepresence. Grey in the sense that it is once again task specific. A user employing such a system to find an unexploded bomb may be satisfied with her sense of presence in that remote space. But the same system may lack the necessary tele-immersion when used to interact with others. Again we are attempting to capture the essence of this sense of remote presence with the term telepresence.<sup>1</sup>

---

<sup>1</sup>We have addressed the ambiguities in the term *telepresence* in the preceding sections. For the duration of this dissertation the term *telepresence* will have the implied definition from Table 3.1.

### 3.3 Human Centered Robotics

The research ideology of this dissertation is in the spirit of the recently identified area of “human-centered robotics” [Asada et al., 1997] 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 gain valuable insights into the social and psychological aspects of human 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 dissertation we concentrate on the design of one such human-centered system whose goal is to enable personal telepresence.

### 3.4 Personal Telepresence

Our intention is to provide a new form of telepresence<sup>2</sup> 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 functions 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 communication cues that are inherent to human communication, understanding, and inter-

---

<sup>2</sup>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. A full discussion of this can be found in section 3.1

action. 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.

We believe 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 telepresence.

### **3.5 Symmetry and Telepresence**

Before proceeding it is worth taking some time to clarify the two different roles or sides of telepresence that cause confusion. Often videoconferencing and some telepresence system provide a reasonably symmetric experience. That is, each side experiences a video and audio feed from the other with both sides viewing and hearing similar views from similar perspectives. However, often in the forms of robotic telepresence we are addressing it is one side “reaching out” to experience a distant side while that distant location may have no view or experience back to the other space. In some other cases such as with P<sub>RO</sub>P<sub>S</sub> there is an experience in both directions (*i.e.* there is two-way audio and video) but the experience is not the same. The two sides have asymmetric interfaces, abilities, expressions, *etc.* In these asymmetric cases of telepresence it is important to make a distinction between the individuals on the two distinct sides of the telepresence. Our solution is to propose two new definitions to correspond to the two sides of the tele-experience (see Table 3.2 and

Table 3.3).

**Remotely Embodied Participant:** In telepresence the individual controlling the actions of a remote physical system or device. This is the user primarily involved in initiating the actions of a system in a remote location with the goal of achieving some degree of embodiment or telepresence within that remote location. This can also be thought of as the controller of the remote system. It is also sometimes called the primary user of the tele-system.

**Table 3.2:** Definition of Remotely Embodied Participant

**Naturally (or Locally) Embodied Participant:** In telepresence the individual or individuals experiencing and/or interacting with a physical system or device controlled by a *remotely embodied participant*. The remotely embodied participant is in a distant location while the naturally embodied participant is embodied in the local space in the normal manner that they are embodied within the real physical world. The naturally embodied participant is also sometimes called the secondary user.

**Table 3.3:** Definition of Naturally (or Locally) Embodied Participant

In terms of the PRoPs we will see that the remotely embodied participant is the driver of the PRoP sitting in front of their desktop or laptop computer interface while the naturally embodied participants are the individual or individuals cohabitating with the PRoP and interacting with the remotely embodied participant through the PRoP.



### **3.6 Importance of the Physical Body**

One early lesson learned in the pursuit of personal telepresence was the importance of the remote body or form. We observed this from watching others interact in real life. This compelled us to emphasize the importance of a remote physical body in our Space Browsers and later PRoPs (see Chapters 6 and 7).

The remotely embodied participant is empowered with a one-to-one mapping to a body-like form in a remote location. Since this remote form stood in as a remote personal proxy for that user, it embodied a form of personal telepresence.

### **3.7 PRoPs and Tele-Embodiment**

Internet video teleconferencing provides an arguably more realistic interface into a remote space than many other CMC connections such as email, telephone, instant messaging. However, it is more of an enhancement to existing telephone communication technology rather than a new form of communication. With video conferencing 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. 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 realized that it was necessary to deliver a more realistic perception of physical embodiment of the user within the remote space being explored. Such as system must

immerse the user in the remote world by providing continuity of motion and user control of that motion. These elements would provide the user the visual cues necessary to stitch together the entire visual experiences into a coherent picture of a building and its occupants. We also wanted to provide the user with the means to communicate and interact with the remote world and its real inhabitants using this new system. Furthermore, we wanted such a system to be accessible to any user on the internet with standard software running on currently existing computer architectures.

PRoPs allow humans to project their presence into a real remote space rather than a virtual space, using a robotic mobile entity rather than a virtual proxy or *avatar*<sup>3</sup> as they are often referred to in 3D worlds. This approach is sometimes referred to as “strong telepresence” since there is a mobile physical proxy for the human at the end of the connection. As a result we coined the term *tele-embodiment* to emphasize the importance of the physical mobile manifestation [Paulos and Canny, 1998]. Tele-embodiment is defined in Table 3.4.

This approach differs fundamentally from more traditional versions of strong telepresence that involve an anthropomorphic proxy or android. Instead, PRoPs attempt to capture, distill, and reproduce certain fundamental human skills without a human-like form. More importantly, the research described in this dissertation is driven by the study and understanding of the social and psychological aspects of extended human-human interactions rather than the rush to implement current technological advances and attempt to re-create *exact* face-to-face remote human experiences. In fact, many believe strongly that

---

<sup>3</sup>Reluctant use of the term avatar here (see related footnote on page 1)

<p><b>Tele-Embodiment:</b> A form of telepresence where a remotely embodied participant is able to have a device or articulated form act personally as their remote body proxy. Typically this device will be a personal representation of themselves, their body, and their actions. Furthermore, such systems are easily identified by naturally embodied participants as representing a single remote human. Tele-embodiment systems must also be untethered, provide independently controllable mobility, and manifest a “reasonably sufficient” number of physically controllable human communication cues. Tele-embodiment is telepresence with a personified, perceptible body that is anthropomorphic in function only, not in form.</p>
--

**Table 3.4:** Definition of Tele-Embodiment

the re-creation of exact face-to-face human encounters cannot be achieved through any mediated communication [Hollan and Stornetta, 1992]. There will also be something lost or cues misinterpreted compared to face-to-face.

To a large degree this definition depends on context. However, in general a remotely controlled robotic arm would not be a reasonable representation of a person and their overall actions. That is, a naturally embodied participant observing such an arm may view it as a reasonable representation of a remotely embodied participant’s arm, at best. More likely, it would be viewed as a pre-programmed mechanical arm. It is unlikely that the representation of only a remotely embodied participant’s arm would suffice as a personal representation of that user’s bodily presence. This raises the importance of the experience being a personal one with a one-to-one mapping of person to PRoP. To clarify we state the following definition.

<b>Personal Telepresence</b> $\equiv$ Tele-embodiment.
--

**Table 3.5:** Definition of Personal Telepresence

## CHAPTER 4

# FOUNDATIONAL WORK

*Simulation threatens the difference between true and false; between real and imaginary.*  
– Baudrillard

Personal tele-embodiment evolved over a number of years.<sup>1</sup> Much of its inspiration derived from several of our earlier internet based telepresence projects and systems. This chapter details several of the most relevant complete systems. These are descriptions of systems towards which the author contributed substantial design and implementation effort. To clarify, this chapter contains work and projects previously designed and implemented by the author. Discussion of other related work is deferred until a subsequent chapter (see Chapter 5).

Three main projects define the early landscape of personal tele-embodiment, Mechanical Gaze (1995), Legal Tender (1996), and Space Browsers (1996). The first two are discussed in this chapter. The scale and relevance of Space Browsers to current personal

---

<sup>1</sup>Portions of this chapter have appeared previously by the author in *Delivering Real Reality to the World Wide Web via Telerobotics* [Paulos and Canny, 1996a] and *A World Wide Web Telerobotic Remote Environment Browser* [Paulos and Canny, 1995].

telepresence system merits its own chapter (see Chapter 6).

## 4.1 Mechanical Gaze

Robots provide us with a means to move around in, visualize, and interact with a remote physical world. We exploited these physical properties coupled with the growing diversity of users on the World Wide Web (web) [Berners-Lee et al., 1992] to create a web based telerobotic remote environment browser in early 1995. This browser, called Mechanical Gaze, allows multiple remote web users to control a robot arm with an attached camera to explore a real remote environment. The environment varies but is typically composed of collections of physical museum exhibits that web users can view at various positions, orientations, and levels of resolution.

Mechanical Gaze came online in 1995 and became the second tele-operated internet based robotic system and the first with a color camera and images. In late 1997 it was taken down. Just prior to its termination it stood as the longest continuously operational online robot on the web.

### 4.1.1 Introduction to Mechanical Gaze

We designed this teleoperated web server in order to allow users throughout the world to visit actual remote spaces and exhibits. It also served as a useful scientific tool by promoting discussion about the physical specimens in the browser such as insects, live reptiles, rare museum collections, and recently discovered artifacts.

The use of an on-line controlled camera eliminated some of the resolution and

image angle selection problems encountered in digitized image libraries. The user had complete control over the viewpoint, and could experience the exhibit in its state at a particular moment in time, under the same conditions and lighting as a viewer who was in the actual space.

In addition, each exhibit had a hypertext page with links to texts describing the object, other web pages relevant to it, and to comments left by other users. These pages were accessed by navigating the camera in physical space and centering on a particular object. The fixed, pre-recorded robot joint positions for the “center location” of each object were recalled as a user selected it for viewing from the exhibition web page. The robot moved directly over the object, captured an image, and delivered it back to the user. The user could further refine their navigation around the object using various web based control mechanisms. The pages can be thought of as mark-ups of 3D objects in the spirit of VRML<sup>2</sup> [Pesce et al., 1994], but where the objects are actual physical entities in a remote space rather than simply models.

Exhibits could be added or removed in a matter of a few minutes, allowing for an extremely dynamic array of objects to be viewed over the course of only a few months. The only limit on the number of exhibits available was the physical dimensions of the robot’s workspace, which was approximately 4000 cm<sup>2</sup>. Users were encouraged not only to check back for upcoming exhibits, but to participate themselves. Users could leave commentary about an item on exhibit, creating dialogue about the piece, as well as give feedback to the owner, artist, or curator of the object. Institutions, museums, curators, scientists, artists,

---

<sup>2</sup>VRML is the Virtual Reality Markup/Modeling Language used to describe 3D worlds online much as HTML is to text. However, as of the writing of this dissertation VRML has been greatly diminished as a key 3D online modeling or development tool.

and individual users were all invited to exhibit objects.

### **4.1.2 Motivation**

Initially, we were driven to develop a useful application for interactive telerobotics. We were inspired by the diversity and growth of the web as the medium for such an inexpensive, publicly accessible tool for remote environment browsing. The restrictions imposed by the Hyper Text Markup Language<sup>3</sup> (HTML) made it difficult to design an intuitive user interface to a complex robotic system. Certainly, we could have chosen to construct custom navigation software for users to download. While this would have allowed us more freedom in the design of the overall system, it would have severely restricted the accessibility of the system. Since we considered the quantity and diversity of users on the web as one of its most powerful aspects, we chose to constrain the development of our system to make it accessible to the entire web community.

### **4.1.3 Background**

One of the early goals of the project was to incorporate methods in which users could remotely examine and comment on actual museum exhibits. At first we were interested in how well such a tool would operate on insect exhibits. We developed a prototype telerobotic browser and presented it at the Biological Collections Information Providers Workshop in January of 1995. At this workshop we received feedback about the uses and implications of such an application to natural science research. Later, in April of 1995 we presented the browser at *Wavelength*, an art installation in San Francisco exploring the

---

<sup>3</sup>HTML is the underlying language used to describe web page layout and content.



science and nature of movement. At these two arenas we were able to learn what elements of the browser were important, not only to scientists performing research, but also to novice users attempting to explore various remote spaces.

#### 4.1.4 Goals

Before designing the system we set forth our goals for the project. Our primary goal was to provide a universal remote environment browsing tool that is useful for the arts, sciences, and in the development of education and distance learning. To meet this goal we agreed upon several elements that we felt were essential to any web-based telerobotic system.

First, we wanted to insure universal unrestricted access to the system. This would allow access to artifacts and objects by a wider audience than previously available. Current access restrictions are usually the result of geographic, political, or monetary constraints preventing the individual from traveling to the object. Likewise, owners and curators of exhibits do not always have the resources or the desire to tour the objects throughout the world. We wanted to develop a tool that would attempt to solve many of these problems by bringing the people together with the objects at a minimum cost.

Rather than a fixed, static display, the browser must allow these users true three-dimensional navigation around objects at varying positions, orientations, and levels of resolution. As David Gelernter suggests in his book *Mirror Worlds* [Gelernter, 1992], such systems that gaze into remote spaces should show each visitor exactly what they want to see. This requires the system to provide millions of different views from millions of different focuses on the same object. Certainly visitors will desire to zoom in, pan around, and roam through the world as they choose. More importantly, they should be permitted to explore

this space at whatever pace and level of detail they desire. Users should also be free to swivel and rotate the image, to get a better look at regions that might be obscured in the initial perspective.

The telerobotics browser should also provide to the exhibit owners, curators, and caretakers a forum to receive feedback and commentary about their exhibit. This same forum should also allow scientists to discuss details concerning classification of specimens such as insects or the origins of a recently discovered artifact. Essentially, some method for leaving comments and creating dialogue should be provided.

Finally, the system should allow exhibits to be added and removed with a minimum of effort, thus providing the possibility of exhibiting a wide variety of objects over the course of a few months. In addition, recently discovered/developed scientific objects should be able to be added for universal browsing within the order of a few minutes.

#### **4.1.5 Why Use Live Images?**

A common objection to our approach is why we did not simply use pre-stored digitized images for browsing objects and spaces. While we agree there are valid uses for pre-stored images, the remote environment browser offered several distinct advantages over conventional image database solutions.

The standard approach to providing remote access to a museum's collection of visual data is to digitize and pre-store images of all artifacts or specimens. This solution requires considerable expense and time commitment to complete the capture, storage and serving of digitized images. We also learned from feedback during our participation in the Biological Collections Information Providers Workshop in January of 1995 that each

researcher had a preferred viewing angle and resolution. Essentially, everyone had different viewing needs. Our telerobotic approach attempted to solve this dilemma by allowing remote scholars to interactively view museum artifacts and specimens on demand at a variety of viewing angles and resolutions. Our interactive viewing solution also relieved museums of the need to store digital images of entire collections over a variety of resolutions.

Our approach allowed immediate visual access to a much larger portion of a museum's collection than currently employed archival techniques. Traditional image capture can take several years for large research collections, with millions of specimens that require special handling. The remote environment browser solution eliminated the waiting period that usually occurs during serial indexing and image capture. The hope was that museums that utilized a remote browsing model would be able to provide remote access to larger portions of their collection materials at a moment's notice. However, we did not carry the Mechanical Gaze project far enough to perform enough research into the usefulness of this technique. Typically, the number of accessible objects increases over time as they are painstakingly archived. Scientists, historians, and researchers agree that the ability to view specimens is more valuable if all specimens are available at the same time. The fewer specimens in a collection that are digitized, the less research value accrues to the resource as a whole.

By allowing researchers to choose their own view and magnification of the specimen or artifact, arguments over which specific view or number of views a museum should provide to remote users should be eliminated or at least minimized. With a three dimensional object there will always be arguments surrounding what view to capture. Unless users can choose

their own view of museum collections' materials, they will not be satisfied with using digital images for research. Even more importantly, some visually oriented research uses, such as taxonomy and morphology can not be supported in the digital environment without the provision of multiple views and magnifications. Useful statistics can easily be gathered by the browser as to which views are more popular among scientists and hence draw conclusions as to the relative importance of particular views and resolutions.

Certainly, dynamic exhibits such as live creatures [All and Nourbakhsh, 2000], moving liquids, and mechanical systems must be viewed using live images. These live views are necessary to study the behavior of such systems. Further discussions about the use of digital images in art and science, as well the implications of their use can be found in several sources [Durrett, 1987, Lynch, 1991, Ester, 1990, Kirsch and Kirsch, 1990].

The sensation of embodiment of an individual in a real life distant location has provided more than enough impetus for people to develop remote telepresence systems. We defer full discussion of this related work until Chapter 5.

#### **4.1.6 Design Overview**

Our design choice for the user interface to the remote environment browser was to mimic much of the look and feel of a museum. We chose this approach, hoping that users would find it familiar to navigate, and thus more intuitive and inviting to use.

As a user entered Mechanical Gaze, they were presented with a chance to view some general information about the project, receive a brief introduction, obtain help in using the system, or enter the exhibition gallery.

Users who entered the exhibition gallery were presented with an up to date listing

of the exhibits currently available for browsing. These were the exhibits that were physically within the workspace of the robot and could be explored. The idea behind the exhibition gallery was to give only a brief introduction to each of the available exhibits. This typically consisted of providing the name of each exhibit, the dates it would be available, the presenter(s), and perhaps a very brief description.

Users who desired to examine an exhibit in greater detail could simply select it from the listing. The user would then be presented with a more detailed description of the exhibit as well as a chance to either browse the exhibit using the robot or request to view the comments corresponding to that exhibit.

#### 4.1.7 History

An interesting coupling of robots and level of detail is found in examining the literary work of two brothers at the early part of the last century. In 1923 Karel Čapek wrote the play *R.U.R. (Rossum's Universal Robots)* [Čapek, 1923]. This exceptional fantasy melodrama is one of the first to address the conflicts involved in replacing human labor with machines and is historically significant for coining the word *robot*. Later, Čapek, along with his brother Josef, wrote *The Insect Play (And so ad infinitum)* [Čapek and Čapek, 1922]. The play begins with a quote.

*So, Naturalists observe, a flea  
Has smaller fleas that on him prey;  
And these have smaller still to bite 'em,  
And so proceed ad infinitum*

While the reference is non-technical, its bio-life theme tied in with our early motivation from work with entomologists. Our original design for the museum browser focused on the incorporation of a robot to view ever increasing levels of detail in insects. In hindsight,

it's clear that the browser has a multitude of other applications. However, we acknowledge Karel and Josef Čapek for so eloquently surmising both of these themes and their interesting robot connection.

#### 4.1.8 Hardware

The Mechanical Gaze system (see Figure 4.1) has operated from both an Intelledex 605T robot<sup>4</sup> with 6 degrees of freedom (DOF) and a 4DOF RobotWorld robot [Scheinman, 1987]. The only noticeable external difference was that the roll and pitch operations were only available when the system was operating from the Intelledex robot. Otherwise, the back-end robotic hardware change was transparent to the user, giving hope to the prospect of such systems running on a variety of different robot hardware in various environments.

Image capture was performed using a camera and frame grabber hardware. Images were received from an RCA Pro843 8mm video camera mounted onto the last link of the robot. The auto-focus feature of the video camera allowed users to view a variety of objects clearly, regardless of the object's own height or the distance from which it is viewed. Typical exhibition spaces allowed users to capture clear images anywhere from 3–30 cm from the surface of the object.

Images were digitized on either a VideoPix frame grabber card attached to a Sun IPC workstation or standard image capture hardware available on an SGI Indy. Eight bit  $320 \times 240$  color images were captured in less than 50 ms. Further computation to convert

---

<sup>4</sup>The Intelledex 605T was a robot manufactured in the 1980's by Intelledex, Inc. of Corvallis, Oregon. Intelledex has since gone out of business.

the image into a compressed JPEG<sup>5</sup> format for incorporation into HTML documents and save it to disk consumed an additional 2–3 seconds. Overall, the time required to capture, convert, and save an image was on the order of 2–3 seconds. Recall that this project was developed in 1994 and 1995 when image capture hardware was still considered quite exotic and expensive.

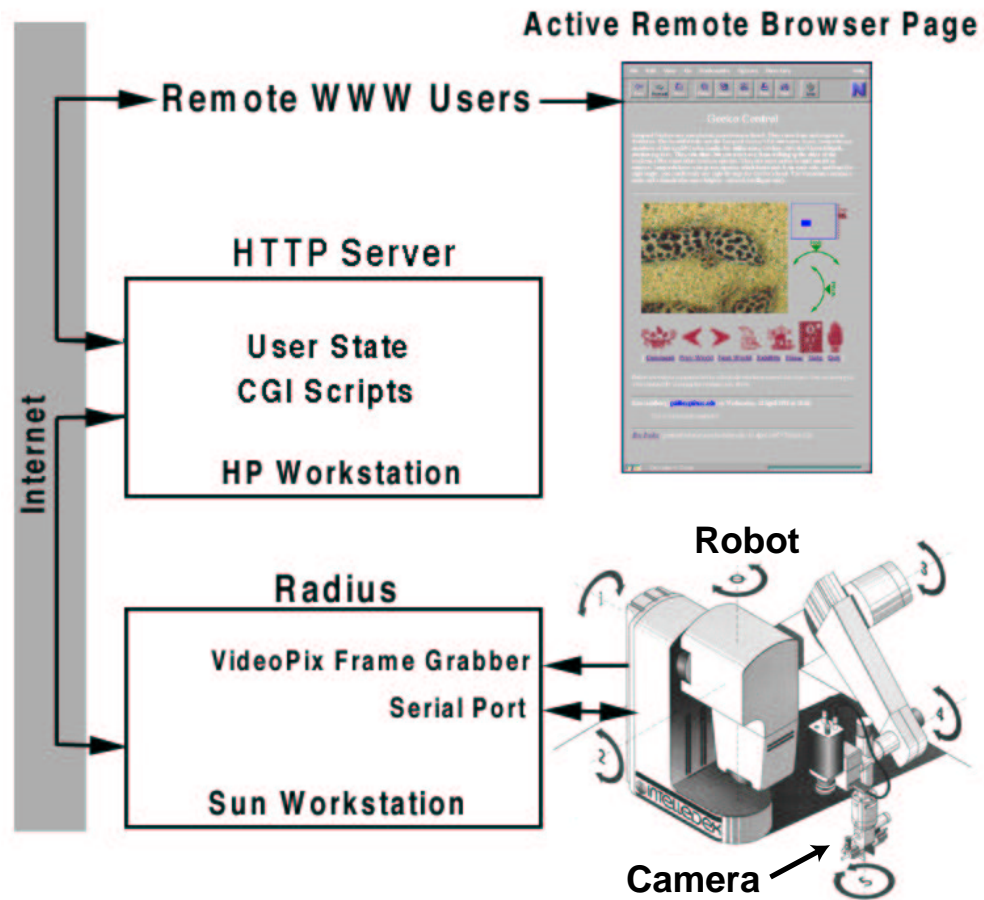
The actual web server containing the custom Common Gateway Interface (CGI) scripts and state information for individual users operated from an HP 715/60 workstation. This machine provided the front end interface to the system by receiving requests from web users. It also employed the services of the other hardware in the system, *i.e.*, the robot and camera, and delivered the results back to the user as a web page.

#### **4.1.9 Robot Interface and Control**

To interface the robot to the web, two separate pieces of code were written. The actual robot motion was performed by a daemon that accepts standardized requests via a socket connection and converts them into the native robot dependent commands. The other code interacted directly with the remote web user by handling administrative issues, resource contention, HTML page layout, and requests to the robot daemon when robot motion was required.

---

<sup>5</sup>JPEG is a standardized image compression mechanism that stands for Joint Photographic Experts Group.



**Figure 4.1:** Mechanical Gaze system architecture with Intellex robot hardware.

### Radius: The Robot Control Daemon

Radius<sup>6</sup> was the name of the robot control daemon<sup>7</sup> that provided a standardized interface to the various robots involved. By standardizing this interface, the rest of the system could be written ignoring any special kinematics or control systems for the particular end robot. Requests that involve control of the robot or camera hardware were handled by

<sup>6</sup>Radius is named after the main robot character in *R.U.R.* by Karel Čapek [Čapek, 1923].

<sup>7</sup>Literally, a daemon (also demon) is an attendant power, spirit, or genius. In a computing context it is often a process that waits in the background, ready to attend to and handle a user's requests. For example, a print server may have a daemon running awaiting print request commands to spool data to a printer.



Radius. Radius listened for these requests on an established socket port.

When a socket connection was made, Radius first checked for authentication using a known encoding. This prevented unauthorized control of the robot hardware. This was particularly important, especially as we move towards devices with the capacity of physical manifestations of energy in a remote environment [Pauline and Paulos, 1997b, Pauline and Paulos, 1997a]. The damage resulting from unauthorized access into such a system could easily cause not only irreparable damage to the robotic equipment and exhibits, but human injury as well. Therefore, measures to prevent at least the most naïve attacks should be included in such systems.

Authorized connections to Radius included a 4 byte message. The message encoded the type of request and a mask. The request type was either a motion command or image capture command. This was followed by several bytes of data depending upon the request type and mask. Radius could also query the robot to determine when all motions had stopped, hence allowing an image to be captured.

When an image grab request was received, Radius employed the use of the available image capture hardware to grab an image, convert it to a  $320 \times 240$  8 bit color JPEG, assign it a unique identification number that is embedded in the image filename, and output it to a temporary space. The unique image number was passed back to the requesting process so that the corresponding image could be displayed in the resulting web page.

Since our interface design was web based, requests were event driven. After a user had loaded an image, the robot was left idle until the user performed another request. Instead of allowing this exclusive access to the robot, leaving the robot idle while the user

contemplated their next action, we used the time to service additional requests from other users. By multitasking, we provided increased access to the robot as well as a more efficient use of system resources. However, we needed a method to guarantee that certain atomic operations were exclusive. For example, a request to move and grab an image, must be exclusive. This insures that no other motion occurs between the time we move the robot and capture the image. If we had failed to implement this, we would have no guarantee that the image delivered back to the user was actually taken from the location that they requested. Using internet socket connections allowed us to enforce the mutual exclusion necessary to insure the correct functionality of Mechanical Gaze even when handling multiple requests. When a request was received by Radius, subsequent requests were queued until the first request had been handled. This insured that requests occur in order and that each user/request has exclusive access to the robot and camera hardware during that time.

### **Navigation Page Construction**

Requests to browse an exhibit were handled by a custom CGI script. Initially, the script was passed a unique identifying internal number corresponding to the exhibit to be browsed. The script would read in the current list of exhibits and extract the relevant information for the exhibit of interest. One of these items was the physical location of the exhibit in the remote environment. Using this information, a socket connection was opened to Radius, the robot control daemon, and a request made to move the robot to the desired location and capture an image.

When the result of that request was received back, the CGI script dynamically laid out the web page. First, it extracted information from the internal list of exhibits. This

provided the name of the HTML file to place at the head of the browser page. The system inserted a line to indicate the amount of time the user has been using the system. Next, it inlined the captured and converted JPEG image, placing it within an imagemap with a unique randomly assigned number. To the right, various robot navigational tools were laid out. Additional web navigation icons were attached below this. These icons allowed users to leave comments about the exhibit, move to the next or previous exhibit, return to the list of exhibits, obtain help, or move back to the Mechanical Gaze homepage. To convey a sense of presence of other users, the system then displayed the last three visitors to the system. The various comments left concerning the exhibit were attached to end of the page. Finally, the CGI script wrote out an internal user file using the same randomly generated unique number from above. This file contained the state information, such as the user, position, time, and other information concerning the page and image just delivered. This number was embedded within the page so that requests originating from this page would reference into this corresponding unique status file. This allowed for subsequent requests to make their reference relative to the correct position that the user last viewed. The final result of a remote environment navigation request was a web page similar to the one depicted in Figure 4.2.

#### **4.1.10 System Utilities**


Mechanical Gaze was a distributed system, employing several different pieces of hardware. To manage these systems as well as maintain the entire system in a functional state, several utilities were developed.


**Benitoite**

Benitoite (blue) crystals with neptunite (black) on natrolite (white). Collected by G.D. Louderback


---

Hello EricPaulos. You've been on for 8 minutes






**Location Status Tool**



**Zoom Tools**



[ [Comment](#) | [Previous](#) | [Next](#) | [Exhibits](#) | [Reload](#) | [Help](#) | [Quit](#) ]

---

**Last Three Visitors:**


Tue 30 Jan at 13:38 PST [EricPaulos \(paulos@cs.berkeley.edu\)](#)  
 Tue 30 Jan at 12:28 PST Guest from proxy.mclink.it  
 Tue 30 Jan at 12:22 PST Guest from pcl1.iq.ufjf.br

---

**Neal ([nfurgurson@gems.vcu.edu](mailto:nfurgurson@gems.vcu.edu)) on Monday, 22 January 1996 at 13:10 PST:**

Some witty soul has drawn a smiley on the table! This is the the first TRUELY interactive website I've seen. I am watching the camera move (with CUSEEME) and controlling the camera arm via the web. Very nice!

---



[ [More Comments Here](#) ]

**Figure 4.2:** The web browser interface for Mechanical Gaze during an exhibit featuring various geologic rock formations. In this case Benitoite crystals. The pan, zoom, and location status controls are on the right while the comments and higher level navigation are below the image.

## **Adding and Removing Exhibits**

Since Mechanical Gaze was dynamic by the very nature that it moved and delivered current images from a remote environment, we also wanted to allow the individual exhibits to be dynamic and change rapidly. The only limit on the number of exhibits available was the physical dimensions of the robot's workspace, which was approximately 4000 cm<sup>2</sup>.

Each exhibit contained an entry in the current exhibits file from which the CGI scripts extracted various information. Included in this file was the number of exhibits along with additional information about each exhibit such as the robot's location for entry into the exhibit and the physical bounding volume available for browsing that exhibit. The bounding volume was described by limits set on the length, width, and zoom controls. There were also limits describing the amount of roll and pitch permitted. If while browsing an exhibit, a user made a navigation request that would move the robot out of the legal boundary for that exhibit, an alert page was presented with a description of the illegal motion and help on how to continue browsing.

A unique directory name for each exhibit was also contained in the current exhibits file. This directory contained an introductory HTML file used to describe the exhibit when users requested the list of current exhibits, a description HTML file containing additional information about the exhibit, a header HTML file to be placed at the beginning of each browser page, and a file containing the running dialogue and comments for the exhibit. Usage statistics corresponding to each exhibit were also located in this directory.

The result of this approach was that adding and removing exhibits was quick and easy. To add an exhibit, one placed it into the robot workspace and provided the

introduction, description, header, and footer files. The addition was immediately active by simply inserting the physical location of the exhibit and its boundaries into the list of current exhibits. Removing an exhibit was accomplished by the even easier task of taking its entry out of the current exhibits list. All modifications to the current exhibits list were effective immediately.

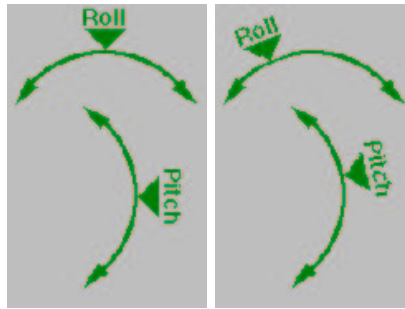
### **User Registration**

One of our goals was to provide all web users unrestricted access to the browser. However, certain features of the system were more effective when reasonably accurate information was known about the user. For example, when leaving comments, it was helpful to tag the message with the name of the user. This helped to identify the user and provided contact information such as an email address and URL pointer to a home page. Allowing users to enter all of this information manually for each comment was not only tedious but problematic. There was little preventing a user from assuming the identity of another user or anonymously dumping pages of garbage text into the commentary. Therefore, we developed a method for users to register themselves by providing a name, email address, and a home page pointer (optional). A password was mailed back to them to be used for registering themselves on subsequent visits. This request for information was not intended to be a violation of a user's privacy. Nor was it intended to be sold or given out. Most importantly, this did not violate our goal of unrestricted access since *anyone* could become a member.<sup>8</sup>

Registered users were granted several additional privileges. When navigating the

---

<sup>8</sup>Anyone with an email address could become a member as a valid email address was required to mail back the system generated password.



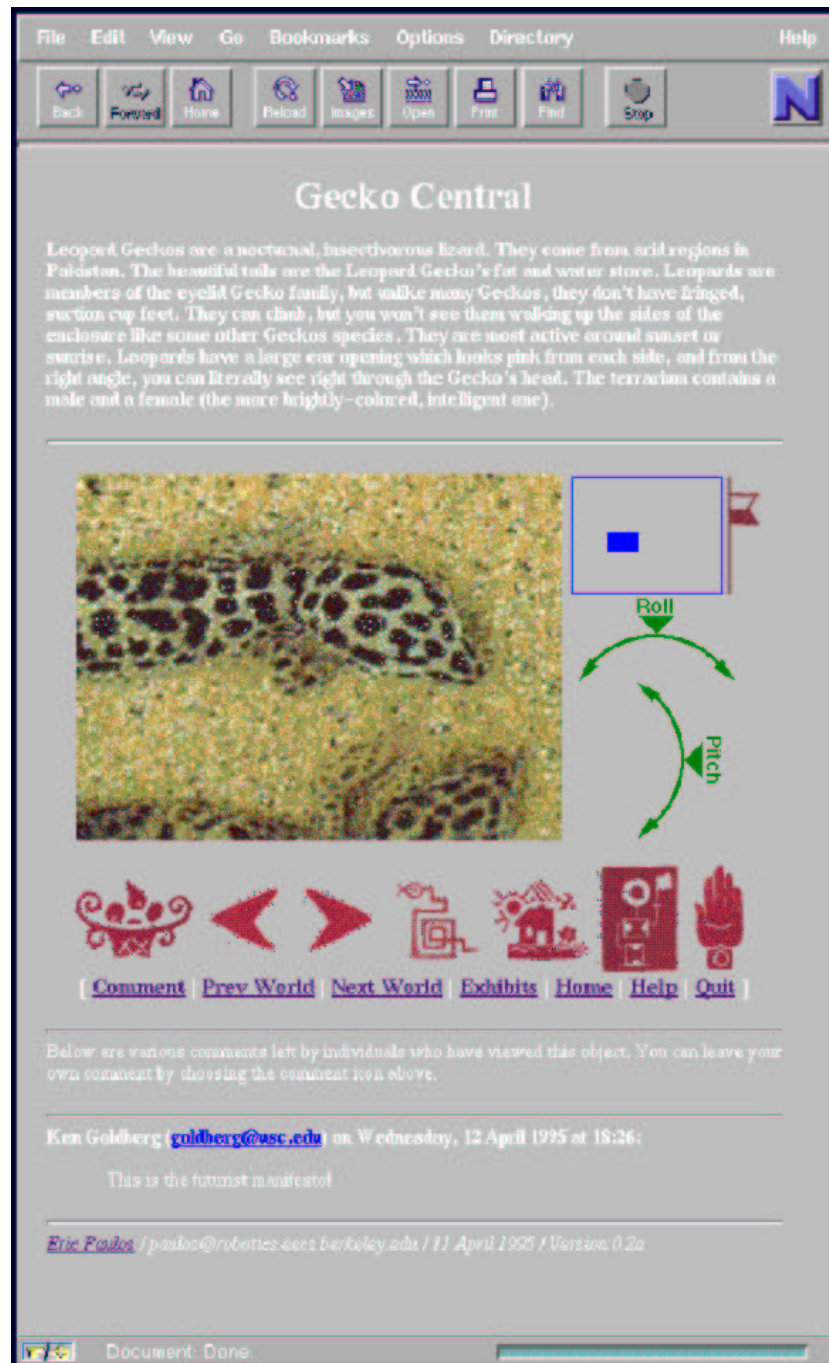
**Figure 4.3:** Two different closeup views of the roll and pitch tools.

robot, they were provided the roll and pitch control tools shown in Figure 4.3. These two tools permitted full control of every robot axis (DOF). For non-registered users these tools were replaced with the simplified zoom in and zoom out buttons to guide the robot shown in Figure 4.2. Also, only registered users were permitted to leave comments about the various exhibits.

#### 4.1.11 Navigational Tools

After receiving a navigation page, a user often wished to change the vantage point of the exhibit and obtain a new image. This modification took place by using any of the navigational tools presented to the user (see Figure 4.2).

One navigation option available to the remote user was to scroll the image. Scrolling moved the camera within the same plane as the current view, captured a new image from that location, and delivered it to the remote viewer in a new navigation web page. This was accomplished by either clicking on a portion of the image (fine motion control) or the location status tool (coarse motion). Fine motion requests brought the selected portion of the image directly into the center of the field of view in the subsequent image while coarse



**Figure 4.4:** Another view of the web browser interface for Mechanical Gaze during an exhibit featuring live gecko lizards. This page demonstrates the additional roll and pitch controls. Also, this image demonstrates an earlier unsuccessful interface tool – a small flag icon that raised up and down a mast used to control the height/zooming. This was later replaced by the more useful thermometer interface described in section 4.1.11 and shown in Figure 4.2.



motions moved the camera to a particular area within the entire defined exhibition space.

Every exhibit allowed a user to zoom in closer to an object for a more detailed inspection, as well as zoom out to achieve a wide angle view. Zooming was accomplished through the zoom navigation tool located on the right side of the image. The camera mimicked the motion of the thermometer indicator. Users could also make selections directly on the thermometer to better control the zooming.

When the system was operating on the Intelledex 6DOF robot, the rolling and pitching tools were presented to registered users. These tools are not pictured in the sample navigation page shown in Figure 4.2, but are shown in Figure 4.4 and separately in Figure 4.3. Choosing a point on the roll or pitch tool would cause the camera to roll or pitch depending upon the selection and deliver the resulting image from the new vantage.

#### **4.1.12 Usage Summary**

Mechanical Gaze continued operation until late 1997. During that time it received a large amount of traffic (and press) as one of the few sites offering a view and control of vantage into a real remote location. We exhibited over a dozen different exhibits from university and private collections. Between exhibits late in its life, we placed a mirror at 45 degrees into the viewing table. For the first time remote users could maneuver the camera over the mirror and gaze through the mirror out into the room<sup>9</sup> housing Mechanical Gaze and its occupants. This proved to be one of the most fascinating features to remote users. Soon email and comments began asking where the room was, what they were looking at,

---

<sup>9</sup>Mechanical Gaze began its life in 127 Cory Hall on the Berkeley campus and eventually moved to 330 Cory Hall.

what a project at the far end of the laboratory was. It was this and other elements that motivated the thought of designing systems that were better equipped to browse a physical space.

## 4.2 Legal Tender

Designed in 1996 in collaboration with Ken Goldberg, Judith Donath, John Canny, and Mark Pauline, Legal Tender became the first publicly accessible tele-robotic laboratory. After giving up their anonymity and agreeing to accept full responsibility for their actions, a remote web user could perform experiments (puncture, burn, stain, *etc.*) on a pair of purportedly authentic US 100 dollar bills. As defined by United States Code, Title 18, Section 333 this is a criminal act. But only if the bills are real, the web site is authentic, and the experiment actually performed. For the first time we could ask the question, “Is a tele-crime occurring?”

### 4.2.1 Motivation

Shortly after Mechanical Gaze began operation, a strange series of emails began to arrive concerning its operation. We expected criticism of its interface, complaints about network delays, and (hopefully) some interest in its use. While we did receive a wide range of feedback on these topics, there was one that was unexpected: the disbeliever.

In 1995, robot tele-operation over the web was in its infancy. Camera hardware and scripts for near real-time interaction were novel and just getting all the robot hardware and interface software working was a challenge. With all of our work with the real system,

we failed to realize that others would doubt its entire existence. After all, we saw and dealt with the intricacies of the system every day. Of course it was real!

Nevertheless, users began to doubt if the system was real. Were live images even taken? Was a robot involved at all? Perhaps the images were all pre-stored and delivered back via a database system. As the Mechanical Gaze users grappled with this question, they began to propose experiments to prove the reality or un-reality of the system. But in fact several of the design issues within Mechanical Gaze worked directly against these experiments. For example, an impressive amount of work went into simply minimizing the latency from a user's navigation request to returning the resulting image. This involved a great deal of fine tuning and optimization with the physical robot kinematics, camera hardware, and local networking software. However, the result was too successful. Users interpreted the quick response time as an indication of the image simply coming off of a pre-stored disk archive rather than a real physical robotic system.

#### **4.2.2 Telepistemology**

Ken Goldberg encountered similar suspicion from the users of his two systems currently in operation at the time, The Mercury Project [Goldberg et al., 1995a] and The Tele-Garden [Goldberg et al., 1995b] (see section 5.2.3 for a further discussion of these two systems). There was a similarity between the questions being raised by the users and creators of these telerobotic systems. Fundamentally, the question revolved around some form of the philosophical question of *epistemology*. Epistemology is the study of the nature, origin, and limits of human knowledge. Ken Goldberg coined the term *telepistemology* to directly address this study of knowledge acquired at a distance [Goldberg and Spaid, 1996,

Goldberg, 2000].

The initial goals of these early Internet telerobotic system were extensions of our “reach” and increasing our access to remote locations and objects. But this “remoteness” coupled with the distributed and decentralized authority of the internet only increased the potential for deception, error, and forgery. As a result these telerobotic systems begin to call into question the issues central to telepistemology: access, agency, authority, and authenticity. A more complete and detailed discussion of telepistemology can be found in *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet* [Goldberg, 2000].

### 4.2.3 Introduction to Legal Tender

It was important to create a situation where a remote user would be called upon to accept responsibility [Milgram, 1974] for their own actions. In this case, their remote tele-actions. Furthermore we wanted to choose a topic where the question of authenticity would be a familiar one that mattered to every user. We chose money. Even those newly familiar with the internet would have little trouble understanding the importance of real versus forged United States currency. Coincidentally, the new anti-counterfeit US 100 dollar bills were beginning circulation in early 1996 and the topic of determining authentic and legal tender was popular worldwide. Thus the title of the project: *Legal Tender*.

Maintaining the anonymity of the Legal Tender web presence itself became important. It was setup without any apparent connection to the people involved or any university affiliation. It operated under the URL<sup>10</sup> [www.counterfit.org](http://www.counterfit.org). It had no ties or links to people or research groups.<sup>11</sup> Its appearance was that of an organization gathering data on

<sup>10</sup>A Universal Resource Locator or web page address.

<sup>11</sup>A `whois` search through a domain registration entity such as Network Solutions did reveal further



**Figure 4.5:** Initial wide angle view of purported US 100 dollar bills in early phases of Legal Tender experiment. The black square represents an area exclusively reserved for the registered user currently operating Legal Tender.

users' perceptions of counterfeit money.

#### 4.2.4 Operation

Remote users were presented with a wide angle image of a pair of purported US 100 dollar bills (see Figure 4.5). Guest users were restricted to moving a high resolution camera around the bill pair and examining the bills.

A user registered by relinquishing their anonymity and email address. A password was generated and mailed back to them through several anonymous re-mailers. This further masked the identity and ultimate intentions of the actual Legal Tender site. Once registered, a user was presented with exclusive access to a small portion of the bill (see Figure 4.6). Only that user could perform experiments and alter that portion of the bill. Experiment selection was important as users were only granted a single experiment. A list of currently available experiments was also presented (see Figure 4.7). Experiments ranged from puncturing the bill to burning the bill with a thermal probe.

Once a user chose to perform an experiment they were reminded that their identity was known. They were also advised that the action they were about to perform was a Federal crime. United States Code, Title 18, Section 333 concerning the defacement of US currency is quoted in Table 4.1.

At this point the remote user could only proceed with the experiment by agreeing to accept all further responsibility for their online action and the resulting remote actions.

Users who agreed to accept responsibility were instructed to wait for up to one minute while

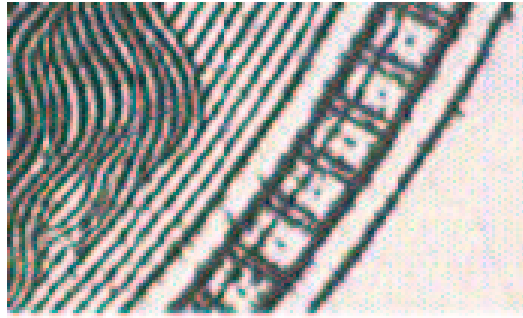
---

information about the project. These searches are more common now and accessible through many web based interfaces. However, at the time of Legal Tender's operation such probes were not commonly known and were accessible only using cryptic UNIX-based command line tools.

### Mutilation of national bank obligations

Whoever mutilates, cuts, defaces, disfigures, or perforates, or unites or cements together, or does any other thing to any bank bill ...shall be fined or imprisoned. [18 U.S.C. 333, 1948]

**Table 4.1:** Mutilation of national bank obligations from United States Federal Code, Title 18, Section 333.



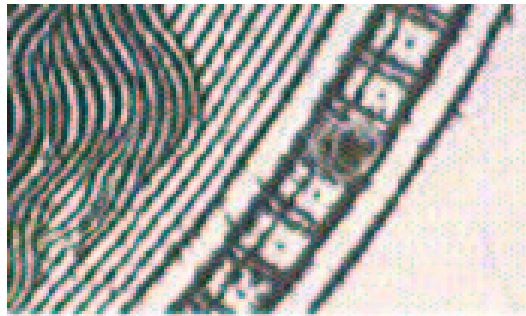
**Figure 4.6:** A typical portion of a US 100 dollar bill presented to a remote registered user for exclusive access.

the experiment was performed. The robot moved the bills into the tele-laboratory and the experiment began. Immediately afterwards another “after” high resolution image was taken (see Figure 4.8). Finally, an updated wide angle view of the bills was taken. After several weeks of use and operation, a typical pair of purported bills would appear as in Figure 4.9.

At the conclusion of the experiment the before and after images were presented to



**Figure 4.7:** A list of available experiments that were available to registered user of Legal Tender.



**Figure 4.8:** A typical portion of a US 100 dollar bill presented after a puncture test.

the remote user for examination and analysis. A series of questions were asked, including “Is this real?”. These answers along with the before and after images are recorded into a publicly searchable archive. A typical experiment image pair is shown in Figure 4.10.

#### **4.2.5 Results**

Legal Tender continued operation through 1996 and most of 1997. During 1996 it was presented at SIGGRAPH [Paulos et al., 1996a] and the Dutch Electronics Art Festival (DEAF) [Paulos et al., 1996b]. At the conclusion of operations the results of all experimental archives were made publicly available. Much of the data was never fully analyzed and remains in files in possession of the author.

The ultimate question of authenticity of the site, bills, and/or experiments has remained a mystery until the writing of this dissertation. In fact, the bills were all real and remain in possession of the author, the site was real, and all of the experiments were real. That is, a real pair of US hundred dollar bills (several times over) were mounted on the end of a robotic arm. Mounted within the workcell were a heated poker and a sharp tip. Upon a request, the robot moved the bills over the appropriate tool, burning or puncturing





**Figure 4.9:** Wide angle view of purported US 100 dollar bills after a variety of experiments by numerous users over the course of several weeks.



**Figure 4.10:** Before (left) and after (right) images resulting from a user's request to perform a thermal test on one of the purported US 100 dollar bills.

a specific portion of the bill. Afterwards, the bills moved above a 45-degree mirror that allowed a wide angle camera to capture the new appearance of the recently modified bills. This images was used to reference into the bills from the main page of Legal Tender. Finally, the bills were moved above a separate high-resolution camera that captured the results of the experiment (*i.e.*, the after image). These before and after images were stored with the user's comments.

## CHAPTER 5

# RELATED WORK

*Art provides exact information of how to rearrange one's psyche in order to anticipate the next blow from our own expanded facilities. The artist shows us how to "ride with the punch", rather than "taking it on the chin".*

*– Marshall McLuhan*

PRoP research is at the intersection of robotics, social psychology, communication, and human-centered computing. As a result it draws on related work from a wide variety of fields. The sensation of embodiment of an individual in a real life distant location has provided more than enough impetus for people to develop remote telepresence systems. This chapter enumerates much of the related work in each of these areas.

### 5.1 Historical Telepresence Systems

Methods of achieving telepresence are not new [Sheridan, 1992]. Even before the word *robot* [Čapek, 1923] had been coined, we find remotely operated mechanical devices, most notably a wireless radio controlled submersible boat by Nikola Tesla in 1898

[Tesla, 1983]. It wasn't until decades later when the need to handle dangerous nuclear materials became important did tele-robotics see a "re-birth". One of the earliest electrically controlled mechanical teleoperational systems was developed by Goertz in 1954 [Goertz and Thompson, 1954]. Many subsequent systems were aimed at safely exploring hostile remote environments such as battlefields, nuclear reactors [Greaves, 1994], deep oceans [Ballard, 1986], mines [Ntuen et al., 1993], and space [Weisbin and Lavery, 1994]. Additional applications for teleoperated surgery [Green et al., 1995] and manufacturing [Draper, 1995] have been explored by several researchers [Sheridan, 1992, Mosher, 1964, Tomovic, 1969, Moravec, 1988].

Most of these system are quite complex, requiring special purpose dedicated hardware to control and interact with the mechanism in the remote environment. In our systems we strived to constrain the development so that they would be accessible to a wide audience without additional, expensive, or extraordinary hardware. In essence, telepresence for the masses.

## **5.2 Telepresence and the World Wide Web**

The spontaneous growth of the web over the past several years has resulted in a plethora of remote controlled mechanical devices accessible via the web.

### **5.2.1 Fixed Cameras**

Some of these early systems employed fixed cameras in remote spaces where users could observe dynamic behavior. The very first of these such systems was the Trojan



**Figure 5.1:** The Trojan Room Coffee Machine image as it appeared in 1994. This  $64 \times 64$  pixel greyscale image was the first “webcam” in operation on the web.

Room Coffee Machine [Trojan Room Coffee Machine, 1994] in early 1994<sup>1</sup> (see Figure 5.1). Remote users watched the consumption and brewing of coffee in a coffee pot in the United Kingdom through a tiny  $64 \times 64$  pixel grey scale image. Later systems allowed users to view a parking lot, observe an office, or keep track of the activities of a favorite pet in its native habitat.

One such novel system was the Wearable Wireless Webcam [Mann, 1996] developed by Steve Mann, which provided web viewers a chance to see the world from the vantage of a camera mounted atop Mann’s head. This project, along with several others including work by Thad Starner [Starner, 1999] and Brad Rhodes [Rhodes, 2000], was the genesis of the wearable computing research community. While the camera mounted atop Steve Mann moved (with Mann’s own movement), the viewers were passive and helpless to change the direction or actions of Steve and his camera.

Another extremely popular early site was the University of Ulm’s Interactive Model

---

<sup>1</sup>The Trojan Room Coffee Machine has actually been available on the internet since 1991 when it was accessible through the X-Windows system as *XCoffee*. Furthermore, it should be noted that the Trojan Room Coffee Machine is still fully operational at the time of the writing of this dissertation.

Railroad [Merz et al., 1997]. Remote users were, for several minutes, in command of a complete model train set. A second camera was later added to provide a closer view. Users would choose a train and could change tracks.

There are now a wide variety of webcams aimed at almost anything or anybody (*e.g.* JenniCam [Ringley, 1996]).

### 5.2.2 Movable Cameras

Systems quickly began to evolve to allow movement and control of the remote camera. This control was accessible directly from the system's web page. One of the first such systems was LabCam [Wallace, 1996] developed by Richard Wallace. His system allowed remote users to aim a pan/tilt camera using an intuitive imagemap interface. Numerous camera systems followed and are now commercially available.<sup>2</sup>

### 5.2.3 Robots on the Web

Progression to intricate control of more degrees of freedom were realized by introducing robots to the web. Ken Goldberg *et al.* [Goldberg et al., 1995a] developed a 3 DOF (Degrees Of Freedom) telerobotic system called the Mercury Project 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. This was the first web interfaced tele-robot, coming online on 1 September 1994. Mark Cox [Cox, 1994] developed an early system for allowing users to request images from a remotely controlled telescope. Users sent observation requests to the fully automatic telescope located on the Pennines of Yorkshire

---

<sup>2</sup>For historical reference, at the time of writing this dissertation, Cannon Camera offers a complete pan/tilt/zoom control camera system with full Java control for around \$1000 US dollars.

England. The system decided when the conditions were right for viewing and acquiring the desired image.

Another early remote robotic system [Taylor and Trevelyan, 1995] developed by Ken Taylor allowed web users to remotely manipulate blocks using a robot with an attached gripper. Soon afterwards, we developed Mechanical Gaze [Paulos and Canny, 1996a], a tele-robotic system where users could control a camera's viewpoint and image resolution to observe various museum artifacts placed within the robot's workspace. Remote visitors created their own dialogue about the various exhibits, which were frequently changed. A complete discussion of Mechanical Gaze can be found in Section 4.1.

By 1995, Ken Goldberg *et al.* had developed another telerobotic system called the TeleGarden [Goldberg et al., 1995b] in which web users are able to observe, plant, and nurture life within a living remote garden. Shortly thereafter the first tele-robotic laboratory, Legal Tender [Paulos et al., 1996a] came online allowing users to assess the authenticity of a pair of US hundred dollar bills through detailed observation and a set of experiments that physically alter the currency. A complete discussion of Legal Tender can be found in Section 4.2.

By 1999 web robots were coming online almost weekly [Siegwart and Saucy, 1999]. Matthew Stein's Puma Paint [Stein, 1998], at Wilkes University allowed users to paint with brushes using an industrial robot arm. The interface had two windows showing live camera views of the work site, and let users manipulate everything from brush saturation, to color, to pressure on the canvas. In addition, completed paintings could be mailed (via US postal service) upon request. The number of robots currently on the web is in excess of several

hundred and far too numerous to enumerate in this text. At least two full day workshops have been conducted recently to discuss this rapidly emerging phenomenon [Siegwart, 1998, Paulos, 1999].

#### 5.2.4 Mobile Robots on the Web

Mobile robots capable of browsing larger areas have also been developed by several researchers. Kaplan *et al.* [Kaplan et al., 1996] developed a simple system that allowed users to drive a small remote controlled car from a simple network interface while receiving back live video. The results were impressive and raised several issues involving user control under varying network traffic conditions and global accessibility, as this system utilized several applications available on only a limited number of computer architectures.

The Learning Robot Lab group at CMU developed Xavier [Simmons, 1998], a web mobile robot that could be instructed to travel throughout a floor of a building at CMU and tell jokes. This system has performed well in this known environment but is not intended to browse spaces without being given a pre-built model. Although models almost always aid in navigation, we chose to design a system that would perform reasonably well even without maps. The trick is to leverage off of the intelligence of the remote user to aid in the guidance problem. As unexpected navigational situations arise, even a well programmed mobile robot will have difficulty negotiating, while a person will not. Recently, there have also been several non-internet controlled mobile robots used for providing information at museums [Willeke et al., 2001].

Unleashing mobile robots on the web soon created serious problem. How should such systems be controlled? These were robots typically designed for use in industry or



research by trained operators. The challenge was to begin to design interfaces for untrained, non-roboticist operators. One early system was Khep on the Web, where remote users control a mobile robot through a maze of mirrors [Saucy and Mondada, 1998]. The system provided two live video views: the vantage of the localized robot and an overhead of the general scene. A user also had control of the overhead camera's orientation and zoom. As with most of these systems, each user gained exclusive control over the robot for 10–15 minutes, after which they were disconnected and unable to reconnect for 30 seconds. This was done to provide “fair” access to the robot as collaborative control schemes provided unsatisfying results.

Another later project at CMU by the Centre for Metahuman Exploration and Big Signal produced systems for control of various robots located in the Antarctic over the Internet [Center for Metahuman Exploration, 1998]. This system was also used for education in classrooms. Online telerobotic systems were becoming increasingly complex and even intentionally dangerous [Pauline and Paulos, 1997b].

Bandwidth increased, latency dropped, frame-rate increased, but the internet was still an unreliable medium. While this posed a minor nuisance when downloading images on a web page, it created a catastrophic event when attempting real time control of a mobile robot around a cluttered, complicated, unknown, environment. One elegant solution was detailed in WebDriver [Grange et al., 2000]. A simple mobile robot extracted useful information from its various on-board sensors including its sonar to present an extremely intuitive Java based navigation tool.

As of the writing of this dissertation, well over several hundred interesting me-

chanical mobile devices and robots are connected to the web. With more spawning daily, we shift the focus of the related work discussion to more specific P<sub>RoP</sub> related systems.

### 5.3 Blimp Related Work

Our early approach to tele-embodiment involved the construction of small internet controlled helium filled blimps (see Chapter 6). We focus on previously constructed blimp related tele-robotic systems in this section.

As early as 1881, the brothers Albert and Gaston Tissandier constructed a 1,062  $m^3$  (37,500  $ft^3$ ) indoor airship propelled by a battery-powered electric motor that appeared at an International Exposition on Electricity in Paris [Tissandier, 1884]. A few years later, Charles Renard and Arthur C. Krebs built *La France*, an indoor airship 1,869  $m^3$  (66,000  $ft^3$ ) which also used electric power.

More recently, there have been several blimp robots built in the last few decades. One outcome of Alan Kay's Vivarium project at the MIT Media Lab in the 1980's resulted in a simple autonomous blimp [Brand, 1987]. Mike Caine and Andy Christian [Flynn, 1993] also developed an autonomous blimp which avoided walls during an event sponsored by the MIT AI Lab in 1988. There has also been some work by the Robot Group in Austin, Texas aimed at developing an outdoor autonomous blimp. However, none of these blimps were designed or developed for telepresence. Most of them were large and intended for outdoor or indoor stadium use and ran autonomously. No one, to our knowledge, has proposed using a blimp as a telepresence device over a network. The originality of our approach is in the combination of lighter-than-air technology for the remote robot with a wide-area network

such as the internet for ubiquitous and inexpensive telepresence.

There is also, for obvious reasons, a tremendous amount of work by the United States Military into Unmanned Aerial Vehicles often referred to as UAVs. Of particular relevance is a project called micro-UAVs which is designing systems about the size of a human hand for over the hill reconnaissance. These systems all carry a small camera and can be remotely (or semi-autonomously) controlled via soldiers. However, most of the work is classified at this time.

Finally, NASA is working on what they call the Personal Satellite Assistant (PSA) [Gawdiak, 1999]. This device contains an on-board camera-microphone-speaker setup as well as a small video screen. While this is not yet possible with small sized, earth based lighter than air systems, it is easy to do in outer space. NASA, with weightlessness on their side, is developing these systems to assist astronauts during space missions within the next few years. A PSA could identify gas leaks, warn astronauts about dangerous temperature changes, or check up on a payload. It could also “sit” by an astronaut, offering advice on an intricate systems upgrade or relay messages from ground controllers. In theory, it could even float up to an astronaut’s night quarters and tell a bedtime story.

## 5.4 PRoP Related Work

Since our public debut of Space Browsers (see Chapter 6) and PRoPs (see Chapter 7) there have been a number of similar mobile internet robotic systems. DEC (now Compaq) Western Research Lab (WRL) has recently developed what they call the Mutually-Immersive Mobile Telepresence Project [Jouppi, 2000]. The project, also titled eTravel, is

composed of a teleoperated robotic surrogate designed from a modified wheelchair base. Its goal is to allow travel to remote locations as a substitute for physical travel. The surrogate has the appearance of its user in the approximate form factor of a person sitting in a wheelchair. Like the P<sub>RoP</sub> there is live video on the eTravel tele-robot. However, the eTravel system, with more hardware than a P<sub>RoP</sub> (multiple cameras and screens), allows for views of the front and both sides of the remote user's head. One of the interesting elements of the system is the inclusion of foveal video-conferencing, which gives a very wide angle view with low resolution at the edges and high resolution at the center. This is tailored to support the human eye's structure which has a high resolution center (the fovea) surrounded by progressively lower resolution peripheral vision that covers a wide field of view. Tests are still inconclusive but the hope is that foveal video-conferencing will be both more compelling and require less bandwidth than traditional constant-resolution video-conferencing. A user controls the system from a specially outfitted "Immersion Room". This is a large room with two wall sized projection screens intended to immerse the user visually in the remote location. Bandwidth requirements for the eight video streams are on the order of several gigabits per second.

*GestureMan* [Kuzuoka et al., 2000] is another recent internet-based mobile robot from the academic research community. While similar to the P<sub>RoP</sub> (*i.e.*, having a pan/tilt camera head, gesturing "hand" and internet controllable base), this research is primarily focused on more intricate and detailed control of the gesturing and hand. It also included more detailed work into laser pointer additions to the gesturing system and pointing cues.

There are also two recent companies producing P<sub>RoP</sub> like systems. ActivMedia is

advertising *PeopleBot* [PeopleBot, 2000], a Human Interaction Robot, for \$10,000 USD. The PeopleBot is a mobile robot 1.25 meters tall with an attached pan/tilt camera head. The system is actually designed less as an internet telepresence robot and more for a variety of autonomous tasks. With its on-board sensing, it's capable of navigating around and avoiding obstacles and people or "playing cat and mouse" with people as their web site claims. It also has on-board speech recognition and text to speech capabilities for interaction with people. This differs fundamentally from PRoPs which are internet controlled telepresence systems rather than the PeopleBot autonomous robot. More clearly, PRoPs are designed to immediately and unequivocally be identified as a remote proxy for another human while the PeopleBot is a quasi-autonomous/telepresence robot/person making their identification to locally embodied users difficult.

Another recent commercially released tele-robot is the *iRobot* [iRobot, 2000]. A ground based mobile robot with a wireless network connection, on-board pan/tilt camera and tilt-able torso/neck, it sells for less than \$10,000 USD. Its sophisticated "flipper" base design allows it to climb steps and overcome simple ground hazards. It also incorporates two way audio and one way video from the robot. It's primary function is for home surveillance and inspection of kids, pets, and babysitters. Its shape, size, and camera vantage are pet or child-like. This physical appearance along with its lack of a screen to support two-way video, high resolution zooming for inspection, and missing pointer hand/arm render it inadequate as a tele-embodiment tool (see definition on page 21). Like the PeopleBot, it also has various autonomous modes which cause it to exist in a similarly confusion quasi-robot/people mode.

## 5.5 Computer Mediated Communication

Computer Mediated Communication (CMC) research has produced important results towards understanding how humans interact with and communicate through technology. Much of this work has been inspirational and extraordinarily helpful in the design of personal tele-embodiment and PProPs.

Ishii's ClearBoard [Ishii and Kobayashi, 1992] is an early example of novel CMC tools. With it, users view and speak to each other through a "window-like" screen showing the image of the other viewer. ClearBoard was designed to integrate interpersonal space and shared workspace seamlessly. One design goal of ClearBoard was to allow a pair of users to shift easily between interpersonal space and shared workspace using familiar everyday cues such as the partner's gestures, head movements, eye contact, and gaze direction.

For nearly a decade Paul Dourish and several others have examined many aspects of CMC. This work identifies other "connection mode styles" for communications using *Media Spaces*. Media Spaces are infrastructures that support continuous tele-connections between remote locations. This differs from traditional tele-connections which are typically designed for single person-person connections. Most of this work investigates various video-presence tools, their acceptance, and applicability both by individuals and groups. RAVE [Gaver et al., 1992] was an Audio-Video Environment media space developed to study many aspects of audio/video CMC. *Portholes* [Dourish and Bly, 1992] was a direct application of this theme for group awareness. The Portholes system was a distributed (PARC and EuroPARC) information service, which captured frames automatically from cameras in mediaspace nodes. The data was sent to each side (replicated) so that it could be displayed

as a panel of frames. Some additional information was also available such as the time the frame was grabbed, email address, and an audio snippet (recorded by the viewer for broadcast purposes). Although the interface was fairly space-intensive and hence a detriment to casual use (as noted in their paper), portholes did provide increased group awareness as well as a place for serious and whimsical interactions that likely wouldn't have occurred via traditionally available communication tools. One of the most interesting findings was the importance of being able to provide for serendipitous interactions. The spontaneity of real life is the richness often left behind in CMC tools. One of the other important results was the ease of human adaptability to the eye-contact problem with video-conferencing. In almost every video-conferencing system there is a physical separation between the camera and video image. Therefore, a user gazing into the eyes of the screen image of a person, hoping to gain eye contact, will appear to be looking down from the vantage of the other user's image.<sup>3</sup> The work by Dourish and others points out that users adapted to this "anomaly" more quickly than expected. Their research confirms that the lack of direct eye-contact in video-conferencing systems is not as detrimental as first time users of systems typically expect.

Similarly, the *Cruiser* [Fish et al., 1992] work examined various methods of remote video interaction and awareness infrastructures. The Cruiser system was designed in an attempt to increase the opportunities for conversation by increasing the number of spontaneous conversations between users. To do this the system allowed users to perform actions such as cruise, auto-cruise, and glance. Cruise allowed users to initiate a sequence (one or more) of 3 second, two-way connections. They disappear if one of the two users

---

<sup>3</sup>This is of course assuming that each user has his camera located atop their monitor.

fails to connect to them. Autocruise was a method of initiating calls between selected users at random times. Finally, glance was a very brief (1 sec), video-only connections to one or a series of people. The system also kept a status line for each user as well as a call history for who had called them.

Even more closely related to this dissertation is work by the Ontario Telepresence Group [Ontario Telepresence Group, 1990] using various *Video Surrogates*. Door Access Control allowed well understood affordances and links to translate into CMC video-conferencing and awareness tools. For example, closing the physical door to your office sent a signal to a conferencing/awareness tool that you were unavailable. Likewise, opening a door indicated a new arrival to be aware of and take notice. Aware of the need for spatially situated cues during video-conferencing and a “body” to represent oneself, the Hydra [Sellen et al., 1992] project was also developed by the Ontario Telepresence Group. Each user is “embodied” within a small desktop device that support two way video and audio. This device is an impressive system allowing meeting style communication to take place around a table or office desk. The effect is that conversational acts such as gaze and head turning are preserved because each participant occupies a place on the physical desktop. Although it is not a mobile system, the P<sub>RoP</sub> work draws heavily from it in philosophy and spirit.

But a Hydra system user could not gesture or even pan and tilt the camera head. GestureCam [Kuzuoka et al., 1994] provided this extension. A master-slave coupling of a camera and input device enabled simple 2DOF gesturing. A laser pointer on the end allowed the remote user to point the remote laser at an object and gesture during the interaction.



More recently this research group has developed GestureMan as described earlier in this chapter (see Section 5.4).

### 5.5.1 Tele-Tactile Experiences

PRoPs are physical systems and their interactions with other people, objects, and environments draw heavily from the results of various other physical interaction systems. One such early system is *Data Dentata* [Goldberg and Wallace, 1993], a one-bit system for tele-handshaking that operated at SIGGRAPH in 1993. A later system, *inTouch* [Brave and Dahley, 1997], employed force-feedback technology to create the illusion that two people, separated by distance, were interacting with a shared physical object. The “shared” object provides a haptic link between geographically distributed users, opening up a channel for physical expression over distance. InTouch is made of two connected objects each consisting of three cylindrical rollers embedded within a base. When one of the rollers is rotated, the corresponding roller on the other distant object rotates in the same way. Two people separated by distance can then play and communicate by moving the rollers or more passively feel the other person’s manipulation of the object. The presence of the other person is thus made tangible through physical interaction with the inTouch device.

### 5.5.2 The Medium Is the Message

Technological advances are increasing our choice of media for transmitting and delivering our messages. This choice of medium through which information is transmitted strongly influences how the information is received, interpreted, and processed by the recipient. Often the technical limitations of the medium also strongly influence the for-

mat and presentation an individual uses when formulating her message, including what information she chooses to convey and what she chooses to suppress. The medium that one chooses is so tightly coupled with its message that it becomes the message. This result is expressed eloquently by Marshall McLuhan's statement that "the medium is the message" [McLuhan, 1963] or more expressively in the quote below:

The electric light is pure information. It is a medium without a message, as it were, unless it is used to spell out some verbal ad or name.... Whether the light is being used for brain surgery or night baseball is a matter of indifference. It could be argued that these activities are in some way the "content" of the electric light, since they could not exist without the electric light. This fact merely underlines the point that "the medium is the message" because it is the medium that shapes and controls the scale and form of human association and action. The content or uses of such media are as diverse as they are ineffectual in shaping the form of human association. [McLuhan, 1963]

How does a person go about choosing the "correct" medium for a given message and why? While a majority of the discussion of this topic lies outside of the scope of this dissertation, the fact that P<sub>RoP</sub>s act as a new medium for human communication and interaction, means that due consideration of P<sub>RoP</sub>s and their message should be studied. There has not been enough development of P<sub>RoP</sub>s to fully understand anything about their social priority ordering<sup>4</sup> or their "message". However, we have examined a small section of the media theory literature to aid in the design of personal tele-embodiment systems and understand how their interpretation and message will affect their use and application. McLuhan himself understood the importance of what we now call CMC (including P<sub>RoP</sub>s). He viewed media technologies as primarily extensions of the mind and nervous system.

---

<sup>4</sup>By "social priority ordering" we mean the natural ordering that occurs when two or more communication modes are presented to a user. For example, when a person walks into the office of another individual that is on the phone, the person will typically terminate the phone conversation to converse with the person standing at their doorway. There is some implied higher priority of the face-to-face communication over the telephone conversation.

Simply put, he described them as extensions of sense: of eye, of ear, of nose, of touch, of taste, of mind. PProPs provide “extended” sensing and transmission of tele-embodiment through their various communication channels. For example, there are audio and video cues as well as non-verbal cues such as gesturing, pointing, gaze, and proxemics, to name a few. McLuhan stresses the importance of this CMC-medium-communication coupling in the following statement:

By putting our physical bodies inside our extended nervous systems, by means of electric media, we set up a dynamic by which all previous technologies that are mere extensions of hands and feet and teeth, will be translated into information systems. Electromagnetic technology requires utter human docility and quiescence of meditation such as befits an organism that now wears its brain outside its skull and its nerves outside its hide. We must serve our electric technology with the same servo-mechanistic fidelity with which we once served our oracle, our canoe, our typography, and all other extensions of our physical organs. But, there is a difference here. Those previous technologies were partial and fragmentary. The electric is total and inclusive. An external consensus or conscience is now as necessary as private consciousness. With the new media, however, it is now possible to store and to translate everything; and as for speed, that is no problem. No further acceleration is possible this side of the light barrier. [McLuhan, 1963]

### 5.5.3 Media Richness

A more concrete example of this phenomenon in practice can be seen in work by Daft and Lengel on the theory of “Information Richness” [Daft and Lengel, 1991]. They studied communication media choices used in business and work settings. Rich media such as face-to-face provided multiple cues, feedback, and a high variety language<sup>5</sup> that enabled people to interpret and reach agreement about difficult, emotional, and conflict-laden issues.

An example of rich media usage would be during salary negotiation. This is not to say

---

<sup>5</sup>High variety language is capable of expressing a wide range of emotions, mood, and undertones of a given message.

that less rich media is not useful. In fact it may be appropriate for routine activities and transmission of literal data. Relaying sales figures is a good example of an appropriate use of low-richness media.<sup>6</sup> The results of their study demonstrated that organizational success was based on the ability of individuals and groups to process information of appropriate richness. For PRoPs it means that society will place personal tele-embodiment into this rich-poor media ordering and as a result will have a profound impact on its appropriate usage in human communication. While its difficult to design for this ordering exactly, a healthy knowledge of this process has and will play into the evolution of personal tele-embodiment.

#### 5.5.4 Social Presence

The concept of *Social Presence* [Short et al., 1976] is another contributing factor in the design of PRoPs. Social Presence is the degree to which a medium is perceived to convey the actual presence of a location and an awareness of the various communicating participants. The research in this field identifies social presence as essential for intense and relational computer-mediated communication.

Individuals prefer to solve collaborative, equivocal tasks through a medium that is able to sustain relationships and facilitate spontaneous, interactive communication. As a result, any system attempting to design tools for supporting such interactions and relationships must provide (1) real-time interactivity, (2) a high social presence, and (3) media richness.

---

<sup>6</sup>McLuhan also comments on this “media richness” with what he calls cool and hot media. He classifies media as either hot or cool dependent on how high their level of interaction. Hot media have very low levels of interaction, whereas cool media demand a high level of interaction.

## 5.6 Social Psychology

As P<sub>RoP</sub>s mature they will find a home somewhere in our social web [Turkle, 1997]. The hope is that by studying the social psychology literature, P<sub>RoP</sub>s can embody the “right” social elements to enable complex themes such as tele-trust and tele-persuasion. Even small successes in these areas will greatly impact the applications and usefulness of tele-embodiment. Tele-work and tele-commuting with personal tele-embodiment tools will become a rich, seamless, satisfying experience. In the following sub-sections we address several of these important issues of social psychology.

### 5.6.1 Trust and Persuasion Online

The media choice also affects trust. A study by Rocco [Rocco, 1998] concluded that online trust breaks down. Groups attempting to reach a consensus using chat and email were unable to instill enough trust in one another to resolve a dilemma, while face-to-face meetings succeeded quickly. There was some hope for online trust but only when previous face-to-face encounters could occur beforehand. Similarly, studies by Connell, Robins, and Mendelsohn have shown what they call a “cyberbias” in actor and observer differences in causal attribution [Connell et al., 1999, Connell et al., 2001]. They used standard getting acquainted experiments varied over a telephone, chat room, and face-to-face media. Their data suggest, among other things, that CMC encourages a degree of solipsism. That is, unlike a simple two-way communication, the computer screen acts, in effect, as a mirror. This mirror inadvertently emphasizes the self as a causal source in the interaction with a disembodied other. Their initial studies suggest that the telephone is the preferred medium

for the tests they conducted so far. Related work focusing directly on email and phone communication has been reported by Naomi Baron [Baron, 1998].

Herbert Clark [Clark and Brennan, 1991] formulates various cost-time-error trade-offs for several media. For example, media lacking appropriate or incomplete richness often omits important cues during communication. This leads to errors and more often wasted time re-transmitting information that may have already been appropriately transmitted in another media. He also explores what he calls the startup costs to initiating a conversation of appropriate richness over limited media.

Numerous other studies [Greenspan et al., 2000, Kraut et al., 1994, Rice, 1992, Williams, 1977] into online trust and persuasion demonstrate the importance of F2F encounters. One goal for PRoPs is to perform as a CMC tool that can be employed during tasks requiring media rich human interactions. That is, PRoPs are designed to excel in situations where subtle communication cues are necessary.

### **5.6.2 The Media Equation**

Personal tele-embodiment involves some form of technologically mediated interaction. Therefore, understanding and employing the work of Reeves and Nass and their “Media Equation” [Reeves and Nass, 1996] becomes extremely important. At the core of their thesis is that humans, whether they admit it or not, apply human social rules, expectations, and responses when interfacing with technology that is obviously not intended to be human-like. For example one of their more eye-catching conclusions is that even people who state that they do not consider a computer as a social actor do, in fact, respond to computers as if they were social actors. People don’t just use media as tools, they also

react to them socially. This also influences the perceived credibility of these systems in significant ways [Fogg and Tseng, 1999]. This means that people will treat the technological interface of PRoPs with various social rules and expectations. PRoPs are systems intended to facilitate social interaction. However, a disregard of these understood “default” media equation responses may work against the social rules the PRoP is attempting to promote.

### **5.6.3 Non-Verbal Communication**

One goal of personal tele-embodiment is to facilitate human communication and interaction at a distance. Much of this relies on identifying and distilling various elements or channels of communication. Most of these are non-verbal cues and we have drawn insight from important work in this field [Feldman and Rimé, 1991, Knapp and Hall, 1972, Burgoon et al., 1996, Goffman, 1981, Goffman, 1967, Norman, 1992]. These references provided us with a fundamental understanding of the manner in which various non-verbal cues manifest themselves in human communication. When designing PRoPs and other personal tele-embodiment systems, we drew heavily from these works. The forms of rich human interaction that we desire from personal tele-embodiment will never be possible without understanding and incorporating these cues into our mediated communication technologies.

## **5.7 Human-Centered Computing**

The work in this dissertation is part of a growing theme of Human-Centered Computing (HCC) [Human-Centered Computing, 1998]. HCC is an interdisciplinary con-

sortium<sup>7</sup> united behind a common goal to design information systems into human contexts and to better understand those contexts using computational tools.

HCC was organized as a response to the unprecedented impact of information technology on peoples' lives. The researchers understood that technology concepts such as ubiquitous computing were fast becoming a reality. Such systems were assured to affect many aspects of the way people live and work. HCC focuses the attention of building these computing systems with a human-centered focus.

---

<sup>7</sup>Computer science, sociology, psychology, information management, education, business, linguistics, electrical engineering, and mechanical engineering to name a few.



## CHAPTER 6

# SPACE BROWSERS

*I, too, desired to go ballooning.*

*– Alberto Santos-Dumont*

Drawing from our previous internet tele-robotic projects and the social literature on presence we set out to design a device called a Space Browser (see Figure 6.1).<sup>1</sup> These telepresence tools were essentially small human sized helium filled airborne internet tele-operated blimps with two-way audio and video capabilities. This chapter describes the evolution and development of several of these inexpensive, simple, networked tele-operated mobile robot (tele-mobot) designed to provide tele-embodiment (see Chapter 3).

In the rush into cyberspace we leave our physical presence and our real-world environment behind. The internet, undoubtedly a remarkable modern communications tool, still does not empower us to enter the office of the person at the other end of the connection. We cannot look out their window, admire their furniture, talk to their office mates, tour their laboratory, or walk outside. We lack the equivalent of a body at the other

---

<sup>1</sup>Portions of this chapter have appeared previously by the author in *Ubiquitous Tele-embodiment: Applications and Implications* [Paulos and Canny, 1997].

end which we can move around in, communicate through, and observe with. However, by combining elements of the internet and tele-robotics it is possible to transparently immerse users into navigable real remote worlds filled with rich spatial sensorium<sup>2</sup> and to make such systems accessible from any networked computer in the world, in essence: tele-embodiment (see Chapter 3).

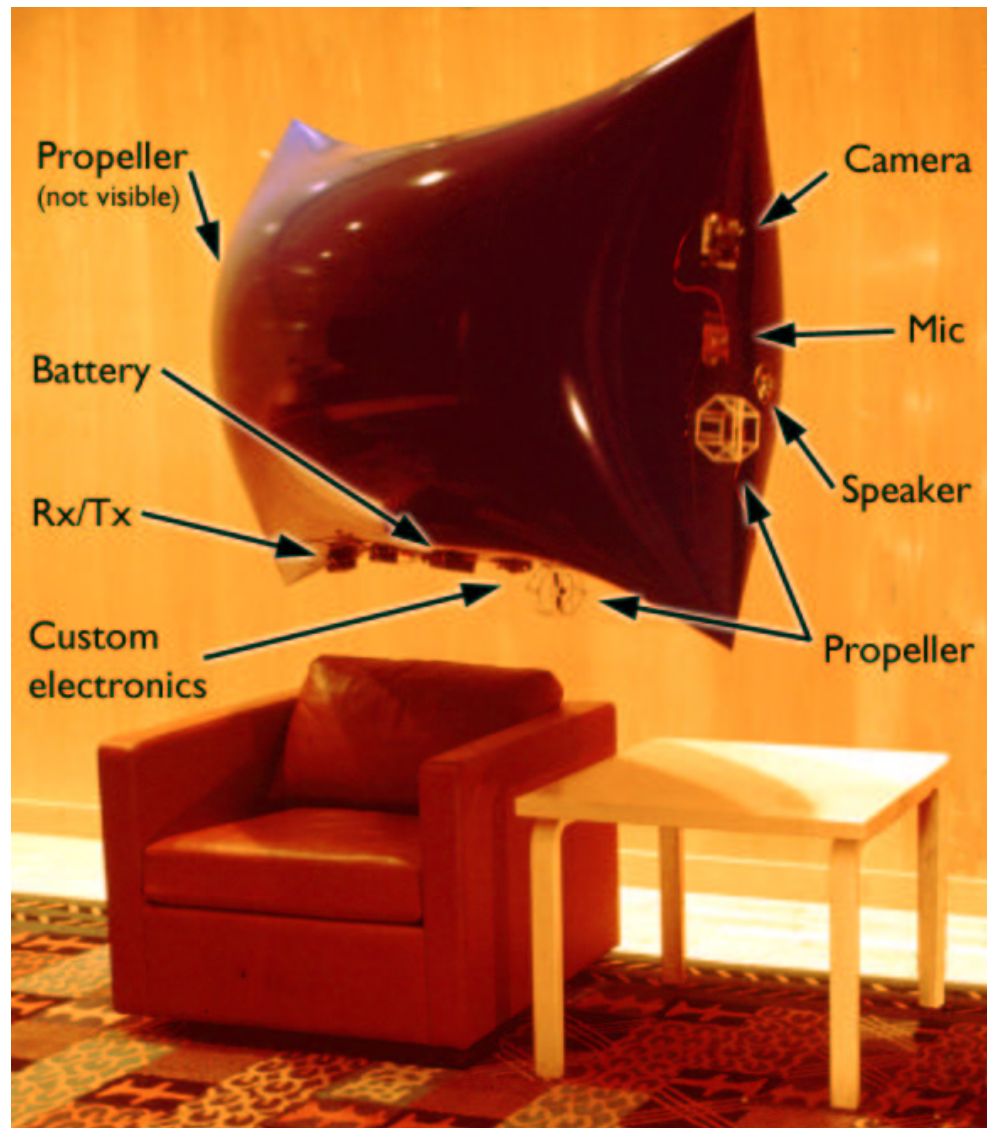
## 6.1 Introduction

Inexpensive, ubiquitous tele-robots provide a form, a physical mobile proxy, that can be moved around in a remote space. We coined the term *tele-embodiment* (see Chapter 3) to emphasize the importance of the mobile manifestation. Tele-embodiment puts today's physical workplaces, stores, homes, museums, and other real physical spaces online. More importantly, tele-embodiment supports most human activities in those spaces, far beyond what is possible with conventional telecommunication. It preserves some of the sensory richness, warmth, complexity and unpredictability of the real world. Many intangibles such as atmosphere, morale, chaos, efficiency and communication channels in an organization can only be perceived through unstructured, spontaneous visits to the groups within it. This was the essence of David Packard's strategy of "management by walking around" as expressed by him below:

It soon became apparent that the instructions the engineering department gave the factory people were not adequate to ensure that every step would be done properly. [Packard, 1996]

---

<sup>2</sup>The parts of the mind concerned with the reception and interpretation of sensory stimuli. More broadly it refers to the entire sensory apparatus



**Figure 6.1:** A blimp tele-robotic Space Browser traversing within a building. A small table and chair are shown for scale reference.

Most of us still shop by wandering the shelves, looking for specials, seeing the item we want, and asking about its features. 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. Ubiquitous tele-embodiment removes the barriers of distance and offers a reasonable facsimile of instantaneous travel to anywhere on earth from any networked computer.

In this chapter we describe one approach to solving this problem using an inexpensive, simple device called a “space browser”. A space browser is an unteathered helium filled blimp with various electronics, propellers, a camera, microphone, and speaker that allows a primary user to browse a remote location by floating and flying around within the remote space.

## 6.2 Motivation

For decades the telephone has been within easy reach to provide us with real-time voice communication [Fischer, 1992]. But that isn’t a substitute for being there. It’s still necessary for us to journey from one coast to another for the experience of a conference, trade show, business meeting, or friendly social visit. More recently, the internet has proved itself as a remarkable communication tool, allowing us to exchange text, video, and images, with anyone whose interests we share, professionally or socially.

Several methods of communication have developed within this new medium. Multi-User Dungeons (MUDs), sometimes also called Multi-User Dimensions [Curtis, 1992], and

Internet Relay Chat (IRC) lines which are mostly text based systems, already allow us to enter and interact with others across vast distances in shared virtual spaces, often modeled after some real space we inhabit.

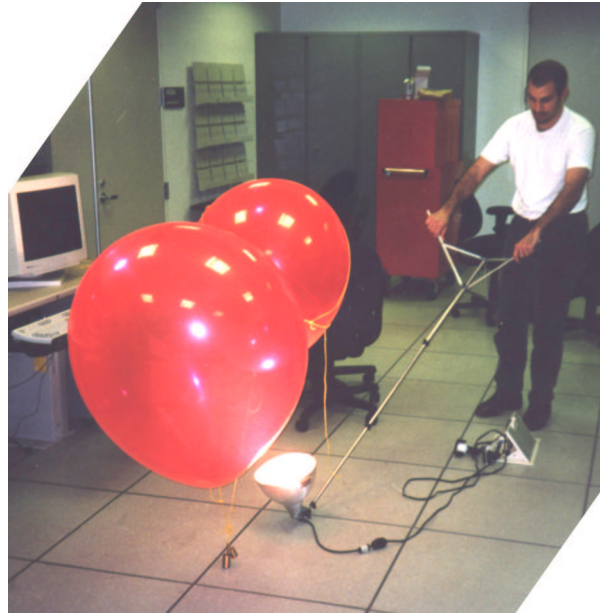
In the last few years these virtual spaces, enhanced by the development of internet based interactive virtual reality standards such as VRML [Pesce et al., 1994] and Moving Worlds [Mitra et al., 1996], have become enriched with three-dimensional graphics, texture-mapped furniture, and metaphorical representations of humans or avatars [Stephenson, 1992]. Entire vivid and inviting three-dimensional virtual worlds are evolving daily on the internet. By their nature they are fascinatingly playful creations, allowing us to exist in a world free from earthly physical constraints and enter a space where we can fly through the air and pass seamlessly through walls. Just as the development of the web has sprouted from the amazing diversity of users operating within it and allowed anyone to publish, so we will find people from all walks of life becoming modern day VR-architects. Many of their creations will indeed be magnificently realistic and interactive. However, we will still lack an interface to the real world in which we live, work, and play.

Furthermore, these virtual systems are only as rich and complex as the models placed within them and often lack the sensory richness, warmth, complexity, chaos, and unpredictability of the real world in which we live. For example, while in real life you may knock over and break a cup, spilling hot coffee onto your friend's lap, current modeling limitations prevent such complex and "scalding" events from transpiring. We expect these virtual world systems to improve dramatically. Newly developed sophisticated behavioral modalities for virtual world avatars combined with more realistic model-

ing tools [Mirtich and Canny, 1995] will be the major contributors. However, there will still exist an enormous gap between the richness of the real world in which we live and these virtual modeled worlds.

Video tele-conferencing on the internet provides an arguably more realistic interface into a remote space, but it is more of an enhancement to existing telephone communication technology than a new form of communication. With video conferencing 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. 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 realized that it was necessary to deliver a more realistic perception of physical embodiment of the user within the remote space being explored. Such a system must immerse the user in the remote world by providing continuity of motion and user control of that motion. These elements would provide the user the visual cues necessary to stitch together entire visual experiences into a coherent picture of a building and its occupants. We also wanted to provide the user with the means to communicate and interact with the remote world and its real inhabitants using this new system. Furthermore, we wanted such a system to be accessible to any user on the internet with standard software running on currently existing computer architectures.



**Figure 6.2:** One of the first images of early space browsing blimp experimentation in the lab in 1994.

### 6.3 Goals

One of our goals in the design and creation of tele-operated blimps was to attempt to provide the primary user with a more immersive experience of tele-visiting a remote location than current systems provided. Furthermore, we constrained our set of solutions by requiring our design to be usable from any networked computer, regardless of any special hardware. Clearly, this second goal was quite challenging to meet. However, we feel that, barring some extremely extenuating circumstances, any system developed today should be accessible to the entire internet community. Finally, our system must be inexpensive, leveraging wherever possible off of existing hardware. The importance of these goals merits a more detailed description which we elaborate in the following subsections.

### 6.3.1 Realism

Drawing from experience in robotics, networking, and interface design, we set out with a goal to design a system that would deliver a more realistic perception of physical embodiment of the primary user within a remote space being explored than current CMC tools. Such a system must immerse the user in the remote world by providing continuity of motion and user control of that motion. As David Gelernter suggests in his book *Mirror Worlds* [Gelernter, 1992], such systems that gaze into remote spaces should show each visitor exactly what they want to see. This requires the system to provide millions of different views from millions of different focuses on the same object. Certainly visitors will desire to zoom in, pan around, and roam through the world as they choose. More importantly, they should be permitted to explore this space at whatever pace and level of detail they desire. Users should also be free to swivel and rotate their vantage, to get a better look at regions that might be obscured in the initial perspective. All of these elements provide the user the visual cues necessary to stitch together the entire experience into a coherent picture of a building and its occupants. Users should also be provided with the means to communicate and interact with the remote world and its real inhabitants using this system.

### 6.3.2 Globally Accessible

The web provides an amazingly appropriate medium for developing inexpensive, publicly accessible telepresence tools. Although the web can feel extremely restrictive at times, especially when attempting to design an intuitive user interface to a complex robotic system, the benefits of global usability and user familiarity outweigh the drawback of not



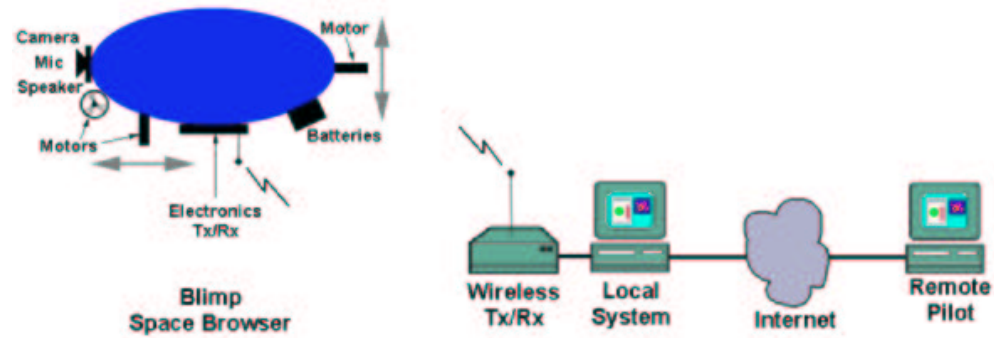
being able to implement a full telepresence system. Certainly, we could have chosen to construct custom navigation software for users to download, or even required users to purchase special head-mounts and hardware attachments to interface to their computers. While this would allow us more freedom in the design of the overall system, it would severely restrict the accessibility of the system. Since we consider the quantity and diversity of users on the web as one of its most powerful features, we chose to constrain the development of our system within the accessibility of web users.

### **6.3.3 Inexpensive**

A final constraint in our design was to develop a system with mostly existing components. We knew that only by designing an inexpensive system could we contemplate the idea of easily and affordably reproducing such systems and using them to populate a variety of spaces and applications. Such widespread use would hopefully result in their adoption into society as a common tool, much as the telephone is today. The ultimate goal: global, ubiquitous tele-embodiment. That is, personal tele-embodiment tools that are so ubiquitous that any location a primary user desires to access is easily visited by a nearby available PRoP.

## **6.4 Space Browsers**

To meet these goals we developed a device we call a Space Browser (see Figure 6.3). Essentially, a space browser was a helium-filled blimp of human proportions or smaller with several lightweight motors directly connected to small propellers and no other moving parts



**Figure 6.3:** Schematic of basic space browser configuration

(see Figure 6.4). On board the blimp was a color video camera, microphone, speaker, simple electronics, and various radio links. The entire payload was less than 500 grams (typically 250–450 grams). Our design choice was to use the smallest sized blimps that would carry only the necessary cargo, thus making them easily maneuverable down narrow hallways, up stairwells, into elevators, and through doorways. We have iterated through countless blimp configurations and have flown blimps ranging in size from 180×90 cm to 120×60 cm. The smaller blimps consumed about the same space as a standing person and were thus well suited for moving into groups of people and engaging in conversation with minimal disruption. Even under full power a blimp moved at human walking pace. This provided not only a realistic sensation of walking but also mapped well onto the expected delays in the video and motor control signals to and from the blimp though the internet. Blimps did not incorporate any other human-like appearance. In the future, people may desire to customize their blimp as they would a virtual world avatar or skateboard. But for now the remotely embodied user does not have control over the form they will take via the blimp. Therefore, for now we have tried to maintain a, hopefully, generic external appearance.



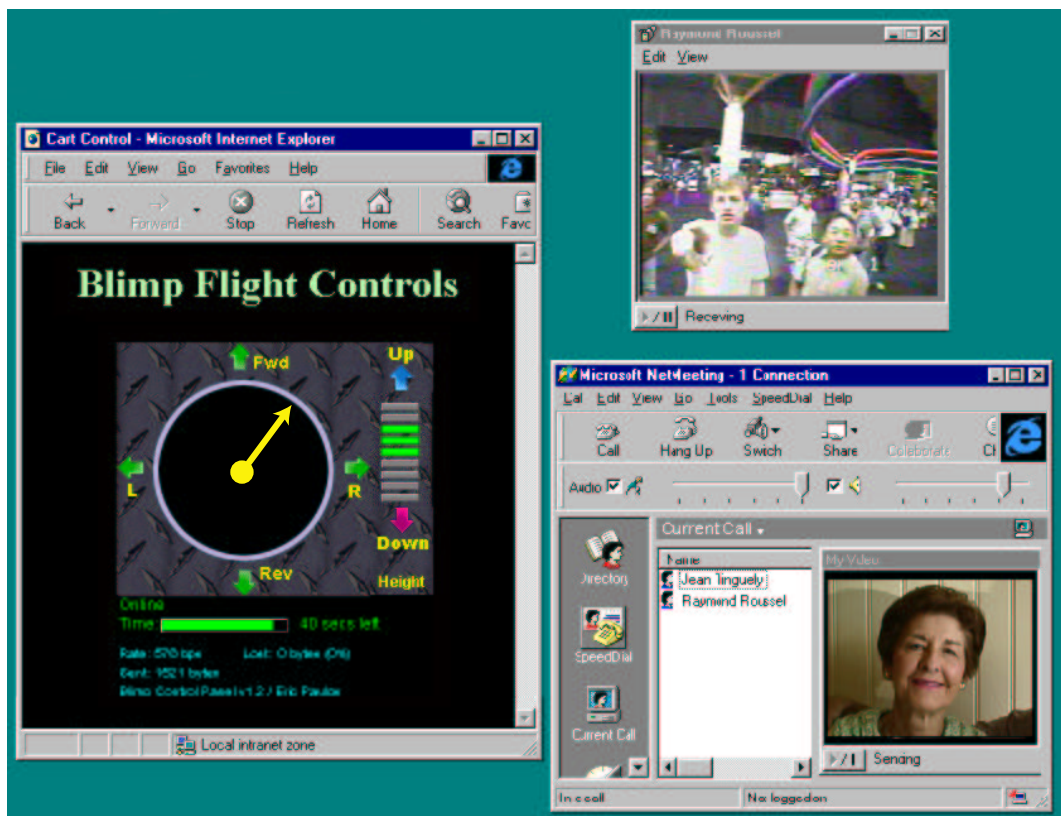
**Figure 6.4:** An medium-sized blue polyurethane space browsing blimp interacting with a person.

A user, anywhere on the internet, could use a simple Java applet running within a Java [Gosling and McGilton, 1995] enabled browser to pilot the blimp (see Figures 6.5 and 6.6). As they piloted the blimp up or down, right or left, the blimp delivered, via wireless communications, live video and audio to the pilot's machine through standard teleconferencing software. The pilot observed the real world from the vantage of the blimp while listening to the sounds and conversations within close proximity to the blimp. The user could choose to converse with groups or individuals by simply speaking at their desktop or laptop machine with the sound being delivered via the internet and then a wireless link to the blimp's on-board speaker. The many sensory stimuli delivered by these tools is remarkably similar to those experienced in the real space.

The system used publicly available free teleconferencing tools initially developed for use on the Internet Multicast Backbone (MBone) [Deering, 1991]. The video tool, *vic*, developed at Berkeley [McCanne and Jacobson, 1995], implemented H.261 video compression and was capable of delivering multicast or unicast network packets. Typical video frame rates of 2–4 fps were adequate for pilots to navigate. Audio was delivered by a similarly developed tool called *vat*. All of these tools were available for most UNIX and PC computers, making space browsers easily accessible to most users.

#### **6.4.1 Advantages**

The blimp offered the possibility of a wide range of spontaneous, group interactions. Telephones and teleconferencing are intrusive media. The recipient must interrupt whatever they are doing to answer the call. The interaction is either one-on-one, or within a pre-arranged group. A space browser cruising by a group can overhear the conversation,



**Figure 6.5:** The Java user interface to the space browser along with live video and audio feed from blimp as it appears on the remote pilot's computer.



Figure 6.6: Closeup view of the Java applet user interface to the space browser.

recognize the group members, and decide if it is appropriate to enter the conversation.

Our aim was not to replace direct human interaction but rather to extend it. Space browsers allowed for mobility in the third-dimension which was a wonderful ability in a building, gallery, or large hall. However, the blimp tele-mobot also offered the user a perception that was in some ways greater than the experience of an actual visit to the remote location. This is because the blimp primary user could travel into places and deliver views from perspectives where no human could travel, such as high above a production facility or conference floor.

A blimp tele-mobot had no problems cohabitating with humans. A collision of a blimp with a fixed obstacle or a person was harmless. Even an out of control blimp posed no real threat to people, while such a statement could not be made for other mobile robots. However, their low weight also meant that they could not be used to transport anything, nor could they open a door, or even push an elevator button.

While most mobile robots with wheels and tracks struggle with varying floor surfaces, and most cannot handle stairs or even a single step, the blimp tele-mobot simply flew over such problems. The blimp also avoided many of the difficult problems encountered by wheeled and tracked mobile robots when they enter a room littered with books, files, or cables on the floor. In fact a large selection of difficult robot motion planning problems were avoided by being airborne.

#### **6.4.2 Disadvantages**

Space browsers were far from perfect in that they were quite high maintenance. Although there was a motor which provided lift for the blimp, it was not intended to be used

continuously to keep the blimp aloft. Instead the helium level in the blimp would have to be carefully balanced so as to maintain a near neutral buoyancy in the air. This was sometimes difficult in buildings with drastically varying air temperatures. Worse yet, although it was not susceptible to small air currents created by people walking by it, air conditioning and ventilation ducts have caused more than a few problems with navigation. In at least one instance a blimp was actually sucked onto an air duct and held captive until the air duct could be turned off. It is difficult to solve these problem since it would almost certainly require more powerful motors to combat these unexpected air bursts. However, larger motors translate into larger blimps and at some point the size of the blimp would severely restrict its ability to navigate, rendering it useless to browse spaces within a building.

Stringent weight limitations allowed for only a small amount of batteries to be carried on-board, yielding flights of about an hour before batteries required replacement. Although replacement was quick and rather straightforward, the process still prevented the blimp from operating in a round the clock manner, as desired. Since the blimp is a robot, we could imagine an automatic system that would cause the blimp to break free from the user and return to some form of docking station to replace its own batteries. However, such systems are extremely difficult to implement and are at the very core of some of the most difficult problems in the field of robotics.

Although novice users have flown the blimp with little or no instruction, the blimp still exhibited erratic behavior. Unlike almost all other robots it was nearly impossible to bring the blimp to a complete halt. We devised several methods to solve this including one which would literally drop an anchor-like device to the ground. Constructing a more



active method to perform this is difficult since accurate location information is not readily available on-board the blimp to aid in such precise positioning. Although the system was simple in design, with the pilot possessing only controls for lift (up/down), turn (left/right), and thrust (forward/back), it often turned into a complex mental challenge to move around. One such problem originated from the fact that a user typically wants to maintain a constant height while flying around, but instead must manually burst the lift motor at regular intervals to maintain that height. To solve this we incorporated a simple lightweight sonar device onto the blimp to help maintain a constant height. Likewise, various simple aerial acrobatics such as nice ninety-degree turns and flying forward without rotating slightly left or right were difficult. In an attempt to solve this we incorporated a simple, inexpensive electronic compass weighing less than 25 grams. Both the compass and sonar were carried on board and thus are not susceptible to the network delays experienced by the user attempting to correct these problems remotely.

Finally, flying is unfortunately not an innate human ability. To compensate for this we developed a blimp pilot simulation tool [Barrientos et al., 1997] which was interfaced directly through the same Java based navigation window. Rather than controlling an actual blimp, the user “flew” an accurately modeled blimp in a virtual three-dimensional world in which all collisions and dynamics reflected real world dynamics. This portion of the system was directly based on *Impulse* [Mirtich and Canny, 1995], a tool for accurate dynamic simulation of physically-based systems. Thus, users could inhabit our blimp flight simulator as they awaited their turn controlling the real blimp, or they could watch the live blimp video feed while its current pilot navigated.

## 6.5 Results

While we unfortunately did not have the resources to conduct a full scale user study of the blimps in action, we were able to gain a wealth of anecdotal information from numerous uses and test flights with real users in a variety of settings. Overall, pilots found the immersive experience of driving a blimp tele-mobot to be very compelling. The motion was smooth and natural. We concluded that people were not threatened by the blimp as they typically approached it almost immediately and engaged the remote pilot in conversation. Most flights took place in Soda Hall, the computer science building on the UC Berkeley campus. However, on several occasions we brought one or more blimps out of our lab and flew them where people could interact with them away from a laboratory environment. We did this not only to gauge the level of acceptance and usability of blimps by humans but also to test out the flexibility of our design in terms of ease of transport, setup, and operation in a remote space. The latter being extremely important if we have hopes of such devices becoming ubiquitous home consumer items.

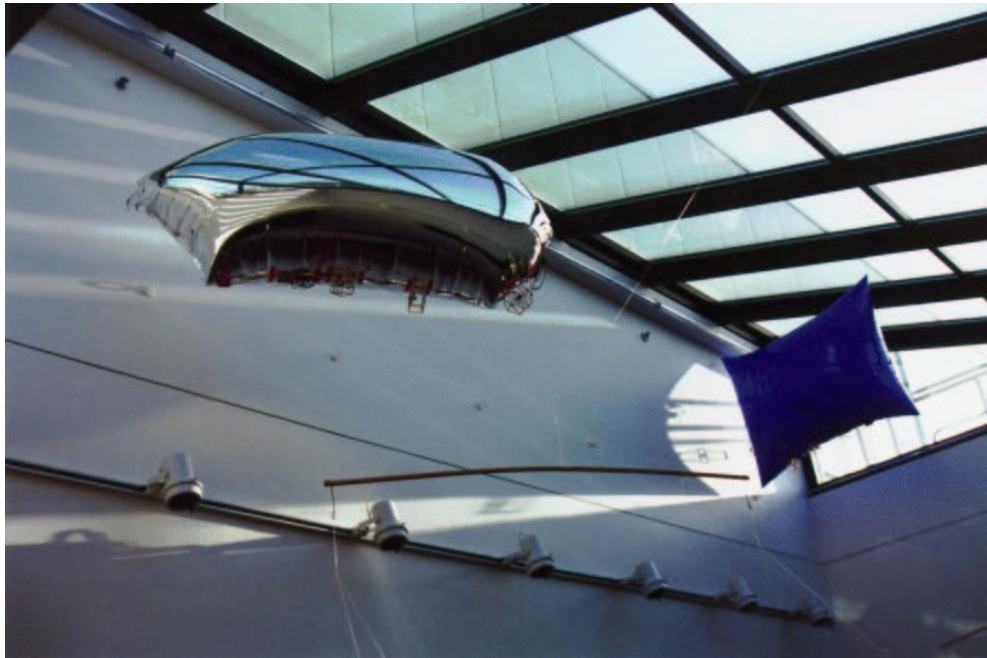
Although it may not be apparent, there is little limitation to the number of blimps that could operate in the same space at the same time. In fact we have flown more than one blimp on several occasions both in our lab and at remote locations. Figures 6.7 and 6.8 contain a collection of a few of the more than ten different blimps constructed. Several flights were made at the Exploratorium, a hands on technology museum in San Francisco, providing the blimp and its remote pilot the opportunity to interact with a wide range of people, including small children. In fact during this test several children accidentally placed their fingers directly into the protective housing for the propeller. Although the plastic



**Figure 6.7:** One of the smallest space browsing blimps designed at just under a meter in length. Using Mylar greatly decreased the blimp size but radically increased its fragility. Impressive blimps with short lives.

propeller struck their fingers with full force, none of the children were injured, reinforcing our claim that space browsers can safely cohabit with humans.

Several blimps also made the trip to the Digital Bayou at SIGGRAPH in New Orleans where they provided a remote proxy for people at the conference and for those that could not attend [Paulos and Canny, 1996b]. Other locations include navigating through and surviving a hostile environment where heavy industrial machinery was in operation, the UC Berkeley Paleontology Museum including browsing a Tyrannosaurs Rex skeleton several stories high, and cross country flights where individuals in New York successfully piloted blimps in California. There was also an *Eyeball* blimp (see Figures 6.9 and 6.10) specifically constructed for the Ars Electronica festival in Linz, Austria where it was flown in 1997. In several of these locations additional publicly available networking tools such as



**Figure 6.8:** A pair of space browsing blimps in operation. The small silver mylar blimp and the much larger blue polyurethane version in the atrium of a building.

the Video Gateway [Amir et al., 1995] were employed to tunnel internet traffic from local networks to more desirable portions of the internet which were capable of routing multicast packets. This allowed anyone connected to the MBone<sup>3</sup> to view and listen to the blimp's video and audio feed.

## 6.6 The Future of Online Blimps

Undoubtedly, countless remotely controlled physical devices will come online in the next few years. Many of them will provide the sensation of tele-embodiment. Technology often moves at a blinding pace and future predictions are difficult to make. However, online

---

<sup>3</sup>The MBone defines the subset of the internet capable of transmitting and receiving multicast packets. Multicast traffic is loosely defined as an efficient one to many broadcast protocol on the internet.



**Figure 6.9:** The *Eyeball* blimp flying within the Design Center at Ars Electronica in Linz, Austria in 1997.



**Figure 6.10:** Another view of the *Eyeball* blimp flying within the Design Center at Ars Electronica in Linz, Austria in 1997.

blimps or some reasonable facsimile will fly into our lives someday soon. These future blimps will have numerous improvements. We discuss a few future challenges for these systems.

Although the blimps performed well under a variety of conditions, there are numerous improvement that could be made. Many of the improvement ideas that come to mind can immediately be eliminated due to weight restrictions. For example, one may wish to add some form of Global Positioning System (GPS) on-board. While this may be possible in the future, current GPS systems weigh on the order of 250 grams.

Some goals are better interactivity, video display of the pilot on-board the blimp, a movable laser pointer that users can use to point at objects, and an eyeball motor for quickly changing the viewpoint of the camera without moving the entire blimp. We have also discussed various ideas that would allow blimps to transport small items, automatically recharge their battery packs, and travel outdoors. The high motion style video generated by the on-board blimp camera calls for the adoption of new video compressions standards such as H.263 which benefit from the incorporation of motion vectors and half pixel interpolation. At some point we imagine much higher levels of control where users can place the blimp into a follow mode which will cause it to follow the person or object within the video image. Likewise, one might develop a system that allows a person to select a feature in the video image which the blimp will then approach. One could also incorporate it as a useful tool into existing research for creating realistic synthetic views of existing architectural scenes by using the blimp as a platform for generating a sparse set of still photographs [Debevec et al., 1996]. Fundamentally, our research draws as much on the sociology of group interactions as on sensing and actuation techniques, and we need the

former to drive our choices for the latter.

## 6.7 Discussion

We believe that many of the technologies needed to support ubiquitous tele-embodiment are in place now, and they are surprisingly inexpensive. Our entire space browser prototype cost under two-thousand US dollars (in 1996). In quantity it should be possible to place such systems within the range of the typical home consumer. Compared to an email or a video-conferencing session the blimp appears to be a richer experience. However, compared to a true visit the space browser lacks many qualities. We do not have enough experience yet to say whether something essential is lost and if so how to compensate for it. Nevertheless, it is clear that these systems represent a new tele-communication form. Our position is that while the telephone and video teleconferencing lack some essential elements of presence in the remote space, space browsers do not. With their low cost and enormous commercial potential, it is quite possible that such devices will become ubiquitous home computer accessories.

Widespread use of mobot proxies will result in many mobot-mobot encounters. In that case, it may make more sense to meet in a real space remote to all individuals. Imagine a lunch break where you meet several friends from around the world to fly together, each with separate space browsers, through the latest exhibits in the Louvre.

Realistically, it will be a long time before the majority of people feel comfortable passing seamlessly between their immediate real world and the reality of tele-embodied worlds. In the mean time, space browsers and other forms of tele-mobots will provide a way



for an individual, no matter how far away, to participate in a, hopefully, rich and natural way, beyond traditional videoconferencing, within another environment.

While space browsing blimps flew overhead, it was clear that to facilitate human communication and interaction they would need to be back down at human eye-level. We explore the design of the next generation of blimps we call PRoPs in Chapter 7.



## CHAPTER 7

# PRoPs

*If all of a sudden someone just stopped the production of Robots...that would be a terrible blow to mankind. – Helena and Dr. Gall in Karel Čapek's R.U.R.*

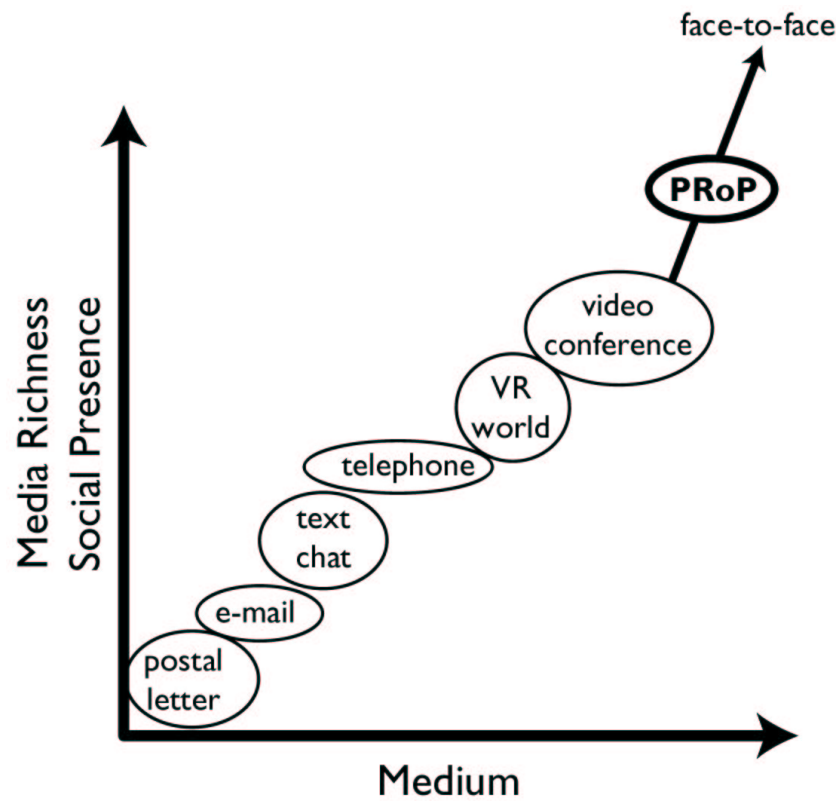
Chapter 6 explored fascinating floating, airborne tele-operated robots that, unfortunately, floated away in mid-conversation. High maintenance and excruciatingly brief operational sessions doomed Space Browsers as useful tools for human telepresence, at least for awhile. A new approach to human communication and interaction using online telepresence was needed. In this chapter we describe the evolution and development of this new system of internet-controlled, untethered tele-robots or PRoPs (Personal Roving Presences). These devices strive to provide the sensation of tele-embodiment (see Chapter 3) in a remote real space. These devices support at least video and two-way audio as well as mobility through the remote space they inhabit. The physical tele-robot serves both as an extension of its operator and as a visible, mobile entity with which other people can interact. PRoPs enable their users to perform a wide gamut of human activities in the remote space, such

as wandering around, conversing with people, hanging out, pointing, examining objects, reading, and making simple gestures.

We are also interested in understanding how PRoPs can capture and transmit what Daft and Lengel refer to as Media Richness [Daft and Lengel, 1991]. This is essentially the ability for people to determine and use the appropriately rich communication media for the task at hand. The concept of *Social Presence* [Short et al., 1976] is another contributing factor in the design of PRoPs. Can PRoPs extend this richness beyond other Computer Mediated Communication (CMC) tools such as standard videoconferencing? Can they support high levels of Social Presence? Although never intended to replace, “face-to-face” (F2F) interactions, can PRoPs move towards this richness and capture some of the subtlety of F2F? A graph of this concept is depicted in Figure 7.1. Studies [Rocco, 1998, Greenspan et al., 2000, Kraut et al., 1994, Rice, 1992, Williams, 1977] of online trust and persuasion demonstrate the importance of these F2F encounters. One goal for PRoPs is to perform as a CMC tool that can be employed during tasks requiring media rich human interactions. That is, PRoPs are designed to excel in situations where subtle communication cues are necessary.

## 7.1 Introduction

We already live in a society accustomed to ubiquitous telecommunications. Telephones are in every office, cellular phones are in many automobiles, and many individuals are reachable at any time via a pager or personal phone [Fischer, 1992]. More recently, the internet has increased our tele-connectivity by allowing us to exchange text, images, sound,



**Figure 7.1:** A graph (not to scale) illustrating the placement of PРоPs in Computer Mediated Communication (CMC) compared to various media in terms of their support for Social Presence and Media Richness. This is not an accurate graph and is only intended to provide a flavor of the role of PРоPs within the CMC landscape.

and video with anyone whose interests we share, professionally or socially.

But obviously something is missing from these tools compared to direct human contact. This difference is so important that we expend a great deal of time and money traveling to experience these direct contacts. There is something about the extended experience of “being there” and the gamut of activities that we can perform when physically present that makes us prefer to be there in person.

We do not believe that we can ever replace true human-human interactions, nor is it our goal to do so. However, we do feel that it is possible to identify and distill a number of human behavioral traits or skills that are inherent to human communication, understanding, and interaction. Employing computer networking and robotic technologies to implement these traits, our goal is to ultimately provide a compelling overall experience for both the remote and local users and more importantly to create a usable system for tele-embodiment.

### **7.1.1 PRoP Design**

In 1996 ground based Personal Roving Presences (PRoPs) first rolled out of our laboratory. They were designed from simple remote controlled toy car bases and augmented with a 1.5 meter vertical pole. Attached to this pole were a color video camera, microphone, speaker, color LCD screen, a pan/tilt “head”, and a movable “arm/hand” for gesturing. Interfaced through free teleconferencing software, a joystick and mouse directed the PRoP through its remote world. With these terrestrial PRoPs, we observed a dramatic improvement from Space Browsers (see Chapter 6) in the quality of remote social interactions. PRoPs provided a visible, mobile entity with which others could interact.

### 7.1.2 Outline

This chapter explores the historical evolution and design of these systems by tracing PRoP development through at least four unique versions. While each of these sections contains information concerning the hardware, software, and interface for each system, a more complete description can be found later in section 7.9. Readers without interest in the detailed development of PRoPs and only in discussion of their current design, analysis, and applications should skip sections 7.3–7.7 entirely. A background discussion behind the compression techniques used by PRoPs is found in section 7.8. A detailed analysis of the elements of the tele-experience can be found in section 7.10 followed by a brief discussion of common PRoP questions and issues in section 7.11. Further details of various PRoP interface, control, and navigation techniques can be found in Chapter 8. Chapter 10 contains a more in depth exploration of the social implications arising from this work.

## 7.2 What's In a Name?

The name PRoP was chosen for a variety of reasons. First, although PRoPs qualify under the definition as a robot, we prefer to make an intentional effort to avoid the use of the word “robot”. Hollywood, science fiction authors, and others have bombarded our society with images of robots as one of two main themes – either blood thirsty sophisticated killing machines<sup>1</sup> (*Blade Runner*, *The Day the Earth Stood Still*, *RobotCop*, *Westworld*) or autonomous companion androids (*Star Wars*). As a result almost every person has an expectation of a robot that falls somewhere between that range. In an attempt to escape

---

<sup>1</sup>This is not so surprising as this is how robots were portrayed in the play *R.U.R.* [Čapek, 1923] in which the word robot was coined.

this preconceived notion, we refer to the system as a P<sub>RoP</sub> even though it clearly takes on a somewhat robotic form. The key word here is *form*. P<sub>RoP</sub>s are not designed to be a new robot in *form*, but instead entirely new systems that strive to provide and facilitate remote human social interaction. An analogy would be if the telephone was simply called a “special speaker/microphone”.

“Robot” also conjures up images of production-line welding machines and 1950’s sci-fi tin-men intent on threatening the human race with extinction. In fact its original (and still current) meaning is the reduction of humans to simple, repetitive machines. The banishing of will, individuality, and emotion in favor of speed, efficiency and precision. The replacement of skilled, situated activity with mindless repetition (a robot keeps trying to drill holes when the work-piece is missing, the drill is off, or the bit is broken). This is the opposite of what future personal social tele-embodiment systems will be. They will support a wide range of behaviors, which will be highly interactive and situated. They will be “anti-robotic” in the sense that “robot” connotes today.

We also wanted to emphasize several themes that are paramount to P<sub>RoP</sub> research. P<sub>RoP</sub>s enable personal tele-embodiment (see Chapter 3). The P<sub>RoP</sub> is a *personal* device that personifies a user in a remote space. It is not an autonomous system nor is it under the control of some collective group [Paulos et al., 2001]. To underscore this individual control and representation we chose the term *Personal*. Unlike traditional video-conferencing, P<sub>RoP</sub>s allow for mobility, (*i.e.*, *Roving*) and have a physical form or body (*i.e.*, *Presence*). They stand in as a remote proxy or delegate for a person, thus they enable a remote presence or telepresence. Thus P<sub>RoP</sub>: Personal Roving Presence. So what’s in a name? Hopefully,



the first step to a new thought process about personal tele-embodiment.

### 7.3 History: Getting Here from There

While Mechanical Gaze (see Section 4.1) mimicked viewing objects in a display case quite well, that was all it could do. Users were unable to glance up, observe other inhabitants, walk around the room, and interact with real people. Overall the experience was unsatisfying and had little in common with an actual visit to a museum. A user needed more than just control of a remote camera but a whole “body” with which to personify herself, browse a remote space, and interact with its inhabitants. From that work we moved towards Space Browsing blimps (see Chapter 6) and coined the term *tele-embodiment* (see Chapter 3) to emphasize the importance of this physical mobile manifestation

Blimps performed well for browsing spaces. However, it was their engagement and interaction with people that proved to be the most interesting application. Unfortunately, human interaction proved to be unsuitable for the blimps. Even with on-board sensing and control, stopping the blimp to chat was an arduous task. Floating in a mild indoor draft, the blimp, as if an inattentive child, would slowly drift away in mid-sentence. Traveling outdoors had more serious consequences - a blatant act of blimp suicide. Even a gentle breeze would hijack the blimp far off into the sky. Frequent helium refills exacerbated the blimp’s already unpleasantly short battery life. Worse, stringent weight constraints restricted the addition of any new hardware for exploring non-verbal communication cues such as gaze and gesture. PRoP development demanded a return to earth.



**Figure 7.2:** One of the earliest PRoP prototype systems. A modified remote control car base with attached hardware (circa 1996).

## 7.4 Early Prototypes

The first few PRoPs were quickly assembled from modified remote control car bases. The standard motors were replaced with gear motors to increase the torque and decrease the speed to a walking pace. Battery power was also increased and a 1.5 meter PVC pole mounted with an attached camera. An image of one variation of this early system is shown in Figure 7.2. Many of the pieces for these early PRoPs came directly from our existing stockpile of space browsing blimp hardware. With less of a concern for weight we were able to use amplified speakers, a two degree of freedom (2DOF) pointing device, and a pan/tilt camera head. A more in depth discussion of the final development of these elements can be found later in Section 7.10.

These early systems, like their predecessors the blimps, used a complicated wireless

configuration. Video from these early P<sub>RoP</sub>s was transmitted via an analog 915 MHz video/audio transmitter. The signal was received at a nearby computer with video and audio capture hardware. This hardware performed any necessary compression and forwarded the signal (*i.e.*, image/sound) on to the remote driver. The driver sent control signals to that same networked computer where they were relayed via another 72 MHz PCM radio link. Audio was sent to the P<sub>RoP</sub> along yet another 145 MHz FM signal.

The software was also adapted directly from that of the blimps. Audio and video were relayed using *vic* and *vat* while control signals were sent using datagrams from Java code. The keyboard was used for mode selection while the mouse drove the base, controlled the head, and moved the hand. But it was becoming clear from even our own experience that radical re-design was inevitable.

## 7.5 P<sub>RoP</sub> 0

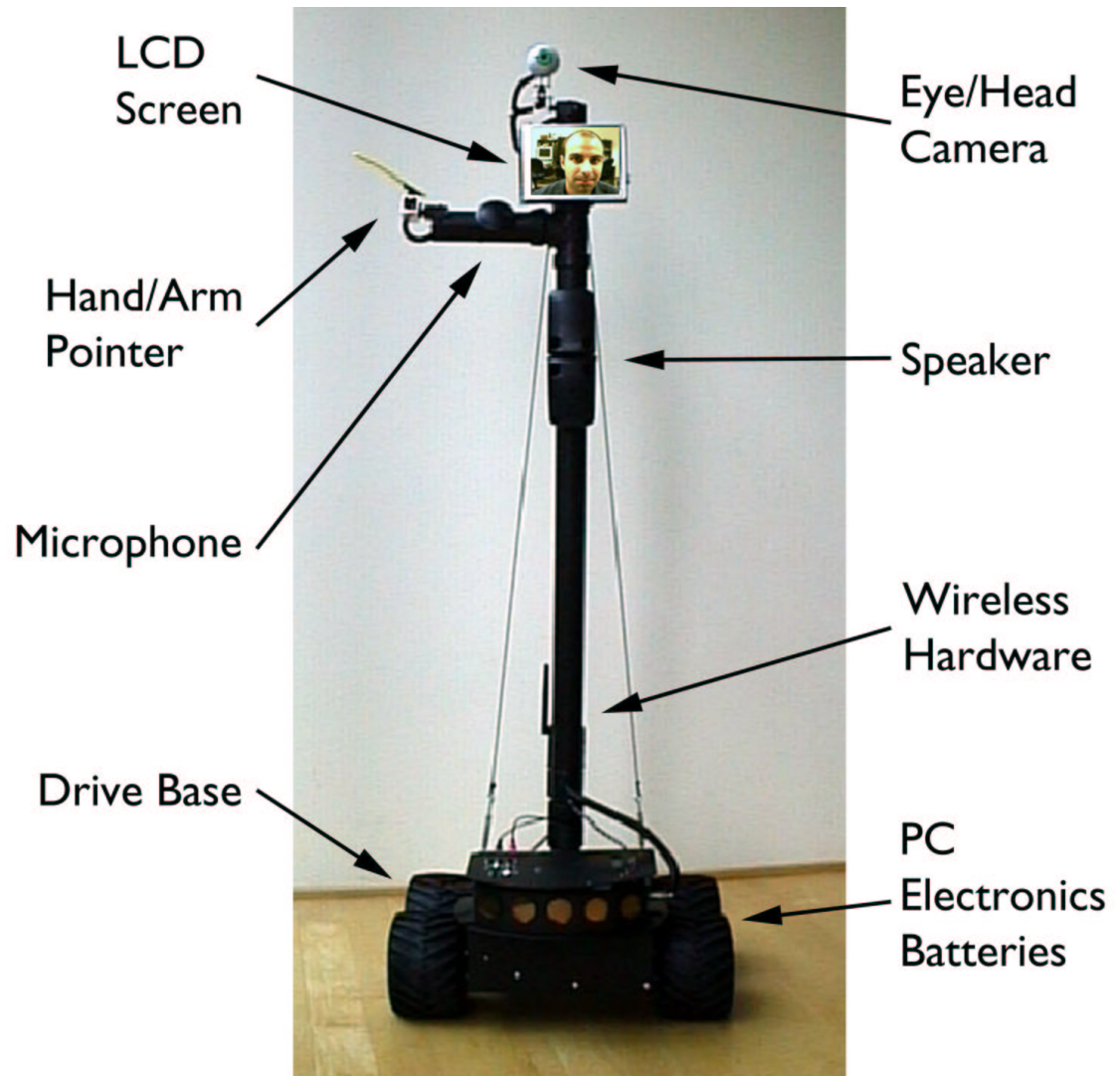
Tuning and maintaining clear signals on all three wireless links was a nightmare and one of the reasons we had originally moved away from the blimp design. After several partial designs, we set out to create the first real P<sub>RoP</sub>, christened P<sub>RoP</sub> 0.<sup>2</sup> It came to life in our 410 Soda Hall lab in the Spring of 1998.

### 7.5.1 Hardware

P<sub>RoP</sub> 0, shown in Figure 7.3 was the first system to move to a single wireless signal. The goal was to move all of the computational hardware on-board. The idea had been with us since working with the blimps but space and weight constraints prohibited development

---

<sup>2</sup>Seeing as this is a Computer Science degree, P<sub>RoP</sub> numbering starts with 0.



**Figure 7.3:** PRoP 0 with camera head, video, LCD screen, controllable “arm/hand” pointer, microphone, speaker, and drive-able base.



**Figure 7.4:** A view of the hand/arm pointer, eyeball camera head, and display screen on PRoP 0.

of such a system. A single board PC/104 format Intel based Pentium computer road along under the hood. It managed two serial connections, one to control the robot base, the other to control the pan/tilt head and pan/tilt hand. A single wireless connection sent and received data on a Metricom wireless modem.

We also removed ourselves from the robot base building business, purchasing a Pioneer 2 base from ActivMedia [ActivMedia, 1994]. The base had a modest payload of about 20kg. The kinematics of the four wheeled system were tank-like. That is, each wheel pair (left and right) were electrically and mechanically linked. Driving the wheels together propelled the base forward and backwards while driving them against each other caused the base to turn or rotate on its axis as a tank. The base also contained seven ultrasonic sonar range transducers and odometry encoders.

Mounted atop a 1.5 meter vertical pole was a speaker, microphone, 15cm diagonal



**Figure 7.5:** Closeup view of pan/tilt eyeball camera mounted atop PRoP 0.

color LCD screen, a color camera and gesturing pointer (see Figure 7.4). Interestingly, most of our previous camera work was with blimps, which required small lightweight cameras. The solution: spy cameras used for clandestine surveillance. By design these cameras were supposed to be undetectable and when we mounted them onto PRoP 0 that's exactly what happened. People often missed the fact that there was even a camera on-board. In an attempt to make the camera more literal we mounted it inside of an eyeball mount (see Figure 7.5).

### **7.5.2 Software**

Improvements were made to portions of the software in PRoP 0 as well. The entire system was now running on the most popular operating system of the time, Windows 95. The most notable improvement was the move to Microsoft NetMeeting 2.1 for audio and video. This allowed us to use an H.263 video codec for the first time. For further discussion

on the advantages of this see Section 7.8. NetMeeting was also freely available, easy to use, and allowed users to correctly tune their audio with an audio tuning wizard.

The control software also migrated to J++, Microsoft's version of the Java programming language. This provided easy access to much of the hardware specific devices necessary for P<sub>RoP</sub> control. It allowed us to add joystick input as well as higher performance serial communication.

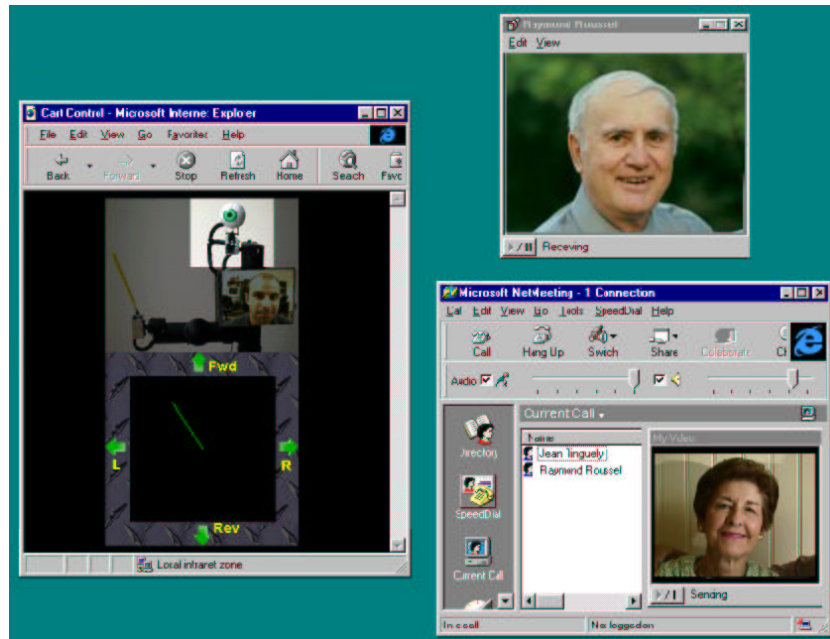
### 7.5.3 Interface

Any Microsoft Windows based PC could be used to control P<sub>RoP</sub>0. This was a major improvement from the previous systems which were mostly UNIX-based requiring installation of *vic* and *vat*. NetMeeting provided the audio/video interface. A window displayed live video from the P<sub>RoP</sub> while a web page, downloaded from the P<sub>RoP</sub>'s web server, contained the control Java applet (see Figure 7.6).

The mouse controlled driving left and right, forward and backward while the joystick controlled the head and hand motions. A user selected the head/hand mode with the two buttons on any commercially available joystick. The joystick "trigger" button controlled the hand panning and tilting while the top button controlled the corresponding head motions. The Java applet displayed feedback about the current mode selected by highlighting the head or hand icon. It also reported the status of the various systems (*i.e.*, current velocity, direction, head pan/tilt, and hand pan/tilt).

### 7.5.4 Disadvantages

P<sub>RoP</sub>0 operated for several years and provided us with the first glimpse of life co-



**Figure 7.6:** PRoP0 interface with Java applet (left) and NetMeeting application (right). Notice the highlighting of the head (*i.e.*, eyeball) in the Java applet indicating head pan/tilt control mode selection.

habitating with PRoPs (see Figure 7.7). PRoP0 also provided us with a tremendous amount of design feedback. Mostly, that we had made some poor choices. The wireless networking hardware we used, a Metricom wireless modem, was only capable of 28–56kb/s data rates. The advantage was that its wireless infrastructure was already deployed throughout the San Francisco Bay Area. Deploying a PRoP simply meant dropping it down on any street or location in the entire region and turning it on. An elegant plan, but not quite reality. The bandwidth was too small and coverage too spotty. While it proved reasonable for reading email or browsing the web, it was unrealistic for supporting continuous media such as a video and audio signal.

Exacerbating the situation was the CPU we used. High-end for embedded single board computers at the time, it operated at only 266 MHz. The constraints of the low





**Figure 7.7:** PProP0 in action interacting with the author (left) and several others (right).

wireless bandwidth required extensive CPU usage to compress the video and audio streams to fit down the narrow pipe. This CPU was simply not up to the challenge.

After adding the body (*i.e.*, pole), screen, pan/tilt motor hardware, speaker, and microphone, the PProP base had reached its payload limitation. The increased weight meant a heavier load on the base drive motors. This, coupled with the draw from the CPU, screen, and other added hardware, resulted in an operation life of less than an hour. Simply increasing the CPU speed was no longer an option. Any speed increase would draw more current and require more batteries, a luxury we couldn't afford. Suddenly we were back to the same problems as the blimp: battery lifetime and weight.

The interface was also problematic. With mode selection a user could not look around while gesturing. And driving was even more of a problem. To stop, a user needed to zero the motion signal, which required looking at the Java applet window while positioning

the mouse in the center of the control box. Worse, walking away from the P<sub>RoP</sub> or accidentally moving the mouse caused the P<sub>RoP</sub> to move in unexpected ways. Later some of this was remedied by sending drive signals only when the mouse button was depressed. This acted as a “dead-man’s switch” but did little to improve its usability. We would have to wait for future P<sub>RoP</sub>s for interface improvements.

Overall, a tremendous amount of time was spent on design modifications and hardware sub-system improvements. As elements were added the weight increased. Similarly, it became clear that higher performance processing would be required. But the current P<sub>RoP</sub> base was not equipped to support the necessary payload or battery capacity for the new computational components. In the end, after almost a year, P<sub>RoP</sub> 0 was abandoned and the design phase began for what would eventually become P<sub>RoP</sub> 1 .

## **7.6 P<sub>RoP</sub> 1**

While valuable research had been extracted from P<sub>RoP</sub> 0 over several years of experimentation, it was time to design a new P<sub>RoP</sub> testbed. A system that would permit new designs and draw upon lessons learned. In March of 2000 P<sub>RoP</sub> 1 sparked to life in our lab.

### **7.6.1 Hardware**

P<sub>RoP</sub> 1 was full of improvements. Constructed from a larger base, the ActivMedia Pioneer 2AT, provided 30kg of payload (see Figure 7.8). The base kinematics were the same but with larger wheels and higher torque motors. It was also packed with

252 watt-hours of hot swappable batteries. This provided enough power for a new 600 MHz Pentium-III based single board embedded computer.

The expanded power and payload allowed us to increase the screen size to 30cm diagonal and run SVGA resolutions. We also moved to a 915 MHz based wireless ethernet system supporting 1.2 Mb/s connectivity and 100 meter range.

During this time there had also been considerable research into alternate camera solutions. Stereo was considered but we were unsure if the stereo imagery would be as beneficial as some of the many other needed improvements. We knew that we needed a wide angle views for navigation and higher resolution narrow zoomed views for inspection and detailed work. Several fixed focus cameras were tried. One problem was that of consistency. That is, when a user toggles or switches to another camera during a task, the field of view change is so disruptive that it often takes a second or two to re-associate oneself with the vantage and familiar landmarks. Also, the fixed focus cameras only operated at a fixed distance. That is, a piece of paper viewed with the high-resolution narrow focus camera would only be in focus at exactly the right fixed distance. This was a task beyond the ability of most P<sub>RoP</sub> users. We knew that we needed to move to an auto-focus camera.

Recall that when we transitioned from working with blimps to P<sub>RoP</sub>s, we used many of the same pieces of hardware. The design constraints for the blimps called for small, lightweight cameras. The ideal choice was small spy cameras designed to be hidden. When we placed these on early versions of P<sub>RoP</sub> 0, the locally embodied users often missed the cue that a camera was onboard. In an extreme move we placed an eyeball around the camera as seen in Figure 7.5. However, most locally embodied users immediately made comments



Figure 7.8: PRoP 1

about this “eyeball” using words like “scary”, “frightening”, and “eerie.” With PRoP 1 this “frightening” eyeball camera was abandoned and a new Canon VC-C3 color pan/tilt auto-focus camera was installed. Not only a much higher resolution camera at 460×350 pixels, the Canon featured user controlled 10× optical zooming, and auto-aperture control.

### 7.6.2 Software

The operating system followed the normal progression of the time to Windows 98. Already relying heavily on much of the J++ hooks into Microsoft’s proprietary operating system (read, non pure-Java), we opted to migrate all of the code to Microsoft C++ using Visual Studio. One hope was for a modest performance gain. The other was to eliminate the use of multiple network socket connections for various data transfers and simplify the code to operate entirely within H.323 (see Section 7.8 for further discussion on this topic).

NetMeeting 3.0 already supported all of the video, audio, and multimedia control using H.323. This meant that any software or hardware supporting H.323 (and there are many) would be able to interface to a PRoP. By using direct C++ calls to NetMeeting’s SDK we were able to send and receive the control signal data along the T.120 data channel. T.120 is part of the suite of protocols encompassed by H.323.

### 7.6.3 Interface

Moving away from a Java applet within a browser allowed some flexibility in designing the interface. A single compiled application would launch NetMeeting, connect to PRoP 1, establish the necessary audio and video connections, and display status information about the control signals.

We completely eliminated the mouse as a control or input device for the system, instead relying entirely on the joystick. The “trigger” button acted as a “dead-man’s switch” to drive PRoP 1 forward/backwards and left/right. The top switch allowed mode selection for the camera head. In addition the “throttle” control found on most joysticks was used to control the camera zoom. Throttle up and down mapped to zoom in and out. From Figure 7.9 it is clear that remote users were able to position themselves at appropriate distances from objects such as a table.

Zooming was a major improvement and finally allowed PRoP 1 users to read signs, drive up to desks, look down, zoom in, and read standard 10-point printed text. And after we added a wide angle lens, navigation became easier and, more importantly, a user could gaze down and see the front of the PRoP base. This vantage was particularly important for aiding in navigation, particularly when a driver had entangled the PRoP against an object or attempted to squeeze through a narrow opening.

#### **7.6.4 Disadvantages**

Increased power, increased bandwidth, increased computational power, all enriched the tele-embodiment experience. But the added hardware came at a cost. Operational time was still at best an hour. Even with the added CPU cycles and bandwidth, our hardware was limiting us to QCIF or SQCIF video image sizes rather than the CIF sizes we had hoped for.<sup>3</sup>

There was also a resentment by users towards installing a separate piece of software

---

<sup>3</sup>CIF stands for Common Image Format. It is a video-conferencing standard to express video image sizes. CIF is defined to be at 352x288, while QCIF (Quarter CIF) is at 176x144, and SQCIF (Sub-Quarter CIF) is 128x96.





**Figure 7.9:** An interaction with PRoP 1. A person and remote PRoP users are collaborating about some passages from a book sitting on the table.

on their computer. Rather than pointing a user's web browser to a particular location for launching the Java applet, an individual was required to install a separate application on their machine.

It was becoming obvious that the overwhelming problem was the base kinematics. The tank design was not designed for easy turning. First, when turning on axis, the motors were at their maximum power consumption. The current drain was often enough to begin to sag the entire power system, often resetting some of the subsystems and exacerbating the debugging process. The problem was even more severe on carpet or a non-slippery surface. However, even on a slippery floor this tank-like turning mode turned the body/pole into an inverted pendulum – swing back and forth as the P<sub>RoP</sub> 1 rotated. Although P<sub>RoP</sub> 1 was never in danger of toppling, it created a dizzying effect for the remote driver as the camera sent back a swaying image. Meanwhile individuals near P<sub>RoP</sub> 1 during this maneuver cringed as they saw it struggle to rotate about its axis and sway in what appeared to be a frighteningly precarious manner. Worst of all, it was this “turn-on-axis” maneuver that users tended to employ most often during navigation. It was clear a remedy was needed.

## **7.7 P<sub>RoP</sub> 2**

P<sub>RoP</sub> 2 came online in early October 2000. Sporting numerous new hardware elements and design changes it became one of the first user friendly P<sub>RoPs</sub>.



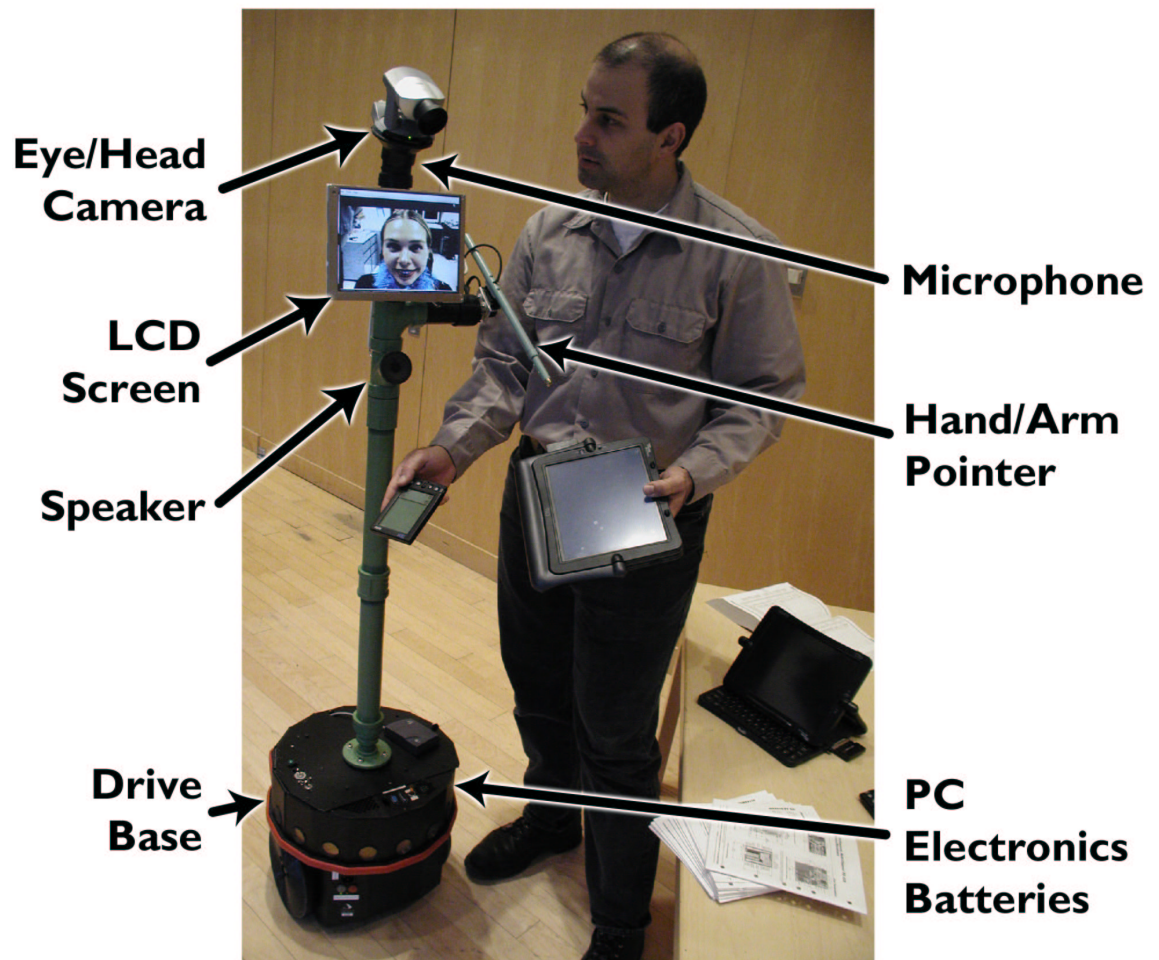
### 7.7.1 Hardware

Discouraged with the kinematic problems of the track-like bases found in all of the previous PRoPs, we designed PRoP 2 around a Nomadics Scout II base (see Figure 7.10). This base consisted of two wheels and a caster. Turning on axis was smooth and simple and the range of terrains was excellent. Floor, carpets, sidewalks, and asphalt were all accessible. The new base was also packed with 720 watts-hours of battery power and a greatly increased payload. The result was well over two hours of operational time under full use and much more if movement was not continuous. Also for the first time the range sonar hardware was accessed and used.

The wireless networking gained a tremendous boost by upgrading to an IEEE 802.11b standard 11 Mb/s ethernet radio card operating at 2.4 GHz. The entire computer science building at Berkeley was outfitted with this network, permitting for the first time, full uninterrupted roaming and PRoP access throughout the entire building. Due to the wireless coverage it was also possible to roam outdoors. Once nearby buildings are setup for 802.11, the PRoP will be able to roam from building to building uninterrupted.

A new camera was added, the Canon VC-C4 pan/tilt camera with 460 vertical by 350 horizontal pixels. This camera had more range of motion ( $\pm 90^\circ$  pan and  $-30^\circ$  to  $90^\circ$  tilt) and higher zoom ( $16\times$ ) than its predecessor, the VC-C3. It was also aesthetically more iconically “eye-like”, especially with the camera in the center of the unit. The PRoP’s Canon VC-C3 camera was offset to the side of the operational unit.

Even the embedded single board computer was upgraded to a newer 600 MHz CPU with a faster system bus, more memory, and superior display hardware. We had



**Figure 7.10:** PProP 2 and its various components as it is used by a remotely embodied user to interact with a locally embodied participant.

experienced problems with P<sub>RoP</sub> 1 rendering live video to the VGA display fast enough. The hardware upgrade in P<sub>RoP</sub> 2 fixed this problem.

### 7.7.2 Software

The plan of P<sub>RoP</sub> 1 being designed with Microsoft's C++ was more problematic than expected. The code was extremely large, messy, and required an unimaginable number of library files to operate.<sup>4</sup> As a result every single piece of the P<sub>RoP</sub> software was re-coded by hand in pure Sun Java 1.3. The operating system was also migrated to Windows 2000, which proved itself much more stable. As an added benefit we were able to install the absolute latest drivers with many optimizations compiled for much of the hardware. For example, several important video capture hardware features were only included in the Windows 2000 driver set.

Sun's pure Java had numerous other benefits. Inside of a browser we could insure that the code would be forced to run in Sun's Java Runtime Environment (JRE). That is, it would use Sun's virtual machine rather than Microsoft's. Thus, the code was removed from the low level details of the operating system and was nearly machine independent. The only problem was that as of the writing of this dissertation, there is no joystick support in Java. Therefore, we had to write a native class and have a user download this single DLL.<sup>5</sup> Support was also added for signed JAR<sup>6</sup> files. This allowed users to access a web page that would download the applet, ask permission to use the joystick DLL, and then operate

---

<sup>4</sup>This was not simply a matter of messy unmanageable coding. The author has been a proficient programmer (first coding in 1982) in many languages including C++ since 1988. Nevertheless he found the management of Microsoft C++ code unwieldy, miserable, and unpleasant to debug.

<sup>5</sup>DLL is a dynamic linked library accessed at runtime.

<sup>6</sup>JAR is a Java Archive format standard.

successfully.<sup>7</sup> A “how to” page was also assembled to help first time users access the P<sub>RoP</sub> controls. The network signals returned to the P<sub>RoP</sub> 0 style sockets rather than layering on the NetMeeting T.120 data channel.

Video and audio signaling remained in NetMeeting. However, extensive testing of the Java Media Framework 2.1 (JMF) looked quite promising for achieving a complete Java solution for the P<sub>RoP</sub>. JMF is a set of Java packages conforming to Sun’s pure Java that support various media actions. It is designed mainly for audio and video capture, transmission, and playback. Of interest to the P<sub>RoP</sub> design is full H.263 video, G.722 audio, and RTP<sup>8</sup> networking support. Although offering impressive performance, JMF was not quite ready to compete with NetMeeting. The author felt that the JMF package was only about 1–2 years from being usable with P<sub>RoP</sub>s, and thus opted to remain with NetMeeting for P<sub>RoP</sub> 2. There is also an open source implementation of H.323 [Open H323 Project, 2000] that would likely provide better performance over JMF in future implementations.<sup>9</sup>

The new wireless network was a major advantage, but used DHCP<sup>10</sup> for IP address allocation. This made locating P<sub>RoP</sub> 2 on the network a bit tricky. To solve the problem, the server software on P<sub>RoP</sub> 2 simply reported its network location to a web page at a known location after each restart. A P<sub>RoP</sub> user need only access this web page to determine if the

---

<sup>7</sup>This was necessary because under Java Sandbox Security downloaded applets executing inside of a browser do not have permission to load a library file such as the joystick DLL. By signing the applet, a user can grant “special” permission to the applet allowing it to access the DLL (*i.e.*, the joystick) and operate correctly.

<sup>8</sup>RTP is a Real Time Protocol using UDP packet transmissions and is well suited for real time streams and multicasting.

<sup>9</sup>Currently this open source H.323 implementation does not operate cleanly on Windows. That is, it does not yet support two way video except under Linux. Furthermore, it does implement the many of the useful H.323 protocols. For example, there is no h.263 video support.

<sup>10</sup>DHCP is a Dynamic Host Control Protocol used to automatically request and configure computers as they come online. One element of this is typically the use of dynamic (or changing) IP addresses after each reboot.

PRoP is online and if so follow a provided link directly to it.

### 7.7.3 Interface

As in PRoP 0 and PRoP 1, NetMeeting was used for the audio and video interface. While the NetMeeting video signal can be embedded within a browser window, it proved more usable for the user to be able to access some of the video property controls hidden when it is embedded.

The move to pure Sun Java allowed for a re-design of the user interface. By far the most attention was given by users to the video window. But the Java window was also used to display several new properties never before available. Real time sonar and odometry readings, as well as heading direction, camera pan, tilt, and zoom status were displayed. Also, the all important battery voltage on-board the PRoP was shown. A low voltage warning was indicated at the appropriate time.

Using a joystick, PRoP 2 control was similar to PRoP 1 . A few keyboard shortcuts were added to auto-position the PRoP and its various components. We had observed that while operating the PRoP users sometimes became confused about their heading. For example, a user would drive for some time, stop and use the head camera to glance around, perhaps locating something or someone of interest. Since the object of focus was in the center of their screen they would command the PRoP to move directly forward. But the camera was not pointed forward so at best the PRoP seems to move erratically and at worst run into a wall or other object. A user could use the on-screen user interface to view the camera position to re-center it, but this imposed extra precision work by the user. Therefore, a few keyboard shortcuts were added to simplify control (see Table 7.1).



**Figure 7.11:** An encounter with PRoP 2. Notice the locally embodied participant focusing his gaze cue onto the screen rather than the camera. This is a common issue with any form of video-conferencing.

Key	Action
<b>s</b>	Center camera facing forward, zoom out
<b>x</b>	Center camera facing down at base, zoom out
<b>a</b>	Zoom in (increments one zoom factor on 16× camera)
<b>A</b>	Zoom in to maximum setting
<b>z</b>	Zoom out (decrements one zoom factor on 16× camera)
<b>Z</b>	Zoom out to minimum setting
<b>!</b>	Set position 1
<b>1</b>	Goto position 1
<b>@</b>	Set position 2
<b>2</b>	Goto position 2

**Table 7.1:** PRoP 2 keyboard shortcut mappings for control signals.

From an initial inspection the keyboard shortcuts in Table 7.1 may appear arbitrary. However, they map in a fairly reasonable way onto a standard QWERTY-keyboard. The “s” and “x” keys are on top of one another, one for forward center, the other below it for down center. Likewise, the zooming “a” and “z” keys are atop one another for a, hopefully, logical mapping on in and out.

The interesting new addition is the set/goto position keyboard shortcuts. We observed a need for this setting during sessions with the PRoP at meetings. The PRoP would operate as expected as a remote user maneuvered to a meeting location and situated herself in the room. At that point there was almost no need to move the PRoP (*i.e.*, in some sense the PRoP was now “seated”). As the meeting progressed it was easy to keep the camera focused on the individual running the meeting. This worked fine when there was a single speaker. However, the second most common form of events at these meetings was dialog between two (or less frequently, more) attendees. For this, PRoP control became extraordinarily cumbersome. Panning and tilting the camera between the two individuals





**Figure 7.12:** From a remote location a primary user interacts and discusses the contents of a poster with a secondary user though PRoP2 .

speaking became laborious. This problem, prompted us to add the ability to easily set two (at least for now) pan/tilt/zoom position and the ability to switch between them quickly. All this did was allow a user to lock (*i.e.*, save) the current position for easy recall later. Again the layout of the keyboard mapping was for two keys directly beside one another. The hope was to have the mapping be for a left and right individual involved in a discussion. More pre-stored settings may be useful but have not been implemented.

Overall, PRoP2 was among the easiest PRoPs to use. Locally embodied users interacted well with these systems (see Figures 7.12 and 7.11).

## 7.8 Compression: Making It Fit

Real-time audio and especially video data are too large to send on current networking hardware. To make transmission possible, the signal must be compressed. The question



of choosing the proper compression strategy is complicated. There are numerous techniques and it is not the goal of this dissertation to enumerate all of the various strategies. Instead we highlight a few basic tradeoffs.

### 7.8.1 H.323

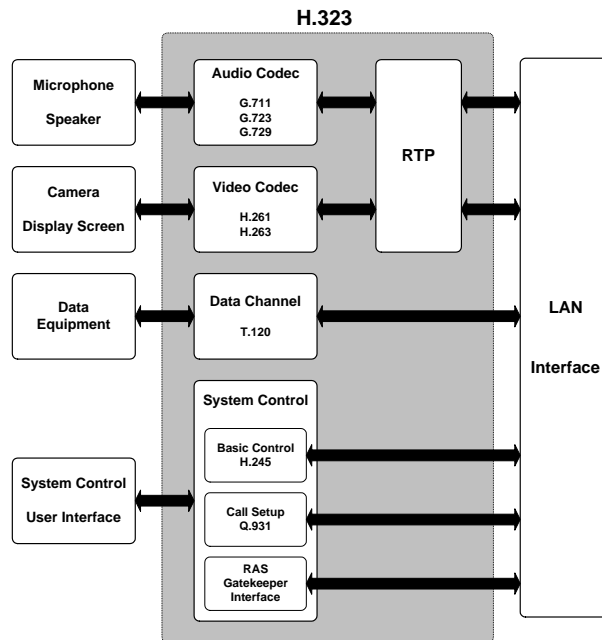
The International Telecommunication Union (ITU) approved one of the most comprehensive and complete standards currently in use. This standard defines how audiovisual conferencing data is to be transmitted across various networks. For example, H.320 is designed to operate on digital ISDN<sup>11</sup> networks, H.324 on POTS<sup>12</sup> lines, and H.323 on non-guaranteed bandwidth packet-switched networks such as the internet.

H.323 is a perfect solution for networks that cannot guarantee quality of service such as the internet. The H.323 standard is actually an umbrella name for a suite of protocols for audio, video, data, and other multimedia. By complying with H.323, multimedia products and applications from multiple vendors can inter-operate, allowing users to communicate without concern for hardware compatibility. Currently, H.323-compliant platforms exist in many sizes and shapes, including video-enabled personal computers, dedicated platforms, IP-enabled telephone handsets, cable TV set-top boxes, and turnkey boxes. This heavily influenced our adoption of H.323 for PRoPs. H.323 also supports bandwidth management and multicast for multi-point conferences. Figure 7.13 shows a block diagram of a typical H.323 system.

---

<sup>11</sup>Integrated Services Digital Network (ISDN) is a network of dedicated bandwidth of either 64 kb/s or 128 kb/s.

<sup>12</sup>Plan Old Telephone Service (*i.e.*, standard analog telephone service used in most homes capable of supporting up to 52 kb/s).



**Figure 7.13:** Block diagram of the H.323 System.

### 7.8.2 H.261 versus H.263

H.323 incorporates two required video compression standards, H.261 and H.263. Both H.261 and H.263 compress the video data by only transmitting a complete image “every so often”. These complete images are denoted as Intra-Frames or I-Frames. In between I-Frames, the algorithm encodes only the differences since the last frame. These difference frames are called Predictive-Frames or P-Frames. Because they only contain difference information, they are an order of magnitude smaller in size than I-Frames. The mechanism by which this differencing is done is beyond the scope of this dissertation. However, several important design issues are noted in the following paragraphs.

First, only data from the past is used to create P-Frames. This is in contrast to

MPEG<sup>13</sup>, another common video compression technique, where data from both the past and the future are used to more optimally create the differencing frames.

The problem with H.261 is that the method it employs to detect differences from frame to frame makes it vulnerable to errors when there is high motion content in a video stream. As a result, video with higher motion (such as a rapidly sweeping hand, turn of the head, or more than one object moving in the picture at the same time) can result in artifacts in the video such as unexpected blocks of color, haphazard lines, or other video “noise” in the displayed image. H.261 is designed to work best with “talking-head” video streams where there is little motion, except perhaps a single head. Motion from a PRoP video feed typically causes extremely poor H.261 performance and quality.

H.263 on the other hand incorporates a more sophisticated motion vector algorithm with a larger search space. First, it allows for future predictive frames like MPEG, making its motion performance better than H.261. Secondly, it accommodates fractional pixels in its algorithm, increasing the video and motion quality over H.261. Finally, it provides many configurable options to allow it to perform well at bit rates that can be lower or higher than H.261 can support.

### 7.8.3 Latency

One final concept related to video streaming is that of latency. This is the amount of time between when an event occurs at one end of a video stream and the time that this event is viewed at the opposite end of the stream. For PRoPs this can be measured as the time elapsed from when an PRoP user says hello until it is heard out the speaker onboard

---

<sup>13</sup>Moving Pictures Expert Group standard.

the P<sub>RoP</sub>.

True real-time response means a latency that is low enough so as to not be perceptible to human beings. This is generally considered to be in the range of 150–200 ms [Bailey, 1989]. Somewhat higher delays than this may be acceptable before they become disruptive to real-time conferencing. Practical systems can have a 400 ms latency and still provide good conference interaction. It is difficult with current P<sub>RoP</sub> hardware and wireless networking to achieve these optimal ranges.

Low latency is directly related to the necessity for high-bandwidth video streaming. All codecs must accumulate some amount of video before they can encode it. Until this minimum amount is accumulated, the codecs cannot encode anything and cannot transmit anything. This means that low-bit rate transmission has the unexpected and very undesirable property of increasing the transmission latency to a very noticeable level.

For example, a typical codec must collect at least 16,384 bytes of video before it can encode and send the first bit of data. At a 384 kb/s rate, about 40 ms will elapse before the codec has accumulated the first 16,384 bytes. Only then can the codec begin to compresses the data, which compacts it for transmission. Even at a nominal 8/1 compression, this means that only 2048 bytes are actually transmitted.

At the other end of the conference is a codec with the requirement that 16,384 bytes of data must come in before it can begin to decode the data and display the video. Using the example from the previous paragraph, at least 320 ms will elapse before the receiving end sees the remotely transmitted signal on their display.

#### 7.8.4 PRoP Compression

The two-way audio/video software used for PRoPs was written to interface directly into H.323 video conferencing standards. Currently, several hardware and software implementations of H.323 exist with much of the code in open source. H.323 supports a suite of protocols for low latency real time audio, video, and data transmission, ideal for use with PRoPs. The current solution uses NetMeeting, a freely available piece of software from Microsoft. The recent Java Media Framework (JMF) also contains a promising implementation of the main H.323 protocols: H.263 (video) and G.729 (audio) as well as a Real Time Protocol network stack (RTP). A migration to JMF would provide a complete PRoP solution in Java. Also, as mentioned in section 7.7.2, there is also an open source implementation<sup>14</sup> of H.323 that would likely provide better performance over JMF in future implementations once a full set of H.323 protocols are supported.

### 7.9 Designing Tele-embodiment: PRoP Architecture

PRoP 0, PRoP 1, and PRoP 2 were all designed from simple mobile robot bases 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 (see Figure 7.10 and Figure 7.14). Onboard all of these PRoPs were a color video camera, microphone, speaker, color LCD screen, a few simple custom electronics, and various drive and servo-motors.

---

<sup>14</sup>As mentioned in Section 7.7.2, this implementation is incomplete.

### 7.9.1 Hardware

The current generation P<sub>RoP</sub>s have undergone numerous major hardware changes during their evolution (see sections 7.3–7.7). They are constructed from a commercially available robot base with an onboard PC/104 format single board embedded computer running a Windows operating system. These untethered devices carry several hot-swappable lead-acid batteries to support complete operation of the computational systems, motor drive, and custom electrical hardware. Kinematically the P<sub>RoP</sub> drive base allows it to maneuver around and rotate in place. Eight or more front facing sonar sensors measure distance to nearby objects. Numerous custom fabricated circuits and in-house C and assembly language programming of single chip microcontrollers provide the interface from the computer to the various motors, sensors, and control hardware.

The computer hardware is typically an MMX enabled Pentium III processor running at 600MHz. Additional hardware supports various audio/video capabilities such as full-duplex audio and live video capture. The audio and especially video compression are computationally intensive operations, necessitating the higher end CPU selection. See section 7.8 for more discussion on the issues surrounding these computations. A speaker and an omni-directional surveillance microphone are also included. Ethernet communication uses a 2.4 GHz Direct Sequence Spread Spectrum (DSSS) wireless interface card fully supporting IEEE 802.11b and wireless ethernet at 11 Mbit/s.



**Figure 7.14:** Using a PRoP to collaborate in resolving the details of an aircraft jet engine problem (envisionment).

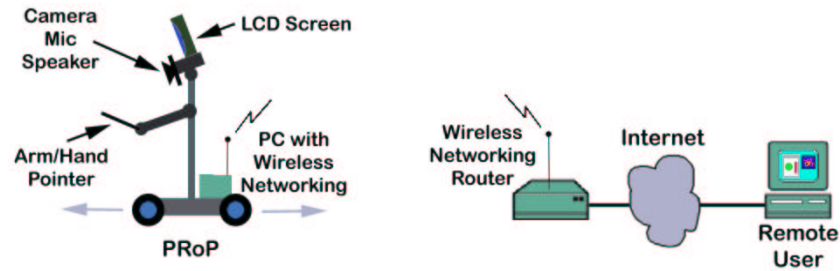
### 7.9.2 Physical Body

The body is a low weight 1.5 meter ABS plastic pole that supports the attachment of various hardware and cabling. A speaker and microphone are located on the stalk of the body. Crowning the pole is a high quality pan/tilt camera that also supports software controlled 16× zoom, auto iris, and auto focus. Located directly below this camera is a 30cm (12 inch) diagonal flat panel color display. Off to one side of this is the “arm/hand” hardware where a two-degree of freedom (2DOF) pointer is attached for simple gesturing. At the tip of this “limb” a small laser pointer is attached.

#### Power Requirements

Operation of all of the computational hardware requires between 20 and 25 watts depending upon various factors such as network load and video image compression level. This includes powering the computer, VGA flatscreen, camera, networking hardware, and hand/arm controller. The power to the drive motors is dependent upon the base type and loading. For example, PRoP 1 requires substantially larger current to execute an on-axis rotation on a carpeted surface than PRoP 2 on a smooth floor surface. As a result of varying operational actions, there is a wide range operational life from the PRoP. For example, PRoP 2 is able to remain online without moving around for almost eight (8) hours. However, under continuous movement at maximum motor loading it remains operational for only about 2.5 hours.





**Figure 7.15:** System overview of the basic PRoP hardware configuration

### 7.9.3 Software

The software consists of two distinct pieces: (1) PRoP interface and control elements and (2) tele-conferencing tools.

#### Interface and Control

A user connects to the PRoP using a standard web browser that invokes a Java applet. This applet reads the various input devices (keyboard, mouse, and joystick) and transmits the data to the PRoP. A Java based server runs on-board the PRoP, serving as the glue between the various PRoP hardware elements (*i.e.*, the camera, base, hand/arm, etc) and the remote networked user. The user's Java applet also receives back various status information from the PRoP and displays it to the user. Currently the network communication is implemented using TCP/IP sockets. An overview of this communication architecture is depicted in Figure 7.15. Complete discussion of the user interface can be found in Chapter 8.

## **Tele-conferencing Tools**

The two-way audio/video software is written to interface directly into H.323 video conferencing standards. H.323 supports a suite of protocols for low latency real time audio, video, and data transmission, ideal for use with PRoPs. The current solution uses Microsoft NetMeeting. Further discussion on compression techniques can be found in section 7.8.

## **7.10 Elements of Tele-Embodiment**

The various PRoP design choices have been guided largely through trial and error experiments with the actual devices (see sections 7.3–7.7). Obvious methods of communication such as audio and video have always been part of our original design. However, after extended use of these PRoPs, it was clear that many important, and often subtle, elements were missing from the experience.

Through this evolutionary development we have been able to identify several behavioral traits which we consider essential to providing the most compelling overall experience for both the remote and local PRoP user. In the following subsections we identify these elements and discuss their role in creating convincing personal tele-embodiment. While some of these traits may seem obvious, there are others that we have found to be surprisingly important and far less evident.

### **7.10.1 Two-way Audio**

Perhaps the most apparent element of communication, two-way audio allows users to engage in remote conversations. Audio is also the channel whose usefulness is most

susceptible to quality degradation from reduced network bandwidth and/or network packet loss. The compression techniques used, such as H.323, provide rate adaption techniques that allow the video signal to degrade before audio. This enables the tele-experience to remain partially intact even under adverse networking conditions.

The importance of background “noise” near the P<sub>RoP</sub> was an unexpected benefit to the immersive tele-experience gained from two-way audio. The experience of using the P<sub>RoP</sub> was noticeably more compelling when the remotely embodied user was 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.* There was not a formal study with the P<sub>RoP</sub> to address this issue of audio and its direct relation to user experience. However, much of the data and design decisions from this section were gained from experience and comments from a small group of users. For example, we noticed that when less background sound was heard, remotely embodied (and locally embodied) users would say things such as, “Are you still there?” or “Can you still hear me?” with a much greater frequency than after we improved the sound levels.

The conferencing software on-board the P<sub>RoP</sub> is designed to provide audio thresholding. That is, when the audio activity (*i.e.*, signal amplitude) is below a certain level, the audio signal cuts out. This reduces bandwidth but creates a walkie-talkie-like transmission. A P<sub>RoP</sub> primary and secondary user would hear something and then have the audio click to silence. This effect is not consistent with real life where a myriad of sounds constantly surround us, as so eloquently expressed in John Cage’s *4’33”* [Cage, 1952]. Although the

connection was not half-duplex, the experience was. Adjustments were made to the conferencing software to remove the audio thresholding and provide continuous transmission of remote sounds.

### 7.10.2 Two-way Video

The video screen allows for exchange of visual cues such as facial expressions while the on-board camera serves as a general purpose tool for viewing a plethora of visual data about the remote space (*i.e.*, Who's there? What does it look like? Is the person I'm talking to even looking at me?, *etc*). This enabling of facial expressions in human communication is vitally important, as pointed out by numerous researchers [Eckman and Friesen, 1978, Schiano et al., 2000, Eckman et al., 1972].

The video signal, like the audio channel, undergoes a lossy<sup>15</sup> compression. Sophisticated video compression algorithms are essential to making such video signals usable over networks unable to support QoS guarantees.<sup>16</sup>

Video is also the most demanding tele-embodiment trait in terms of network bandwidth consumption and processor usage (see section 7.8). The effects and tradeoffs of video and image quality and its resulting perception by humans has been extensively studied by several researchers [Reeves and Nass, 1996, Bailey, 1989]. Surprisingly, in many cases the overall quality of the resulting video is far less important than the ability of that video to

---

<sup>15</sup>Lossy is a term used to describe a type of data compression in which information is lost during the compression process. Data that undergoes a lossy compression is **not** guaranteed to contain the same data as the original data. While these forms of compression may appear useless, they are not and in fact are used commonly. JPEG image compression is one such commonly used lossy compression technique.

<sup>16</sup>Quality of Service networks are able to guarantee bandwidth and latency bounds for data. For example, when you lift a standard telephone receiver you are allocated a full 64 kb/s channel whether it is in use or not. This insures a quality of service for the duration of the telephone call.

provide subtle information about the motions, actions, and changes at the remote location. Reeves and Nass demonstrated that humans have little difficulty interpreting themes and specifics within video imagery despite substantially degraded fidelity in the overall signal.

There are also lessons and tradeoffs of frame rate verse video quality. When a primary P<sub>RoP</sub> user is navigating or when significant amounts of activity are occurring within the camera field of view, the importance of high video frame rates dominates over the resulting video quality. We learned that during times of “high video activity” the resulting compressed video signal should convey to the remote user at least an approximate representation of the remote motion and/or activity. However, during periods of small temporal video activity such as when the user is conversing with an individual, examining an object, or reading a sign, it is clearly the overall quality of the video signal that dominates over frame rate. The conferencing software used onboard the P<sub>RoP</sub> provides for these quality verse frame rate trade-offs. The conclusion is that for a wide range of tasks, humans seem capable of transparently adjusting to this lower quality video signal. We observed similar results in our own user testing (see Chapter 9).

An example of the quality verse frame rate tradeoff can be found in an actual task we observed users perform. On several occasions P<sub>RoP</sub> drivers lost their way in a familiar hallway. They performed a two-step process to orient themselves. First, using the high-frame-rate low-quality video, users maneuvered the P<sub>RoP</sub> towards a door name-tag or room number sign. Accurately positioning the P<sub>RoP</sub> in front of such a room marking is prohibitively difficult without the use of high-frame-rate video to provide visual feedback to the user. Second, when motion stopped, the user requested high-quality low-frame-rate

video which is used to easily resolve the name on the door (or room number), thus identifying their location in the building. The desired video tradeoff setting can often be inferred from various sensing. For example, when the PRoP is in motion, frame-rate should dominate.

In addition we noticed that with only one-way video, PRoPs were mistaken as tele-operated surveillance tools or autonomous reconnaissance drones. Both of these tasks are far from the intended application of PRoPs. This video-asymmetry was eliminated when a LCD screen with a video feed from the remote user was mounted on-board the PRoP at head height. Individuals near the PRoP could then easily understand that not only was the PRoP controlled remotely by a real person but more importantly they could identify the person much as they would in a real face-to-face encounter.

The two-way video is also an appropriate mechanism for transmitting various non-verbal cues from the remote PRoP pilot such as facial gestures and expressions. When bandwidth limitations exist, the screen can still be used to display a still image of the remote user. This still image, although less than ideal, succeeds in conveying at least the identity and existence of the remote user.

Finally, two-way video enables mutual observability. That is, both individuals (PRoP driver and remote individuals) can see one another. Without this ability, an awkward sense of surveillance arises. For example, Space Browsers lacked the ability of two-way video. As a result several individuals mistakenly remarked on it as a sophisticated security surveillance system.

### 7.10.3 Mobility

Mobility, and in fact all of the remaining behavioral traits, are notable PRoP enhancements to standard video conferencing. So 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 through a relatively simple joystick interface. It also provided enough freedom for users to maneuver within (and outside of) buildings, traverse stairways, and pass through doorways.

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 handle what we call dextrous human motions. In particular PRoPs do not need 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 during waking hours. 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 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.

The most recent PRoPs satisfy many of these goals. However, stairs and curbs are yet to be surmounted. Tradeoffs in other portions of PRoP design were made rather than solve the stair/curb navigation problem. In fact, curb and more obviously stair navigation

will likely require new user interfaces. The solution is more than just mechanical.

### **Proxemics**

Mobility allows users to position themselves with respect to a group or individual. As well as providing browsing and exploring capabilities to the P<sub>RoP</sub>, this also permits first order proxemics to be expressed [Hall, 1966]. Proxemics is the study of the cultural, behavioral, and sociological aspects of spatial distances between individuals. An example is the distances that people keep between them when performing particular tasks. The spaces can loosely be defined as intimate space, personal space, social space, and public space. The type of interactions and communications that take place in each region are largely cultural. However, it is important that a person can situate themselves within the proper space when interacting.

With P<sub>RoPs</sub> it's clear that there is some form of social proxemics occurring. That is, a primary user can position the P<sub>RoP</sub> far or near an individual to create additional tension or seriousness. It's also possible for the remotely embodied user to present the appropriate elements of a threat as in the real world to another locally embodied person through P<sub>RoP</sub> positioning. However, two questions need to be answered. First, can a P<sub>RoP</sub> user feel confident in positioning the P<sub>RoP</sub> at the desired social distance for the particular type of interaction intended. Secondly, the proxemic mapping through a P<sub>RoP</sub> is far from ideal. What is gained and lost in this important proxemic cue with P<sub>RoPs</sub>? This dissertation does not undertake this challenge but highlights an awareness of it. We have attempted to insure that the P<sub>RoP</sub> camera lensing and placement provide a reasonable view of the remote space, allowing a P<sub>RoP</sub> driver to judge distance and space reasonably.



#### 7.10.4 Directed Gaze

A small movable “head” (*i.e.*, a camera on a controllable pan-tilt platform) mounted atop the PRoP facilitates the ability to quickly glance around a room to get a feel for its size, occupants, *etc.* This device also delivers a gaze direction non-verbal cue to remote individuals. The device is similar to the GestureCam [Kuzuoka et al., 1994], which allows a remote participant in a conversation to have direct control of her visual field of view. This relatively simple PRoP “head” provides a vitally important element of human communication; namely, the direction of attention or gaze.

The gaze cue 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. This aids in back-channelling used in human communication.

#### 7.10.5 Deictic Gesturing

We learned quickly that gestures are very important for human communication [Goodwin, 1986b, Goodwin, 1986a, Goodwin, 1981]. Remote users immediately found the need to point out a person, object, or direction to the individual near the PRoP. Although the movable head could be used as a crude substitute, it lacked the correct visual gestural cues of pointing and was often ambiguous to individuals watching the PRoP.

A simple 2 DOF pointer attached near the location of an “arm/hand” portion of the PRoP facilitates these simple pointing gestures. These motion patterns allow 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. To preserve meaning, nuance, and richness in these gestures, continuous input devices (*i.e.*, joystick, mouse, and tablet) were used.

We found that 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. We have also been exploring several optional “arm/hand” designs to accomplish basic gesturing functions. More complex gesture interfaces and mechanisms are beyond the scope of this dissertation.

One simple gesturing system we built consisted of a laser pointer mounted on the “hand/arm”. The laser control is simply an on/off switch at the client applet. This device allows the remote user to point at the object which is the focus of attention. It also allows a limited form of “gesturing” of the laser spot, by drawing on a nearby wall. This kind of gesturing is frequently used by humans giving lectures with laser pointers in darkened rooms. The speaker’s actual hand gestures are invisible, but many speakers improvise gestures such as circling an important feature, or zig-zagging the spot, or indicating slow rises or falls. Because much hand gesture is encoded in motion rather than dexterity, those gestures survive and can even be interpreted quite well when manifested instead by a tiny red spot.

There has also been a significant amount of research into gesture recognition [Long et al., 2000, Rubine, 1991, Starner and Pentland, 1995]. These systems typically aim to identify a human motion, typically made with a mouse, and interpret it as a known ges-

ture. 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.

### **7.10.6 Reflexivity**

The ability of a user to experience their own existence and actions through the PRoP turns out to be an extremely important element in providing a compelling tele-visit.

When users could point the camera downward and actually see the wheels and base of the PRoP there was a noticeable improvement in the quality of the immersive tele-experience. Likewise, the experience was enhanced when users could steer the wheels, move forward and backwards, or position the pointer while visually watching the actions that resulted from their commands. Imagine if you were able to give commands to your arms and legs but never sense the result? Clearly, the experience would be lacking a significant element, which we call *reflexivity*.

There is also reflexivity in the audio channel. This form of reflexivity is actually a major hinderance. The ability to hear one's own voice when speaking is a useful mechanism to regulate the tone or volume. However, since we are dealing with audio compression, lossy wireless networking, and limited computational resources, the audio is reflected back with a substantial delay. This is an intrinsic property of all video-conferencing applications. The best solution is an investment into echo cancellation hardware microphones and software systems.

These limited resources can wreak greater havoc on the immersive experience. This occurs when network delays cause the user to feel detached from the PRoP. The lag between moving the control joystick or mouse and seeing the results can sometimes be several seconds. By then the PRoP may have wandered far from where the user intended. The impression that the user gets is that the PRoP has "a mind of its own" which is exactly the opposite of an immersive experience. Expected technological advances in computational power, networking bandwidth, and latency will eventually solve several of these problems. Promising tests of PRoP 1 and PRoP 2 have demonstrated remarkably high quality control

with user thousands of kilometers away or even in other countries (see Chapter 9). However, substantial transmission delays are going to be a fact of life on the internet for at least a few more years. That means that real-time control of PRoPs over the internet will continue to be cumbersome.

### 7.10.7 Physical Appearance and Viewpoint

Employed as a communication tool in public settings, subtle physical design choices often hinder the effectiveness of PRoPs. This is especially true for the secondary users cohabitating with the PRoP. Careful attention to its overall height, shape, and color are important.

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.

#### Height

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. It does not suffice to simply have a camera someplace in the room where the PRoP is currently located.

We learned early on that PRoP 0 was too high as it caused intimidation as it gazed down on others.<sup>17</sup> However, one early design of PRoP 1 was too low, requiring users to awkwardly bend down or squat to communicate with it. What is the correct height? People come in a variety of shapes and sizes. The answer depends heavily on the context, task, individual controlling the PRoP, and the people with which the PRoP interacts.

## Color

Height is only one of several of these traits. We have also experimented with color choices. Color has enormous cultural, gender, and religious connotations. A color choice may imply dominance or passivity (among other things). Hue and saturation play an equally important role. For example, highly saturated colors are often used in children's toy design and may not be an appropriate choice for a serious meeting with investors. Similarly, one may want to appear relaxed and playful with a different color choice for a party or casual social event. Appearance often matters in life, and there's no reason to expect PRoPs would be immune to this social affect. Chapter 10 provides some further discussion on this issue.

There has been much trial and error as well as anecdotal evidence for and against many of the elements in the current design. For example, we have experimented with several different color choices. PRoP 1 was painted with Rust-Oleum brand Painter's Touch Colonial

---

<sup>17</sup>It's worth noting that in some situations an individual may desire to cause intimidation and thus choose a PRoP height and style accordingly. For these "special ops" physical appearance is even that much more vital.

Red Gloss #1964 and PRoP 2 with Rust-Oleum American Accents Leafy Green Satin #7934. Admittedly, color choice is an entire research field and we have been unable to perform extensive studies into the choices made for PRoP color. The choices were not arbitrary. Reasonable attempts were made to choose colors that were gender neutral, highly visible, yet not overly lustrous or associated with cautionary messages (*i.e.*, bright Caltrans<sup>18</sup> orange or glossy yellow). The goal was to make the PRoP easily stand out from most backgrounds without tending towards gaudy or obnoxious colors.

## Shape

The current PRoP shape is largely dictated by ease of design and modularity (*i.e.*, interchangeable parts). This has led to simple ABS, aluminum, and PVS piping and off the shelf hardware. It is clear that even a small amount of design work into the shape of the PRoP would be extremely beneficial. However, we are engineers and scientists, not designers. Thus the necessary PRoP design challenges exist but are outside of the scope of this dissertation (see Chapter 11).

### 7.10.8 Browsing and Exploring

When designing PRoPs we found that even with all of the previously discussed traits, it is essential to allow a remote user to wander, explore, and travel throughout buildings and spaces in much the same manner as humans normally do. It is this higher level browsing and exploring behavior that is perhaps the most important element of tele-embodiment.

---

<sup>18</sup>California Department of Transportation which erects bright orange equipment on California roadways while working.

PRoPs should allow users to search for an individual or a particular location such as a laboratory, office, or conference room. They should support exploration and wandering, where the user has no specific target. This behavior is intended to mimic the action of walking around a location noting the names and functions of rooms, wandering around looking for people they want to visit, or checking out the progress of experiments in a laboratory.

More importantly, browsing and exploring are autonomous operations performed by a remote user and do not require any support from individuals in the remote space, beyond the availability of the PRoP itself. The benefit of this is that PRoPs can be installed and used in a location with little overhead and disruption to the inhabitants of the remote space. A remote PRoP user can also be given a tour of the remote space by one of its local occupants. In fact either or both the visitor and guide may be PRoPs. Overall, a wide gamut of human activities can be performed without any local assistance such as attending meetings, seminars, conferences, discussions, *etc.*

### **7.10.9 Hanging Out**

A surprising but important social function is the ability to simply hang out. We know that physical presence often improves the length, quality, and priority of one-on-one and group interactions. This is a purely social phenomenon. In many work situations, individuals are willing to talk to you for a much longer period in person than they would be willing to do over the phone. When phone and in-person communications collide, most people try to terminate the phone call so they can return to their live interlocutor. We would like to better understand the factors that influence this preference and see where



a PRoP presence fits into the priority ordering. Tests planned beyond the scope of this dissertation will likely provide some insight into this phenomenon.

A visit to a remote place is an extended activity in which a person shares space with others. During this time, the task being performed by each person changes, and there may or may not be tasks in common at any given moment. These tasks serve as additional stimuli for communication between the individuals, leading to multiple communication episodes with different subjects.

## 7.11 Discussion

PRoPs, personal tele-embodiment, and the ideas that surround this research are filled with fresh new topics. Inevitably the work raises 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 crowded arena of techno-jargon. However, it seems productive to use a common term to distinguish the growing research in the area of traditional CMC and telepresence. Obviously, methods of achieving telepresence are not new, nor are systems that allow tele-communication. Similarly, techniques of human communication have been studied 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 work represents our best attempt to convey this direction of

personal tele-embodiment research.

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

While standard 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 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 realized the importance of immersing the P<sub>RoP</sub> 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, P<sub>RoPs</sub> attempt to

achieve certain fundamental human skills *without* a truly anthropomorphic 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 *exactly* re-create 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 a 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.

With PRoPs in so many forms it quickly became a problem deciding on useful methods for controlling all of the systems onboard these systems. In the next chapter we discuss specific challenges involved in designing interfaces for these systems.

## CHAPTER 8

# CONTROL, NAVIGATION, INTERFACE

*Everything moves continuously. Immobility does not exist. Don't be subject to the influence of out-of-date concepts. Forget hours, seconds, and minutes. Accept instability. Live in Time. Be Static – with movement. For a static of the present moment. Resist the anxious wish to fix the instantaneous, to kill that which is living.*  
– Jean Tinguely

The wide range of expressive degrees of freedom requires considerable attention to the interface design of PRoPs. Failure to support easy user access to these features will cause them to be disregarded by the user and hinder remote user interactions [Laurel, 1990, Laurel, 1993]. There has also been recent work into how to initiate and/or transition between various online telepresences, virtual reality, and the real world [Koleva et al., 2000]. Several control interfaces and navigational tools designed for PRoPs are described in the following sections. Several of the more elaborate control schemes are beyond the scope of this dissertation but are described here and further in Chapter 11.

## 8.1 Let's Get Moving

Basic control of the PRoPs is exacerbated by a traditional design challenge: map a low DOF input device into a high DOF output system. For example, a basic PRoP requires at least seven control signals: velocity, steering, head pan, head tilt, pointer pan, pointer tilt and camera zoom. But the basic input device is a 2DOF mouse. This is not to say that high DOF systems are impossible to control. Almost everyone knows at least one computer gamer who can dexterously control the most complicated spaceship or action hero (or villain) with only a mouse and a few raps on the keyboard. Basic control of PRoPs originated with simple mouse control with keyboard or mouse buttons used for mode selection. This was considerably easier with the previously designed tele-operated blimps where one would rarely need to modify the height and there were no steerable cameras or pointers on-board.

## 8.2 Joystick

As described in Sections 7.3–7.7, PRoP control evolved quickly to incorporate a joystick with buttons used for mode selection. Previously difficult tasks such as moving straight ahead or swinging the head back and forth were dramatically improved. The joystick offered vitally needed tactile feedback about the position of the control signal input with respect to the range of the device (*i.e.*, where the center position was). With the mouse, a user could only determine this visually by looking at the screen that displayed the control signal output.

### 8.3 Keyboard Shortcuts

Currently, a user drives the PRoP using the standard joystick left-right and forward-back controls. The “hat” or point-of-view (POV) switch found on more modern “gamers” joysticks directs the head pan-tilt motion. Zooming is controlled through the joystick’s throttle or Z-axis controller. There are also several keyboard shortcuts. The *A* and *Z* keys are used to step the zoom in and out, respectively. A complete description of these shortcuts and interface tools can be found in section 7.7.3.

### 8.4 Sonar Feedback

There has also been some initial work into incorporating the sonar data on-board the PRoP into the navigation system. Such a system is capable of providing simple collision avoidance and wall following operations. All of the sonars are currently mounted approximately 10 cm off the ground. Therefore, they are unable to detect collisions of the body or arm with objects of the ground. For example, a table or desk is not correctly detected. The addition of more sonars or some force sensing resistors on the body are possible solutions.

There are drawbacks to these ultrasonic transducers<sup>1</sup> or sonars, curtailing their continuous use. Sound that is generated above the level of human hearing range is called ultrasound. Although ultrasound typically starts at 20 KHz, these sonars operate close enough to the human range of hearing to be quite loud and often disruptive to conversation. When 16 sonars are all in operation firing twice a second the audio distraction is certainly not negligible. There is also future discussion of the use of sonar feedback in Chapter 11.

---

<sup>1</sup>An ultrasonic transducer is a device that is capable of generating and receiving ultrasonic vibrations.





## CHAPTER 9

# EXPERIMENTS AND EVALUATIONS

*The experiment requires that you continue.*

*– Prod 2 to subject in Obedience to Authority by Stanley Milgram*

PRoPs are novel technologies. The social conventions surrounding them are unknown and need to be studied. For example: How will they be used? When are they appropriate? Where do they fit into the social priority ordering (*i.e.* will a person interrupt phone call to speak with a PRoP enabled user)? Can a PRoP user be more persuasive than a user with video-conferencing alone? How does PRoP size, location, weaponry, and positioning affect its influence and persuasion ability [Milgram, 1974]? How about compared to face-to-face (F2F)? Many of these questions will remain open for some time, as PRoPs have only recently entered the CMC landscape. However, this dissertation does undertake several initial experiments to investigate several basic issues: (1) usability and acceptance (see section 9.1), (2) network effects (see section 9.2), (3) airport traveler’s survey (see section 9.3), (4) student group study (see section 9.4), and (5) tele-lecture (see section 9.5).

## 9.1 Experiment 1: Usability and Acceptance

Our first usability experiments attempted to assess the efficiency of the interface and control mechanisms for P<sub>RoP</sub>s as well as their appearance and acceptance. The experiment consisted of a brief explanation of the P<sub>RoP</sub> controls and abilities. The user was then instructed to explore a building they are familiar with, locate a set of landmarks, and converse with another individual.

### 9.1.1 Users

The experiment involved seven individuals (4 female / 3 male) ranging in age from 26 to 55. Employees and visitors to a public city library were given the opportunity to sign up to participate in a 30 minutes session sometime within a one week period. Involvement was not in any way linked to their employment or library privileges. Individuals were informed that they would be asked to test and evaluate the interface and usability of a remote controlled robotic system. No payment was offered nor was the experiment conducted during their work hours. Computer familiarity varied from an experienced network administrator to a user who had spend almost no time on the internet. Only one user had any familiarity with internet based video-conferencing while two had significant experience with chat rooms and instant messaging systems. There was one avid gamer in the group. A more detailed analysis of the profiles of the users is given in Table 9.1. Although gathering data from hundreds of users would be nice, we found a wealth of useful feedback from even the small group we did examine. Restricted and limited access to the test facility was another hindrance to having a large test group.

User	Age	Sex	Computer Experience	Video Conf Experience	Comp. Gaming Usage
1	46	F	Intermediate	None	None
2	55	M	Intermediate	Some	Some
3	26	F	Advanced	None	None
4	43	F	Beginner	None	None
5	42	M	Expert	None	None
6	29	M	Intermediate	None	Extensive
7	38	F	Intermediate	None	None

**Table 9.1:** Profile of participants in Experiment 1: Usability and Acceptance. Levels used for computer experience are Beginner, Intermediate, Advanced, and Expert

### 9.1.2 Setup

The tests were held in a public building, a city public library, with no affiliation to a university, company, or research laboratory (see Figure 9.1). In particular the building contained no “high-tech” facilities or technology related items that would distract the users. PRoP 1 was used in these experiments. PRoP 1 was setup in a large 220 m<sup>2</sup> (2500 sq ft) room with hard floors. This room was in the basement of the public library and was typically used for classes, public meetings, and workshops. The room contained several large desks, chairs, a piano, and a blackboard. The PRoP stood turned on ready at one end of the room. A separate smaller room behind a closed door contained a standard computer (500MHz Pentium-III with 512 MB running Windows 2000) with joystick, camera, microphone, and headphones. This computer was running on a private network with no other network traffic. A Proxim wireless bridge base station established a 2Mb/s 802.11 link to the PRoP. The sonars and sonar visual feedback were disabled during this test leaving the users with only the most basic PRoP controls. This was done because we were interested in basic usability



**Figure 9.1:** Initial setup of PRoP 1 for experiment 1 in a large room in the basement of a public library. This room is typically used for classes, public meetings, and workshops

without additional software or online visual aids as provided by sonar. One reason is that the sonar, although typically helpful, can cause confusion with too much data or invalid readings. It was also felt that the additional time to learn the sonar interface and/or its usage would overrun the 30 minute limit we were enforcing for each individual.

### 9.1.3 Method

Individually, each user was brought into the room and shown the PRoP. They were advised that they would be using a new research project and be asked to evaluate it for a 30-minute period. They were also informed that they could pause or discontinue the experiment at any time and for any reason with no questions asked. The audio, video, pan-tilt-zoom camera, and mobility features were briefly explained. They were then escorted



**Figure 9.2:** User number 2 during Experiment 1. This individual is maneuvering the PRoP to complete the first task.

into the smaller room behind a closed door and shown the typical computer setup with A/V hardware (see Figure 9.2). Video from the PRoP camera was already on the user's screen. The experimenter explained and briefly demonstrated the joystick and keyboard controls for using the various features of the PRoP and then left the subject alone in the room.

The user was asked to perform two tasks. First, to maneuver the PRoP over to a table where some books were placed by another individual and to read the text on them. A confederate sat on the opposite side of the table with the books and interacted with the PRoP once it arrived at the table. The confederate communicated to the remote user via the PRoP. This task required basic navigation skills around several obstacles and use of the head pan-tilt-zoom controls. This test focused on getting the individual comfortable with

the PRoP controls as well as assessing its use for reading documents and collaborating with others around a table or workspace.

During the second task, the user was instructed by the table confederate to find another location, typically near the blackboard or piano, and meet them there via the PRoP. When the individual arrived a brief discussion ensued for several minutes after which the experiment was ended and a post evaluation discussion occurred.

#### **9.1.4 Results and Discussion**

The experiment was designed to be casual. The first questions focused on the initial basic reaction to the PRoP's appearance. More detailed discussions centered on the controls, UI, applications, missing elements, and specific "annoying" features.

Surprisingly, even with minimal instruction and unfamiliar hardware (*i.e.*, the PRoP) all of the users were able to complete all of the tasks within the time allotted (*i.e.*, 30 minutes). Only one PRoP collision occurred during the course of the tests. The user easily recovered after they became aware they had backed into an object. Recall that presently there is no sensing or camera view from the rear of the PRoP making backing up in unfamiliar locations particularly precarious. Most users rotated the PRoP in place to turn as permitted by the kinematics of the robot base. The pace of the PRoP's motion, preset to a maximum of 50cm/s, was "fine" for most users with only one individual noting it was "too slow". There was a general consensus that the joystick control was natural and easy to use. At least two individuals found the head pan-tilt control "too sensitive" and hence difficult to hone in on objects of interest while one user paid particular praise to the head control functionality. All seven users employed the keyboard controls and found them

“natural” and “sensible”. Interestingly, all of the individuals used the head to “look down” when navigating the PRoP in tight spaces, such as approaching the table or a person to converse with, even though no formal instruction was given about such control usability.

Every user was able to read the text on the book and used the zooming tools. Users repeatedly and easily positioned the PRoP from 3-20 cm from the edge of the table even though no instruction to do so was given. The time to complete the first task varied from 2-5 minutes.

Universally, every user in the study ranked the audio and video network delay as the most annoying feature in the system. A video delay of 0.5 to 1 second was measured in this setup. Only one user listed the “picture quality” of the system as a problem. Two individuals explicitly mentioned their surprise at being able to discern various people’s facial expressions with the PRoP’s video. Although all the users were able to operate the system successfully, several users expressed difficulty navigating as a result of the video delays. The conversations during the second portion of the test also contained a 0.5 to 1 second delay. The microphones used contained no echo cancellation; hence users would hear themselves slightly delayed on the PRoP. All of the users again commented on this but said that by the end of the test they found themselves adapting to the delay. They did this by allowing extra time between responses during dialog. At least one user found it extremely difficult to construct lengthy verbal responses due to the aural confusion resulting from hearing their own voice feeding back. Another found the PRoP motor sounds less than subtle.

Users were fairly split about missing elements. Extra navigational tools such as some range data to nearby objects and “a camera on the feet” were mentioned. One user

was specific in needing a split screen, using one camera near the base to navigate while the other camera could be used to glance around from head height. Another noted the difficulty in judging distance. The other missing element was a hand or hand-like tool. But again users were split when asked if such a hand would be frightening to individuals near the PRoP. Most people explained that if the hand were offering something as it approached, as in friendship or some form of greeting, the intimidation of the hand would be minimized or eliminated. That is, a PRoP initially carrying something in its hand would be acceptable. The hand addition came up most often for aiding in communication such as shaking hands. However, at least one individual noted that it was needed to perform remote work with the PRoP.

Several other PRoP appearance themes were discussed. Overall people found the PRoP friendly and non-threatening. One individual made particular praise of the shape and fact that there were no protrusions beyond the extent of the base. Two individuals, whose jobs involved interacting with children, independently pointed out that although they could imagine numerous applications for PRoPs in their work, its height was a major hindrance for interacting with children. For them, shifting their own height to be eye-to-eye with children was a necessary element of successful communication that the PRoP did not facilitate. Another physical feature was the “big bulky-shiny tires” that made the PRoP look intimidating to one individual.

Finally, most people could easily imagine using a PRoP for browsing places and visiting with people. However, there seems to be a universally agreed upon “acceptable use” policy. Users in general felt comfortable using it with people they already knew or those



expecting the arrival of a PRoP - but not in other situations. In a more concrete example, one user was excited about a PRoP to go to a cousin's party saying "he would love it" but not to a store noting, "it may catch people off guard and frighten them."

## 9.2 Experiment 2: Network Effects

The majority of PRoP usage, including Experiment 1, operated within local networks. However, our objective goal has always been that PRoPs would be usable by "Anyone, Anywhere." Experiment 1 attempted to explore the "anyone" issue, but could people actually navigate around a crowded building from anywhere on the internet? Experiment 2 explored this question.

### 9.2.1 Users

The experiment involved four individuals (3 male / 1 female) ranging in age from 26 to 43. Computer usage experience varied from expert to intermediate. While one of the users had viewed and used the PRoP previous to the experiment, the remaining 3 had no prior PRoP experience. In fact, two of the individuals had never even seen the PRoP nor been to our lab.

### 9.2.2 Setup

PRoP 1 was used exclusively for this experiment. Each user was sent a brief email explaining the functionality of the PRoP (*i.e.* various joystick and keyboard controls). At a predetermined time, the individual connected to the PRoP using a Java applet and standard

video-conferencing software. The software setup was typical of what a remote user would encounter. Individuals used their own computers and hardware as a remote user would. Remote connections were made from distances of over 9000 km (5600 miles) between the United States and Europe as well as locally within the San Francisco Bay Area. Diverse networking topologies such as T1, DSL, and cable modems were employed during these tests.

### **9.2.3 Method**

Once connected and basic communication established, the P<sub>RoP</sub> user was led around a laboratory environment by a local guide. The guide pointed out projects and objects in the room with the P<sub>RoP</sub> user free to communicate regarding any of them. Next the P<sub>RoP</sub> user maneuvered through a rather cluttered lab and engaged other individuals and groups in conversation about their work. During the session the guide helped direct the remote P<sub>RoP</sub> user through the tasks. These requests were at a very high level. For example, the guide would say, let's go into the hallway. However, it was up to the P<sub>RoP</sub> pilot to decide when, which direction, and through which door. After several minutes the P<sub>RoP</sub> left through an open door and down a hallway as they walked with a confederate guide. After several spontaneous encounters within the hallway with people, the P<sub>RoP</sub> user maneuvered outside of the building. During these encounters with others, the guide would leave the nearby vicinity of the P<sub>RoP</sub> so that conversation would proceed through the P<sub>RoP</sub>. This also forced the P<sub>RoP</sub> pilot to answer questions about the P<sub>RoP</sub>. Finally, near the end of each hour-long session, the P<sub>RoP</sub> user re-entered the building and said goodbye before terminating the visit and logging off the P<sub>RoP</sub>. A sequence of some of the images of this

interaction can be seen in Figures 9.3 and 9.4.

#### 9.2.4 Results and Discussion

This test was run near the extremes of expected PRoP operation, with users from Europe, over 9000 km away, attempting to visit a lab, meet people, and converse with them. Not surprisingly, network delays were again listed as the major usability problem. Peak delays of up to 2 seconds were reported in this experiment. Since the interactions in this experiment were more spontaneous, users specifically noted that the delay made gaining people's attention difficult. Typically, a person would be walking down a hallway towards the PRoP. By the time the remote PRoP user saw the person and called out, "Hello, how are you doing?" the intended recipient had already walked passed the PRoP. Again we observed people adapting and compensating to this with earlier, shorter verbal call-outs. This corresponds with what Clark refers to as the various "costs of grounding" in communication, particularly in regards to the "conversation start-up costs" [Clark and Brennan, 1991]. Grounding is the process of augmenting and maintaining a common ground between conversing parties. It implies communication, diagnosis (to monitor the state of the other collaborator), and feedback (acknowledgment, repair, *etc.*). Clark and Brennan established that the cost of grounding varies according to the medium.

In every case the users were able to complete all of the operations within the time allotted. Again we found this separate set of users mentioning that they adapted to the delays for navigation. An exciting result is that the users from Europe were able to successfully use the PRoP. Although one of the users connecting from Europe had never seen the PRoP, used the system, or visited our lab, the test went off without even a single



**Figure 9.3:** PRoP 1 during Experiment 2 as operated outside the computer science building at UC Berkeley. The remote user is controlling the PRoP from Denmark in this image.



**Figure 9.4:** Close-up view of PROP 1 during Experiment 2 while under the control of a remote user in Denmark. You can see his video image on the screen while the inset image is the view he is receiving.”

collision.

One user's occupation was as a journalist and described using the P<sub>RoP</sub> as a "thrilling experience" that "really felt like I was visiting your lab at Berkeley." This individual strongly felt that using the P<sub>RoP</sub> provided them with the important tools that they needed for completing their job as a journalist. That is, to look around a lab or location, assess the current situation, approach particular individuals, and interview/interact with them appropriately. This user has subsequently used the P<sub>RoP</sub> to aid in gathering information for news stories at our university. Another user from Europe who had never been to the United States remarked at the end of the session, "Well I guess I've been there [to the United States] now."

While this experiment was fairly unstructured we learned that P<sub>RoP</sub>s are usable over great distances and with large latencies. We were very encouraged with the rapid learning curve of these new P<sub>RoP</sub> drivers. The feeling was that many of the interface choices that we had made were easy to comprehend and use. While the first test was in a fairly controlled environment, it was reassuring to see the P<sub>RoP</sub> function on the real internet under conditions we would expect it to face if deployed in an office, laboratory, or shop floor.

### **9.3 Experiment 3: Airport Traveler's Survey**

Partially to gauge reaction to the P<sub>RoP</sub> concept we decided to conduct a survey of people who would be likely early adopters of P<sub>RoP</sub> technology. We also hoped to perform some application studies during this survey. We had been working too close to the flame so to speak and needed to gain some insight from people away from academics and research.

We went to potential early customers of PRoPs in this study.

One group of people that may use PRoPs are those that travel often. Our claim is that many of the goals of these trips could be adequately satisfied with a PRoP. Therefore, we went to an airport where we could be sure we would find individuals that would be traveling. We asked them how the introduction of a PRoP may influence their travel and lifestyle.

### **9.3.1 Users**

The survey involved 25 individuals (16 male / 9 female) ranging in age from 26 to 56. Computer usage experience varied from expert to intermediate. None of the users had any prior knowledge of the PRoP nor were any of them paid for their participation.

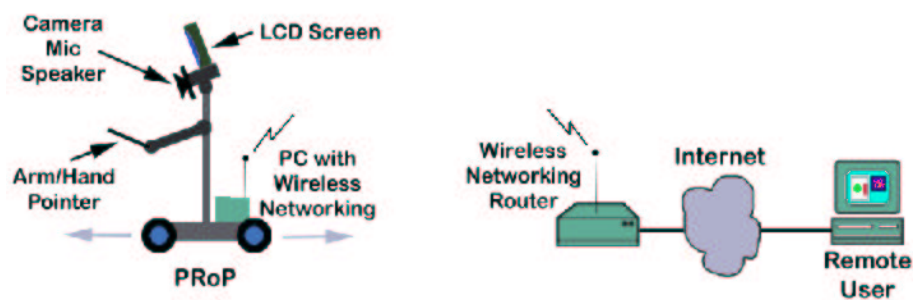
### **9.3.2 Setup**

A group of three individuals with expert knowledge of the PRoP were provided with a two page color leaflet (see Figure 9.5). These people were distributed around the San Francisco International Airport. Each of them were instructed to dress and act casually and represent themselves as students working on a project that they needed feedback about.

### **9.3.3 Method**

These individuals approached airport patrons at random within the domestic travel area, explained that they were students that needed some feedback about a project they were working on that involved allowing people to travel to distant locations through a novel form of tele-presence rather than physically traveling. After a 2-3 minute explanation of the





**Figure 9.5:** Images from the color leaflet used to explain the PRoP concept to people encountered at the San Francisco International Airport.



system and viewing of the images of the PRoP to aid their description, a series of questions were asked. The questions are listed below:

- Would you use something like this? When? Why?
- What sort of business are you in?
- Why are you traveling today? Conference? Meeting? Sales? Customer visit? Product support? Site Inspection?
- How often do you travel?
- Do you travel to the same place? For the same job? To visit the same customer or contact?

Then a 10–20 minute session of free form discussion about the project occurred. During this time novel applications for the PRoP were solicited from airport patrons.

#### **9.3.4 Results and Discussion**

A table of the evaluation of the PRoP questions and feedback is listed below for most of the participants. The missing participant feedback data is either due to individuals that terminated the discussion (typically to catch a flight) or lack of any real data. For example, some of the participants consumed the entire session asking questions about the PRoP without answering any of our questions. Other individuals provided such little data that it is not included. Either they were unable to figure out what the PRoP did (even when provided the photo in Figure 9.5) or their answers were nonsensical or incomplete (*i.e.*, “Can you use the batteries in it to charge your laptop?”).

<b>Age</b>	40
<b>Sex</b>	Male
<b>Job</b>	Lecturer at UCSF
<b>Use PRoP?</b>	Maybe
<b>Application</b>	Remote teaching or rural doctors
<b>Comments</b>	

<b>Age</b>	42
<b>Sex</b>	Male
<b>Job</b>	Electrical Power Sales
<b>Use PRoP?</b>	Maybe
<b>Application</b>	Customers could use PRoP to tour his power plant
<b>Comments</b>	User of PRoP needs a strong personality in order to not be ignored

<b>Age</b>	40
<b>Sex</b>	Male
<b>Job</b>	Field Engineer / makes robotic equipment for fabs
<b>Use PRoP?</b>	Yes
<b>Application</b>	Use to allow field engineer to troubleshoot from remote location / use to support environmental hazard team / send in PRoP if there has been a leak
<b>Comments</b>	Loved it / wanted to discuss further and introduce the concept to other field engineers / he travels over 50 percent of the time

<b>Age</b>	40
<b>Sex</b>	Female
<b>Job</b>	Parole Officer from Quebec
<b>Use PRoP?</b>	Maybe
<b>Application</b>	She would use it to visit ex-cons on parole in the community
<b>Comments</b>	Her husband is a salesman and she thinks he would not use it as he needs more face to face

<b>Age</b>	50
<b>Sex</b>	Male
<b>Job</b>	Architect
<b>Use PRoP?</b>	Maybe
<b>Application</b>	Inspect work on a construction site / also saw application for real estate developers
<b>Comments</b>	Would need to be very durable and mobile for use in construction / tends to visit the same site over and over

<b>Age</b>	35
<b>Sex</b>	Male
<b>Job</b>	Investment banker
<b>Use PRoP?</b>	Yes
<b>Application</b>	Conferences and trade shows would definitely use it
<b>Comments</b>	Would not use for client work, needs face to face

<b>Age</b>	40
<b>Sex</b>	Male
<b>Job</b>	Web server sales
<b>Use PRoP?</b>	No
<b>Application</b>	
<b>Comments</b>	Not much better than video-conferencing

<b>Age</b>	55
<b>Sex</b>	Male
<b>Job</b>	CEO of manufacturing company
<b>Use PRoP?</b>	No
<b>Application</b>	
<b>Comments</b>	Needs face to face time to make it happen

<b>Age</b>	45
<b>Sex</b>	Male
<b>Job</b>	Financial Services
<b>Use PRoP?</b>	Yes
<b>Application</b>	Visit remote offices
<b>Comments</b>	Travels often to other offices in his company

<b>Age</b>	42
<b>Sex</b>	Male
<b>Job</b>	Chemical Engineer
<b>Use PRoP?</b>	Yes
<b>Application</b>	Site inspection
<b>Comments</b>	Travels to Europe inspect and setup chemical processing equipment / would use to support critical remote site for company

<b>Age</b>	32
<b>Sex</b>	Male
<b>Job</b>	Project Manager
<b>Use PRoP?</b>	Maybe
<b>Application</b>	
<b>Comments</b>	Would like to see lighter weight portable version

<b>Age</b>	35
<b>Sex</b>	Female
<b>Job</b>	Board of Directors
<b>Use PRoP?</b>	Yes
<b>Application</b>	Tradeshows and remote collaboration with others
<b>Comments</b>	Visits same location every month / would use as inter-office proxy

While users never experienced the actual PRoP during this survey, it was encouraging to see people's enthusiasm about the project and their own view of where PRoPs fit into their lives. From the feedback, PRoPs used for tele-inspection and remote collaboration are promising application areas. The PRoP could support early troubleshooting of a problem at a remote plant before sending personal out. Even if a person was ultimately required, a PRoP enabled remote expert could eliminate much of the wasted troubleshooting time of the expert enroute before their arrival. Many of the people felt uncomfortable using it for face-to-face sensitive tasks such as sales or customer interactions. This was expected but it was important to establish that with an independent study. People who were already in a technical field appeared more likely to adopt the PRoP to substitute for some of their travel.

## 9.4 Experiment 4: Student Group Study

With the success of Experiment 1 we wanted to carry out a better study on how usable the PRoPs really were by a larger group of people. Our goal was to provide a less structured forum for the experiments so that users would use the PRoP in a more free form manner rather than being focused on a particular task.

### **9.4.1 Users**

In a loosely structured setup, a group of 15 non-computer science students (6 female / 9 male) participated in this study. This group ranged in age from 19 to 23 and most had intermediate levels of computer experience. This time all of the users had experience with videoconferencing, specifically NetMeeting. None of the users had viewed the PRoP until the day of the experiment. However, all of the users had met each other previously and had collaborated on projects before. These students were a collection of undergraduate research assistants that maintained videoconferencing computer equipment used in running experiments in social psychology. None of the participants were paid.

### **9.4.2 Setup**

With all of the users in the same room, they were given a 5 minute explanation of the PRoP input and control devices as well as its functionality. PRoP 2 was used exclusively for these experiments. Each user was instructed to learn to use the PRoP well enough while operating it to explain it to the next user. They were also told that they would be responsible for using the PRoP in a later experiment and would need to explain and demonstrate its usage to others. The PRoP was placed in a hallway separate from the room where the individual worked with the computer to control the PRoP.

### **9.4.3 Method**

The users were broken into roughly two groups. The first group watched as the first individual used the PRoP. On the other end the second group interacted with the PRoP. They were given no formal instructions on what to do. However, this group spoke

with their colleague on the other end and led them around the building, into rooms, and down hallways. They introduced them to others in the group. After 15 minutes, a new user took over, receiving explanations from the previous user. That user then rotated out to the P<sub>RoP</sub> interaction group. Similarly, a user from that group moved into the P<sub>RoP</sub> driving group. This proceeded without intervention from the experimenter, except to synchronize user rotations.

#### **9.4.4 Results and Discussion**

The results were encouraging for this test. First, users were able to learn basic control of the P<sub>RoP</sub> after 15 minutes and pass sufficient knowledge of this on to subsequent users. Also, again we observed a group of people who had never used the P<sub>RoP</sub> being introduced to it and controlling it successfully. The locally embodied users interacted comfortably with the primary P<sub>RoP</sub> user. There were some navigation problems, typically from users losing track of their forward direction when panning and tilting the camera. Also, a user would often leave the P<sub>RoP</sub> in some non-standard “home position” when handing off usage to the next driver which caused large amounts of confusion.

We noticed that since these users all had experience with NetMeeting and hence the delays associated with network videoconferencing, we received no complaints about network delay or response time. The dominant complaint was that the system moved too slowly when driven. The remotely embodied P<sub>RoP</sub> driver felt that they could not keep up with the movement and gait of the locally embodied users within the remote location. Similarly, the locally embodied users complained that they were bothered by always having to wait for the P<sub>RoP</sub> to catch up to them when maneuvering throughout the building.

Again, we found encouraging results of novice P<sub>RoP</sub> users being able to operate the P<sub>RoP</sub> in an unfamiliar location. We also wanted to know if people would be able to explain the P<sub>RoP</sub> to others. In fact they did. Most of this work was preparatory for future experiments when groups of users will need to perform tasks and get to know each other. We'd like to be able to understand how long it takes people to learn to operate the P<sub>RoP</sub>. We'd also like to know at what level of proficiency they control it. From this survey we found 5 minutes a reasonable time to explain the basic P<sub>RoP</sub> function and operation and 10 minutes adequate to establish basic confidence using the P<sub>RoP</sub>. Users blazed into cluttered classrooms and down hallways with little problem – until switching off to a new user. It was then that the learning started over again for the next new participant.

## 9.5 Experiment 5: Tele-Lecture

Often the claim has been that the P<sub>RoP</sub> could be used to attend a meeting remotely or deliver a remote presentation to an audience. This experiment was designed to measure P<sub>RoP</sub> performance in exactly such applications.

### 9.5.1 Users

The author of this dissertation and an admittedly expert P<sub>RoP</sub> user were invited to deliver a 90 minute presentation to an audience of 50 or more individuals.<sup>1</sup> The audience was composed of academics (students and professors), mainly in the fields of computer science, electrical engineering, communications, sociology, and psychology. They ranged in

---

<sup>1</sup>This was for Terry Winograd's 10 March 2000 Seminar on People, Computers, and Design at Stanford University. The talk was titled "Social Tele-embodiment: Understanding Presence".

age from 18 to 57 with a 30/70 ratio of females to males. Almost all were expert computer users with previous use of some form of videoconferencing.

### **9.5.2 Setup**

The presentation was in a lecture hall. The topic was a lecture on the contents of this dissertation. The primary PRoP user for this experiment was the author of this dissertation. In this experiment PRoP 0 was used. Prior to the arrival of the audience, the PRoP was placed out of view and the primary PRoP user left the room to control the PRoP from a separate location. The host introduced the speaker at which point the PRoP maneuvered out and began the presentation. During the lecture, the PRoP moved around the stage area of the room (approximately 25 square meters) as the PRoP hand gestured along with with the head. During the talk, live video of the primary PRoP user's face was placed on the onboard screen.

### **9.5.3 Method**

For the first hour of the presentation the PRoP was used to lecture. After that point, the remotely embodied PRoP user (*i.e.* the speaker) came out and presented directly (*i.e.* without using the PRoP). At the end, questions were taken from the audience.

### **9.5.4 Results and Discussion**

The effectiveness of the lecture at presenting the material was measured by the level of confusion and questions and confusion from the audience. Subjectively, we would say that these were consistent with questions during an in-person presentation. Also, the



timing of the presentation was similar to an in-person presentation.

However, some interesting comments were made by the audience. One individual had apparently attributed a height to the PRoP speaker based on the actual height of the PRoP. When the remotely embodied PRoP user presented himself in person, this individual was surprised at his height. Overall, this test was less formal than the others but demonstrated that PRoPs can be used to present material to a group in a classroom or lecture style setting. One worry is that, since the PRoP is still quite novel, its appearance for a lecture distracts from the content of the lecture. That is, rather than listening to the presentation, members of the audience are wondering what abilities the PRoP has, where the remote driver is located, how long the battery life is, *etc.* It may be useful to explain a bit more about the PRoP at the start of such lectures in the future, at least for some time. Since this presentation, a new feature has been added to the PRoP to easily interface it to a projector screen for lectures.

## 9.6 Summary

While most people who see a PRoP are fascinated with it and the remote user for a short period of time, typically 10-15 minutes, we are more interested in how to design PRoPs to fit into the daily lives of humans. We designed a series of experiments, questions, and surveys to observe how people would react to them, use them, and adapt to them. Most of the results were very encouraging and have established a groundwork of design choices for future PRoPs.

We believe PRoPs will be an important part of the CMC landscape in the coming

years. As a result, there will be important studies done into the usability and social effects of PRoPs. There will be in depth studies of various gesturing methods and techniques, attempts at measuring trust using PRoPs, optimal robotic kinematic designs for PRoP bases, color and shape choices, *etc.* However, for now these tests are beyond the scope of this dissertation (see Chapter 11). We expect this work to be presented in a collection of follow-up studies now underway. However, the brief introduction of PRoPs to the work during the experiments in this question have raised a number of social and ethical questions concerning their use. We confront them directly in the next chapter.

## CHAPTER 10

# SOCIAL IMPLICATIONS

*[My project] is something we're all intimately interested in: the reshaping of the human body by modern technology. — Vaughn in Crash by J.G. Ballard*

Two-way communication between humans with PRoPs creates a myriad of interesting remote experience possibilities as well as social dilemmas. In the previous chapter we observed several experiments raise some of these issues. We address several of these topics in this chapter.

### 10.1 Human Acceptance and Interaction

PRoPs are strange devices and are often met with equally inquisitive looks. The major challenge will be to negotiate their acceptance into the social world of humans. A device that is unapproachable, seemingly odd (read “frightening”), or awkward to be around will fail in its quest to facilitate human communication and interaction. Much of the resolution of this problem will emerge from the well understood area of social psychology coupled

with research into technological acceptance. PRoPs must appear friendly and amicable to interact and cohabit with. Their actions should be smooth, calm and predictable.<sup>1</sup> Will PRoPs be welcome or loathed guests at a talk or meeting?

One related psychological experiment undertaken in the 1970's by Stanley Milgram [Milgram, 1977] involved the use of what he called *Cyranoids*, so named for their relation to the character Cyrano de Bergerac in Edmond Rostand's play by the same name [Rostand, 1923]. Cyranoids are people who do not speak thoughts originating in their own central nervous system. Rather, the words that they speak to a listener originate in the mind of another person, called the source, who can listen to the space around the cyranoid and transmit words to the cyranoid by means of a wireless transmitter. A tiny radio receiver inconspicuously fitted into the cyranoid's ear is used to receive the words from the source individual. The similarity is apparent in PRoPs act as cyranoids for the pilot of the PRoPs, or what Milgram would call the source. Milgram carefully constructed experiments where individuals would interact with a cyranoid with characteristics differing significantly from those of the source. In other words he wanted to measure how people assign identity. Will they base it on the physical appearance of the person in front of them even though every word and action is originating elsewhere from another person. Or will they assign the identity to that remote person controlling the actions.

For PRoPs, we are interested in the capability of people to interact with PRoP proxies (*i.e.*, cyranoids) in a natural way. Milgram concluded from his experiments that people are unable to separate the cyranoid from the source and always assume that the

---

<sup>1</sup>This is a problem if a PRoP user wants to be quick and abrasive. But for now their acceptance outweighs encompassing the full gamut of human emotions.

individual with whom they are interacting is the cyranoid. Remember, that this is with near perfect information since the individual can ask any question and observe every detail of the cyranoid's actions.

The main result we can draw from Milgram's work is that appearance is extremely important in human interaction. Appearance is so important that people in his experiments continue to believe that the person with which they are interacting is the person in front of them even when every element except the vocal-movement is removed from that person to another distant person. Therefore, if P<sub>RoP</sub>s too closely resemble humans, locally embodied users communicating through them will misinterpret the correct source of the transmission. Current P<sub>RoP</sub>s are quite far from resembling a human and thus we avoid this Cyranoid-delemma. However, Milgram's results are impressive and an important reminder about the importance of physical appearance and the confusion it can cause for users.

The next few years will see a wealth of various robots come into our social arena. Some will be robotic pets or office assistants [Kaminsky et al., 1999], but many will be personal robots. The role that these systems play in our lives will be largely dictated by our emotional and physical reaction and interaction with them. These are, unfortunately, areas of research most computer scientists are untrained to tackle. Designers, social psychologists, and others will need to make concerted efforts to shape the form, actions, use, and applications of these technological systems early in their development.

## 10.2 Privacy and Security

As fascinating as a floating blimp or cruising P<sub>RoP</sub> is, it is somewhat disconcerting when you wonder who may actually be at the other end of the connection. Should that person be allowed access to the space occupied by the P<sub>RoP</sub>? We envision a system to easily provide secure access to P<sub>RoPs</sub> using techniques not too different from those currently used to limit access to individual files and computer hardware. People would invite individuals into their local space by issuing them a digital token that would authenticate and validate individual access to the local P<sub>RoP</sub> and hence the space. Also, although they are designed to be small and agile so that they can navigate within a building, P<sub>RoPs</sub> have no super human abilities that enable them to pass through walls, push elevator buttons, unlock doors, or for that matter even open a closed door. This means that P<sub>RoP</sub> accessability can be filtered using the same methods as people. A locked door still means don't come in and neither a person nor a P<sub>RoP</sub> can thwart this barrier.<sup>2</sup>

## 10.3 Authenticity

While developers of tele-robotic tools on the internet strive to create the most realistic impression of “being there” in the remote space, users are all too often left wondering if what is on the other end is real. After all, they have little information to go on. In fact, since all of the information delivered to them is digital as a result of the underlying transport mechanism, any images and video could easily be faked by cleverly fetching them from a catalog of pre-scanned images. In fact there have been a number of creators

---

<sup>2</sup>This is not an issue for an adept locksmith.

of purportedly tele-robotic sites over the last decade that have been exposed as charlatans (see Section 4.2).

We find that in a typical tele-robotic system, users often perform a series of what appear to be *ad hoc* tests in an attempt to convince themselves that the tele-robotic system is in fact authentic. For example, visitors to the Mercury Project [Goldberg et al., 1995a] (see Section 5.2.3) would move to look down on a watch, only to return a few minutes later to check that it had actually advanced as expected in real life. Several users of our system have in fact claimed that implementing a real system would be much too difficult and that what we have presented is merely a hoax. Our belief is that users are developing their own form of what we call the *Tele-robotic Turing Test*, named to reflect its obvious relation to a test described by Alan Turing in 1950 [Turing, 1950]. In the original Turing Test, an experimenter is connected to a human and a machine via a simple terminal. Through dialogue conducted only through the terminal, the user must determine which of the two candidates is the machine and which is the human. If after a reasonable amount of time the interrogator cannot make a decision, the machine is labelled intelligent. Similarly, tele-robotic experimenters must go through a set of possible queries of the remote robotic system. However, unlike the actual Turing Test, the choice is not between a machine and a person. Instead, the user is weighing what they believe can be digitally faked using existing technology and their knowledge of the real world from their daily life. We developed Legal Tender to directly address this issue of authenticity in relation to web based tele-robotics (see Section 4.2)

Perhaps at the core of this debate is the philosophical issue of epistemology, which

is a branch of philosophy concerning the nature, origin, scope, and limits of human knowledge. When users engage and question such remote systems we claim that they are in fact exhibiting a new form of this philosophical question, which we call *tele-epistemology*, originally coined by Ken Goldberg [Goldberg, 2000, Goldberg and Spaid, 1996] (also see Section 4.2.2).

## 10.4 Responsibility

Trust is one of the most precious commodities on the internet. Its rarity must surely be due in large part to the lack of consequences of actions online. We can sling insults and flames at others in cyberspace, but they can filter us out. We lack the ability to threaten and harm. The possibility of “hurting” them or affecting them in any physical way is practically nil. The consequence is that trust is a nearly unobtainable commodity online. Face-to-face remains the best way to build it.

This presents a dilemma for P<sub>Ro</sub>P<sub>s</sub>. They must be safe to interact with. Who would choose to unleash a harmful robot, controlled by others within their home or office? However, it is clear that P<sub>Ro</sub>P<sub>s</sub> embodying only these limited docile abilities may be limited in the types of emotions and actions it can express.

But that situation is changing. Telepresence projects have broken down the safety barrier between cyberspace and real space. Scientists and technicians at Survival Research Laboratories (SRL) along with the author of this dissertation built a telerobotic system in 1997 and again in 1998 capable of automatically targeting, arming, and delivering live explosive ordnance [Pauline and Paulos, 1997b, Pauline and Paulos, 1997a]. It was controlled



from the public internet.

Lethal telerobots are an extreme and rare occurrence on the net. But as ubiquitous computing connects cyberspace to more kinds of sensors, actuators, controllers, and appliances, our reach to others in the real world will be greatly extended. The internet has long ceased to be a safe playground from a psychological point of view [Turkle, 1998, Turkle, 1997]. We may need to abandon our assumptions of physical safety as well.

When control of a physical mechanical system is accessible by anonymous individuals, great precautions must be taken to insure the safety of people and objects sharing the space with the PRoP. We are all aware of the interest in hacking into computers and manipulating, stealing, or destroying digital data. One can easily imagine the fascination of taking control of a potentially dangerous device to use to one's own ends.

The tele-robot's abilities, physical attachments, and even the basic construction must all be considered. Even an out of control PRoP must safely interact with humans and property. Since we desire these tele-robots to co-inhabit with humans, this constraint is of the utmost importance.

The design challenges for safe robots are numerous and may take years to resolve. For example, even a seemingly un-threatening and "safely" designed, but slightly heavy, tele-robot may accidentally be knocked over near a stairwell, causing it to tumble recklessly down the stairs, and impact an innocent human in the stairwell. In 1981, Kenji Urada learned this lesson the hard way when he unfortunately became the first person to be killed by a robot.<sup>3</sup>

---

<sup>3</sup>It's important to note the circumstances at the Kawasaki factory where the incident took place. At the time, the robot was malfunctioning, Kenji climbed over a protective rope and accidentally hit the on switch which caused the robot to deliver the fatal blow.

The importance of the safe co-habitation requirement cannot be overstated as it relates fundamentally to the acceptance, approachability, friendliness, and interactivity of PRoPs and humans. Unlike many other robots, it is vital that safety be a primary concern when designing PRoPs. This issue is so important that we propose a variation on Asimov’s first law of robotics for PRoPs.<sup>4</sup>

At no time should a PRoP ever be capable of physically injuring a human being regardless of the action or inaction of the remote tele-operator.

**Table 10.1:** First Law of Personal Tele-embodiment.

## 10.5 Intimacy, Trust, and Persuasion

Along the path to trust and intimacy, there will be studies of communication and persuasion skills. Intimacy is the farthest point along this road for our studies with PRoPs. If researchers don’t succeed at the first two, the third will almost surely fail. Or will it? Our intuition at first was like most people: we balked at the conjunction of notions of “trust” and “intimacy” with “robot” and “telepresence”. Hubert Dreyfus takes a strong stand on this issue:

Even the most gentle person/robot interaction would never be a caress, nor could one use a delicately controlled and touch-sensitive robot arm to give one’s kid a hug. Whatever hugs do for people, I’m quite sure tele-hugs won’t do it. And any act of intimacy mediated by any sort of prosthesis would surely be equally grotesque. [Dreyfus, 2000]

---

<sup>4</sup>“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 [Asimov, 1950]

It is hard to disagree with this point of view. So we will agree, and sidestep this argument by declaring that future telerobotic systems won't be "robot-like" or "prosthesis-like" at all. Robots and most prosthesis were machines designed for interaction with objects not people.

Let's look instead at today's "social machines". There is a new generation of interactive devices targeted at young children, soft toys with computer cores and capabilities like touch sensing and speech. They include Tickle-me Elmo, Actimates Barney, and Furby, the Model-T of furry automatons. None of these devices can hug or caress, but will respond to a variety of touch from children. They have been carefully designed to match the behavior of pre-school children. They have no behavioral autonomy, but can participate in reasonably complex interactive behaviors (like games of hide-and-seek) that are driven by a child. That is, they are capable of situated activity. More recently several "Mental Commit Robots" as they are called have been designed to "hug back" [Shibata and Tanie, 2001]. There will surely be many studies over the next few years about the educational and social value of these toys [Turkle, 1998, Turkle, 1997]. We won't speculate on the outcomes. But they have been a commercial success, and many children feel strongly attached to them.

A toy that responds to touch by talking or playing encourages the child to use touch to communicate. Without touching back, it reinforces a social bond through the other behaviors it is capable of performing. It is hard to argue that this is a false intimacy. While the automaton's behaviors are simple and programmed, the child's are not. The toy's voice (actually an actor's voice) is repetitive, but entreats the child to participate in activities like game-playing, drawing, and singing. There the child can stretch her skills

at improvisation and expression. The toy participates in those activities as partner or spectator and can provide scaffolding for complex activities. The social behavior of the child is encouraged by the toy, almost certainly strengthening her bond with it. Imagine how much richer this interaction could be if the voice were a parent or teacher's voice, and the interaction were truly spontaneous. Now imagine that as well as sensing a hug, the toy could hug back, again under control of another person. When the behavioral gamut is rich enough and the possibility of two-way touch is there, how important is the "quality" of touch?

## CHAPTER 11

# FUTURE WORK

*Teach us to make Robots. – Fourth Robot speaking in Karel Čapek's R.U.R.*

Personal Tele-embodiment is in its infancy. This dissertation has described the evolution of several of these early systems. In this chapter we provide discussion on future advances with P<sub>RoP</sub>s.

There are wealth of future areas of P<sub>RoP</sub> research that can be explored. P<sub>RoP</sub>s are in their infancy and we will certainly see a myriad of applications, designs, and models. In this chapter we briefly highlight several future P<sub>RoP</sub> improvements. While there are clearly whole new directions that P<sub>RoP</sub>s can splinter into. We will try to maintain focus on some more near term incremental upgrades to the current fleet of P<sub>RoP</sub>s.

## 11.1 Physical Design Improvements

The design space of the P<sub>RoP</sub> form demands extensive serious research and attention. An entire research project needs to examine issues of shape, color, texture, height, size, weight, mobility, *etc.* During this process, various social issues of human communication must also be addressed. This typically involves not just engineers and computer scientists but also designers and social psychologists. One impressive piece of work that nicely incorporates many of these issues is the VisiPhone [Donath et al., 2000].

The current choices made for these P<sub>RoP</sub> traits are well intended and not without thought but do lack formal research. It is certainly possible in the rush to their design that these design choices ignore well understood and accepted practices from social psychology and design.

While the literature is researched and experiments on P<sub>RoP</sub> design are undertaken, a closer look at the “arm/hand” design should begin. Many of the issues here are still open questions? Should it look like a hand? How like or unlike a hand is appropriate? Should it be able to grasp something? Carry something? What is a reasonable range of motions? Is a laser pointer attachment appropriate or too threatening? How can such an attachment be more appropriately augmented? What are appropriate non-verbal cues to perform with the arm/hand?

However, in terms of nonverbal cues, there are numerous important cues that have yet to be explored. So far no attempt has been made to address posture. Future P<sub>RoPs</sub> should experiment with posturing. Humans are able to discern less than a few degrees of off-axis shift in other humans and we expect the same performance with P<sub>RoPs</sub>. Leaning

or shifting towards a person, even slightly, contains a myriad of nonverbal connotation.

Less subtle is the current shattering of the remotely embodied user's immersive experience when their voice echoes back to them from the PRoP's speaker, delayed, through the network. While we observed individuals adapting to this, we are well aware that it makes long conversations nearly impossible. While short utterances are manageable, conversing or lecturing with unbroken<sup>1</sup> speech for more than about 10 seconds is nearly impossible. Humans have no previous or learned experience speaking when they can hear their own voice delayed.<sup>2</sup> Users almost always stop talking when the feedback becomes impossible to ignore – typically 5-10 seconds. The good news is that a solution exists. It requires some software and hardware changes to improve the currently non-existent noise and echo cancellation properties of the microphone. Luckily, with rapid adoption of video-conferencing tools, cellular telephone headsets, and speech input devices for computers, this problem is being formidably attacked. A sufficient solution certainly exists and could be found with some amount of research.<sup>3</sup>

While too much audio feedback is a problem, additional video feedback is desperately needed. It is clear from our own use and the feedback from others that a separate camera view is necessary. This led us to identify the need for at least two levels of video resolution or “Telepresence with Extra Eyes” [Yamaashi et al., 1996]. The system should provide a wide angle near or pointed at the base of the PRoP to aid in navigation while the other camera should be used for gazing, glancing, navigating, and social interaction.

---

<sup>1</sup>By unbroken we mean that the PRoP driver is speaking without pausing.

<sup>2</sup>This is why radio stations using a 5-10 second broadcast delay to enforce censoring ask callers to the station to turn their radio volume down or off.

<sup>3</sup>Currently, most instant messaging systems include a voice chat mode with the same software echo cancellation that is needed for the PRoP.

The exact solution is not yet clear. One method would display the remote user with a split-screen display showing the lower camera view in the bottom portion of the display.

## 11.2 Force Feedback

There are also a number of interesting approaches that could be made with a force-feedback joystick. For example, fluctuations of the magnitude and frequency of the joystick vibrations as well as its perceived mass spring properties may allow a user to receive rudimentary tactile feedback about the state of the remote PRoP (*i.e.*, moving, idling, controlling head, moving backwards, *etc.*). It may also be useful for transmitting proxemic cues (see section 7.10.3). For example, the sonars provides useful distance data to nearby objects (and people). The force-feedback joystick can provide a nice tool for wall following and gauging distance to people. Moving closer to an individual (or object) could tighten up the PRoP controls making the PRoP driver well aware of the closing proximity to others. The pilot would have little doubt that they are encroaching into a person's intimate space, perhaps intentionally.

## 11.3 Haptic Integration

We turn finally to the science of haptics (touch) for clues to the future of social telepresence. As for other areas of telepresence, most haptics research is about interaction with objects rather than people. Much is known about how well we can sense surface shape, texture, and temperature. Much less is known about the role of haptics in social communication.



First some basic facts about touch. While the fingertips, tongue and lips are extraordinarily sensitive, most of the rest of the body is a rather ordinary touch sensor (compared to what can be built artificially). Shaking hands, kissing, and sexual intimacy stretch the limits of our sense of touch. But other social contacts are dampened through several layers of clothing. Is a hug ineffective because it passes through two layers of woolen sweater? Or for that matter, is a handshake ineffective through a set of gloves? We would prefer to shake a real hand, but when we cannot is there social value in a “low-resolution” version? Does it retain the same connotation as a real handshake and if not how useful is the resulting haptic transmission? Future PRoP research should explore this important extension to tele-embodiment.

## **11.4 Interface and Navigation Evolution**

The interface has evolved rapidly over the years but this has led to an incremental layering approach to its design. A unified effort to design a simpler, more usable interface would be invaluable. Particularly useful would be the incorporation of point-and-go and map-building navigation aids. Point-and-go interfaces would allow for easier navigation of the PRoP to locations of interest while map-building tools would build simple maps of a PRoP’s travels, augmented with images and visualization tags. Other important interface improvements would be some form of intuitive visual feedback of sonar, heading, and wheel encoder data.

## 11.5 Go There Now: Point and Click Interfaces

Progression from mouse to joystick has dramatically improved usability, but there are still substantial burdens associated with interfacing into a remote world. Anecdotal observations reveal extreme control difficulty and tedium in two fairly common navigational tasks. The most common tasks are requests of the form “Hey I want to go over there” or “I want to get to the end of the hall” or “Look there’s Jack, go over there so I can say hello” or “Move over to that door”. Currently, P<sub>RoP</sub> navigation is in its infancy and requires the remote user to provide continuous control signals (with constant corrections) to achieve these fairly common goals. Recently, work has gone into a prototype system using a single, simple pointing gesture [Crisman et al., 1998] on an image to direct the P<sub>RoP</sub>. However, this system, has yet to be fully instrumented on the P<sub>RoP</sub>.

The point-and-click interface works by allowing a user to select a point of interest within the video image. From that the P<sub>RoP</sub> is automatically guided towards the selected goal. The actual implementation of the system uses one of two different approaches: visual servoing or direct calculation.

### 11.5.1 Point-and-Click: Visual Servo

Analysis of a sequence of video images from a camera as it moves can reveal the direction of motion. Image two images taken some small interval apart as the P<sub>RoP</sub> maintains a constant motion towards an object. If we calculate motion vectors for the pixels in these two images, every point will appear to be moving away from a single point called the *focus of expansion*. The focus of expansion, also noted as the intersection of all

displacement vectors, is the location towards which the P<sub>RoP</sub> is moving. By calculating this point from video images, and servoing the P<sub>RoP</sub> to correct errors in its motion, it can eventually arrive at the desired goal location.

This technique comes from the well established field of computer vision and visual servoing [Hutchinson et al., 1996]. However, the computational resources and quality of the compressed video signal make performing this form of visual servoing difficult on current P<sub>RoPs</sub>. We do expect this form of navigation to be used soon on P<sub>RoPs</sub> as they also provide the ability to track, and hence follow, people (see section 11.6).

### 11.5.2 Point-and-Click: Direct Calculation

The other method to perform the point-and-click operation is to calculate the intersection of the selected point in the image with the physical ground plane of the robot. Viewing a single high quality image, a user selects the location they desire to move toward. They do this by directly selecting on the image. The point selected translates to a ray in space originating from the camera plane. All of the camera geometry is known as well as the pan/tilt angle and lens properties. From this we can calculate the intersection of this ray with the ground plane and move the P<sub>RoP</sub> towards that new location.

In our test setup this version of the point-and-click interface was developed (see Figure 11.1). This interface consisted of an image display area, control buttons, and a status display. A user selects the point-and-click action, which captures a high quality image and displays it for the user to inspect. While it would be nice to incorporate the streaming video from the P<sub>RoP</sub> into the interface, several issues currently make this difficult. It is unclear if the video quality from such a signal would even be of high enough fidelity for this



**Figure 11.1:** Two views of the prototype point and click interface. Before selection (left) and after completed motion (right).

version of the point-and-click interface. More importantly, since the video signal is handled completely by NetMeeting we have no real software hook to enable us to construct our own interface incorporating it. The eventual incorporation of the Java Media Framework (see section 7.8) will provide us with much greater flexibility in the design of the point-and-click interface. A more appropriate integration with the streaming video would then likely occur.

In the test setup, when the user selects the point to move to in the image, motion begins immediately until either the goal is reached, the sonars encounter an obstacle, or the user stops the motion manually either by selecting another point or switching to manual control via the joystick. While the automatic motion is being performed, the calculated distance and bearing are displayed on the bottom portion of the interface. In this simple

version the distance in meters is calculated and numerically displayed. Once a user initiates the motion, the P<sub>RoP</sub> then runs closed loop towards the goal without requiring any input from the user. Of course the user can interrupt the system at any time to steer the P<sub>RoP</sub> to a new location. This method is subject to errors during control that it cannot correct for during its motion. The visual servoing technique does not have this same problem. An early example of this can be found in the WebDriver project [Grange et al., 2000].

## 11.6 Navigation Automation

The design of this interface brings up the autonomous robot navigation problem. This challenge is at the very core of traditional robotics research and it is not the goal of this dissertation to solve that problem. However, recall that P<sub>RoP</sub>s are not autonomous, but controlled remotely with some limited autonomy to aid with navigation. Therefore, the solution to the navigation automation problem relies on direct user intervention to assist the P<sub>RoP</sub> in extremely difficult navigational situations. Initially, point-and-move interfaces will require that the goal be directly visible and have an obstacle free path or a path that can be achieved with a simple obstacle following motion planning algorithm. This control will not require actual features to be selected, but will instead use geometric constraints based on camera properties and position. Again the objective is to develop a hardware and software solution that will succeed most of the time in automating an extremely common navigation task. Expected failure modes include instructing the P<sub>RoP</sub> to move to a location in an inaccessible room<sup>4</sup> (*i.e.*, through a window or closed door), across stairs, onto a table,

---

<sup>4</sup>Obviously, any location requiring “super-P<sub>RoP</sub>” powers will fail. That is, locations inaccessible by a P<sub>RoP</sub> even when controlled by the most skilled pilot.

or across a densely cluttered room of objects.<sup>5</sup>

Other frequent navigation tasks are of the form “ok, I’ll follow you there” or “sure let’s talk while we walk to the café”. This follow-mode automated control is an extension of the previous point-and-click interface from above. However, in this case the target is (1) an actual feature and (2) moving. Successful implementation of this feature into the P<sub>R</sub>oP control architecture will be more difficult, contain more unknowns, and likely be less accurate. Sonar and simple vision processing techniques are likely solutions paths to this goal. For example, a full implementation of the visual servoing technique discussed in section 11.5.1 would allow for tracking and following operations.

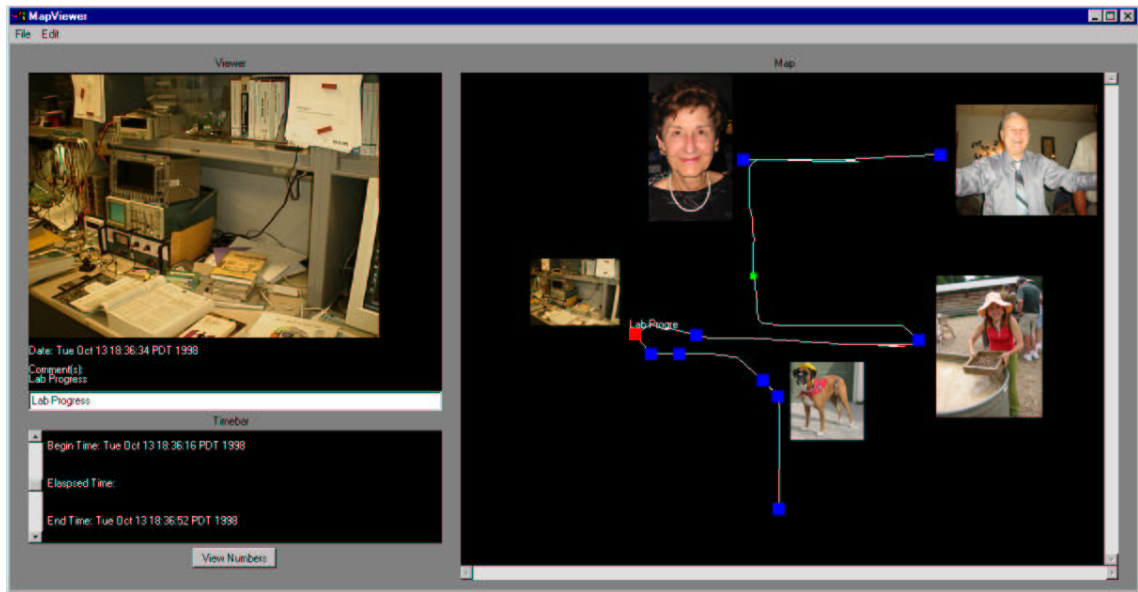
## 11.7 Smile: Creating A Tele-visit Visual Scrapbook

A goal of this research is to allow people to be immersed into real remote spaces and allow them to explore and interact with the inhabitants there in much the same way as they would if they were truly present. However, there are some augmentations to P<sub>R</sub>oPs that can provide the remote individual with an experience beyond what is possible with an actual visit. We have proposed and started construction on portions of this research. Most of this work is left for future P<sub>R</sub>oP development. We leave this section as an outline to designing such interfaces.

Essentially, rough position information from the P<sub>R</sub>oP’s odometry data allows a simple map to be constructed. An individual captures images that are automatically embedded into a simple map. One can imagine a user browsing the hallways and laboratories

---

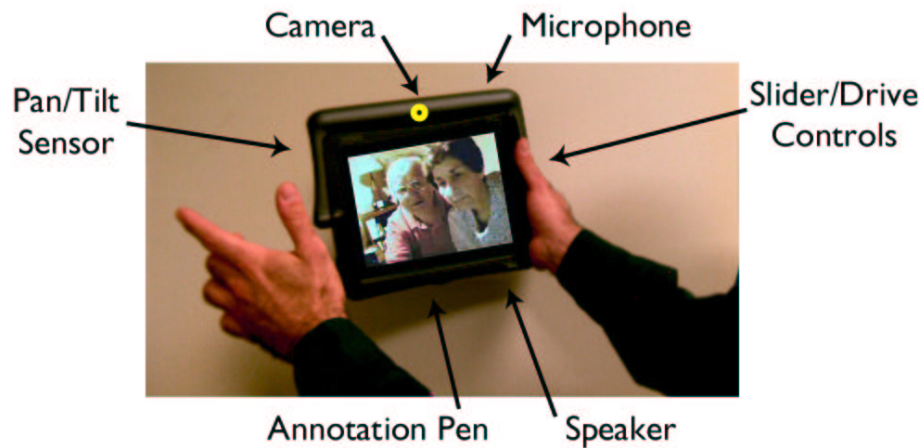
<sup>5</sup>This will typically fail because such a room will appear to the P<sub>R</sub>oP navigation hardware as a single continuous insurmountable obstacle. This is a resolution deficiency in the sensing hardware.



**Figure 11.2:** The prototype tele-visit visual map construction tool. The path is shown with “interesting landmarks” as images. Higher quality images are displayed in the box on the left with an associated date/time stamp and comment.

of a building. The PRoP would build a crude, but usable, low-fidelity map augmented with images of people, events, and items of interest encountered during the visit. The map can be edited and comments added either interactively or off-line. The overall result is a temporal and spatial remembrance tool of an individual’s tele-visit. One of the early prototype systems we constructed is shown in Figure 11.2. The data collected during such visits can easily be recalled from storage even when the PRoP is off-line, creating a sharable history of an individual’s visit.

The importance of “activity detection” systems would prove useful in the construction of such maps. “Activity detection would automate the map/visual construction process by transparently identifying times of “importance”. The idea is that by analyzing the sensing hardware already on-board the PRoP, fairly accurate predictions about these



**Figure 11.3:** Panel PC interface to a PRoP.

“important” moments can be made. For example, if the PRoP is stopped, exhibiting high levels of pointer gesturing, and high throughput of bi-directional audio data, it is likely that a conversation is taking place and the map building software could automatically capture several high quality images and place them into the corresponding map position, thus recording the event. The hope is that the resulting maps will mimic the episodic style of encounters and events that are typical of the human interaction routine as well as their cognitive remembrance model [Richardson et al., 1994, Rhodes, 2000]. It is also useful for automatically generating meeting and activity summaries and indices [Li et al., 2000].

## 11.8 Panel PC Interface

There has also been some initial work into a panel PC interface for the PRoP (see Figure 11.3). The panel computer PRoP interface would act as a “window” into the remote location. An attached camera would transmit the user’s image to the PRoP display. An attached sensor would detect panning and tilting of the screen and in turn move the



pan/tilt PRoP camera. Driving the PRoP would involve using a simple sliding control on the panel. A freeze button would acquire a high-resolution snapshot and allow the user to annotate the image with a pen. Finally, the hypothesis is that gestural I/O mappings will be simplified with sensors directly attached to the body and easily worn such as a watch.

## 11.9 Getting to Know You Experiments

PRoPs are fascinating systems and will inevitably find a place in our social arena. Without definitive user testing of their true social role, we will be unable to make claims about their effectiveness. Much of this experimentation is beginning now, with what we call “getting to know you” experiments. These are modeled after previous user experiments [Connell et al., 2001] using face-to-face, chat rooms, telephones, and video-conferencing. The excitement is that we will be able to have some rough comparison of the PRoP versus these other CMC methods. Of course the PRoP will perform worse in some situations and better for others. But for which, how, and why are the questions that will be the most interesting. By identifying where PRoPs excel, we will also be able to begin drawing conclusions about the role that PRoPs may play in our future society. Experiments are beginning now to test pairs of users as they get to know each other across the PRoP medium. In addition, these users will participate in a prisoners dilemma style game, which will measure trust with the PRoP and its users.

## 11.10 Application Exploration

As fascinating as the PRoP seems, there is still a lack of formal studies of where its primary applications will be found. Will it be tele-teaching? Remote inspection? Tele-collaboration? Tele-work? The application space is quite large but identifying where the PRoP fits is still an active area of research. A few images of some envisioned remote collaboration and inspection tasks are shown in Figures 11.4, 11.5, 11.6, 11.7, and 11.8.

Finally, third generation wireless mobile cellular communication will be a reality in the next few years. These networks will be ubiquitous and, more importantly, capable of transporting the network traffic required by PRoPs. This will place PRoPs directly into our everyday world (see Figure 11.9).



**Figure 11.4:** Remote inspection and troubleshooting over an aircraft repair task (envisionment).





**Figure 11.5:** Discussing a problem with the propeller operation on an aircraft (envisionment).





**Figure 11.6:** Troubleshooting another problem with the aircraft (envisionment).



**Figure 11.7:** An inspector discusses repair procedures (envisionment).





**Figure 11.8:** Training a repairman to perform a critical repair using a PRoP (envisionment).



**Figure 11.9:** A PRoP user has their Tarot cards read on Telegraph Avenue in Berkeley as third generation wireless becomes available greatly increasing the range of PRoPs (envisionment). Photo: Peter Menzel



## CHAPTER 12

# CONCLUSION

*It is necessary to begin at the first page and end at the last. The Author.*  
– *Raymond Roussel in La Doublure*

There is little argument that today's telepresence systems are inferior to live contact in a number of ways. This dissertation has proposed several early steps towards a partial solution to this crisis. Before taking a stand on whether telepresence will be able to provide a sense of intimacy in the future, we first argued that the initial step is to build systems that faithfully support the body's role in face-to-face interaction. Today's systems fail to do this. Most of the flaws are not mysterious. Important cues such as gaze, proxemics, and haptics are being explored by the CMC research community. But they are prototypes, and those systems rarely provide more than one enhancement (one extra cue). Rarely is there an explicit presence or proxy for remote participants, which aggravates the asymmetry with live participants. Once we work past this stage and build CMC systems with a gamut of body-centered behaviors, we will have a much better idea of how well they evoke a sense

of presence.

We began this dissertation by highlighting the importance of the body in human interaction. This claim is so important that we presented the term *tele-embodiment* (see Chapter 3). The body is essential as our means of knowing the world and of knowing others. We have argued that we can best interact with others at a distance by recreating the affordances of our physical body with this tele-embodiment. The research into this dissertation has sought to make the fusion between pilot and robot as direct as between a jet pilot and her aircraft, or a puppeteer and puppet, or a skater and the ice. What arises from this fusion is not a human+robot hybrid, but a new kind of embodied person – a PRoP.

The major contribution of this dissertation is to set the initial groundwork for exploration and development of PRoPs and PRoP-like systems, to initiate and guide their role in a larger human-social setting, and to set the stage for their future applications, forms, interfaces, architectures, and designs. This dissertation details a history of design over several years and systems (see Chapters 4, 6, and 7) towards an understanding of personal tele-embodiment. An existence proof of such systems, if you must. We have undertaken a series of early experiments and tests to validate our claims and further understand the role of PRoPs in daily human life (see Chapter 9). This dissertation work has initiated the discussion of their role, impact, and implications (see Chapter 10). It has established, through real life demonstrations, that these system can and will be a part of our future (see Chapter 11). Finally, we view ourselves as the harbingers of this research with a firm dedication to guiding its focus in these early formidable years. This dissertation exists to

provide that vitally important contribution.

Reflecting back on PRoPs, these new tele-embodied humans, what does this imply about the “self”? Others make judgments about us from their perceptions of our actions. Our bodies communicate far more about us than we can do with text or speech alone. When the medium (*i.e.*, the body) changes, people’s perceptions of us change. Attributes such as dominance, aggression, wisdom, friendliness are all influenced by the body. The internet has already liberated many individuals to explore different identities in text-based VR, changing gender, age and social background. But text-based interaction requires constriction of the self. It requires abstraction and symbolic thought, the use of shared representations and tropes. It cannot be fairly said to be a rich form of embodiment. But the possibility of rich embodiment in new forms does exist with telepresence, and PRoPs are our attempt to explore its full potential.

Our claim remains that PRoPs already 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. More importantly, our initial user tests have demonstrated important feedback as well as promising results concerning the usefulness and usability of PRoPs in our future.



# Bibliography

- [18 U.S.C. 333, 1948] 18 U.S.C. 333 (1948). United States Code, Title 18, Section 333.
- [ActivMedia, 1994] ActivMedia (1994). Activmedia Incorporated Robotics. 46 Concord Street, Peterborough, NH 03458, 603-924-9100, <http://www.activmedia.com>.
- [All and Nourbakhsh, 2000] All, S. and Nourbakhsh, I. (2000). Insect telepresence: Using robotic tele-embodiment to bring insects face-to-face with humans. *Autonomous Robots*.
- [Amir et al., 1995] Amir, E., McCanne, S., and Zhang, H. (1995). An application level video gateway. In *ACM Multimedia*, pages 255–266.
- [Anderson et al., 1997] Anderson, A., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., and ad J. Van Der Velden, A. F. (1997). The impact of VMC on collaborative problem solving. In Finn, K. and adn S. Wilbur, A. S., editors, *Video-Mediated Communication*, pages 133–155. Hillsdale, NJ: Lawrence Erlbaum Associates.
- [Asada et al., 1997] Asada, H. H., Bekey, G., Tarn, T. J., Kazerooni, H., Leifer, L., Picard, R., and Dario, P. (1997). Workshop: Human centered robotics. In *IEEE International Conference on Robotics and Automation*.

- [Asimov, 1950] Asimov, I. (1950). *I, Robot*. Garden City, N.Y. : Doubleday.
- [Bailey, 1989] Bailey, R. (1989). *Human Performance Engineering*. Prentice-Hall.
- [Ballard, 1986] Ballard, R. D. (1986). A last long look at titanic. *National Geographic*, 170(6).
- [Baron, 1998] Baron, N. (1998). Letters by phone or speech by other means: the linguistics of email. *Language and Communication*, 18:133–170.
- [Barrientos et al., 1997] Barrientos, F., Paulos, E., and Canny, J. (1997). Blimp behaviors. In *SIGGRAPH Visual Proceedings*, page 99.
- [Berners-Lee et al., 1992] Berners-Lee, T., Cailliau, R., Groff, J.-F., and Pollerman, B. (1992). World-wide web: The information universe. *Electronic Networking:Research, Applications and Policy*, 1(2), Westport CT:52–58.
- [Brand, 1987] Brand, S. (1987). *The Media Lab: Inventing the future at MIT*. Penguin.
- [Brave and Dahley, 1997] Brave, S. and Dahley, D. (1997). inTouch: a medium for haptic interpersonal communication. In *ACM SIGCHI*, pages 363–364.
- [Burgoon et al., 1996] Burgoon, J., Buller, D., and Woodall, G. (1996). *Nonverbal Communication: The Unspoken Dialogue*. McGraw Hill.
- [Cage, 1952] Cage, J. (1952). 4'33". Composed by John Cage and first performed by pianist David Tudor at Maverick Concert Hall, Woodstock, New York, on August 29, 1952, for an audience supporting the Benefit Artists Welfare Fund.

- [Čapek, 1923] Čapek, K. (1923). *R.U.R (Rossum's Universal Robots)*. Doubleday, Page, and Company. Premier performance in Prague, Czecholovakia, 1921.
- [Čapek and Čapek, 1922] Čapek, K. and Čapek, J. (1922). *The Insect Play (And so ad infinitum)*. Oxford University Press.
- [Carpenter, 1998] Carpenter, A. N. (1998). *Kant's Earliest Solution to the Mind/Body Problem*. PhD thesis, University of California, Berkeley.
- [Center for Metahuman Exploration, 1998] Center for Metahuman Exploration (1998). <http://www.bigsignal.net> (Accessed on 10 December 2000).
- [Clark and Brennan, 1991] Clark, H. and Brennan, S. (1991). Grounding in communication. *Perspectives on socially shared cognition*, pages 127–149. Edited by Resnick, Levine and Teasley, Washington, DC: APA Books.
- [Connell et al., 2001] Connell, J., Mendelsohn, G., Robins, R., and Canny, J. (2001). Don't hang up on the telephone yet! In *ACM Conference on Supporting Group Work (GROUP)*.
- [Connell et al., 1999] Connell, J., Robins, R., and Mendelsohn, G. (1999). The cyberbias: Do people make causal attribution in cyberspace? In *Poster, American Psychological Association*.
- [Cox, 1994] Cox, M. (1994). Robotic telesopes: An interactive exhibit on the world-wide web. In *Second International World Wide Web Conference*.
- [Crisman et al., 1998] Crisman, J., Cleary, M., and Rojas, J. (1998). The deictically controlled wheelchair. *Image and Vision Computing*, 16(4):235–249.

- [Curtis, 1992] Curtis, P. (1992). Mudding: Social phenomena in text-based virtual realities. In *Directions and Implications of Advanced Computing (DIAC) Symposium*.
- [Daft and Lengel, 1991] Daft, R. and Lengel, R. (1991). Organizational information requirements, media richness, and structural design. In *Management Science*, volume 32, pages 554–571.
- [Debevec et al., 1996] Debevec, P., Taylor, C., and Malik, J. (1996). Modeling and rendering architecture from photographs. *SIGGRAPH Conference Proceedings*.
- [Deering, 1991] Deering, S. E. (1991). *Multicast Routing in a Datagram Internetwork*. PhD thesis, Stanford University.
- [Donath et al., 2000] Donath, J., Karahalios, K., and Viegas, F. (2000). Visiphone: Connecting domestic spaces with audio. In *ICAD*.
- [Dourish and Bly, 1992] Dourish, P. and Bly, S. (1992). Portholes: Supporting awareness in distributed work groups. In *ACM SIGCHI*.
- [Draper, 1995] Draper, J. (1995). Teleoperators for advanced manufacturing: applications and human factors challenges. In *International Journal of Human Factors in Manufacturing*, volume 73, pages 457–458.
- [Dreyfus, 2000] Dreyfus, H. (2000). Telepistemology: Descartes' last stand. *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*.
- [Durrett, 1987] Durrett, H. J. (1987). *Color and the computer*. Academic Press.



- [Eckman and Friesen, 1978] Eckman, P. and Friesen, W. (1978). *The Facial Action Coding System*. Consulting Psychologist Press.
- [Eckman et al., 1972] Eckman, P., Friesen, W., and Ellsworth, P. (1972). *Emotion in the Human Face*. Pergamon.
- [Ester, 1990] Ester, M. (1990). Image quality and viewer perception. In *SIGGRAPH 1990 art show*, pages 51–63.
- [Feldman and Rimé, 1991] Feldman, R. S. and Rimé, B. (1991). *Fundamentals of Nonverbal Behavior*. Cambridge University Press.
- [Fischer, 1992] Fischer, C. S. (1992). *America Calling: A Social History of the Telephone to 1940*. University of California Press.
- [Fish et al., 1992] Fish, R. S., Kraut, R. E., and Root, R. W. (1992). Evaluating video as a technology for informal communication. In *ACM SIGCHI*, pages 37–48.
- [Flynn, 1993] Flynn, A. (1993). *Mobile Robots: Inspiration to Implementation*. AK Peters.
- [Fogg and Tseng, 1999] Fogg, B. and Tseng, H. (1999). The elements of computer credibility. In *ACM SIGCHI*, pages 80–87.
- [Gaver et al., 1992] Gaver, W., Moran, T., MacLean, A., Lovstrand, L., Dourish, P., Carter, K., and Buxton, W. (1992). Realizing a video environment: Europarc’s rave system. In *ACM SIGCHI*.
- [Gawdiak, 1999] Gawdiak, Y. O. (1999). Personal satellite assistant.
- [Gelernter, 1992] Gelernter, D. (1992). *Mirror Worlds*. Oxford University Press.

- [Goertz and Thompson, 1954] Goertz, R. and Thompson, R. (1954). Electronically controlled manipulator. *Nucleonics*.
- [Goffman, 1967] Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Doubleday.
- [Goffman, 1981] Goffman, E. (1981). *Forms of Talk*. University of Pennsylvania Press, Philadelphia, Pennsylvania.
- [Goldberg, 2000] Goldberg, K., editor (2000). *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*. MIT Press.
- [Goldberg et al., 1995a] Goldberg, K., Mascha, M., Gentner, S., Rothenberg, N., Sutter, C., and Wiegley, J. (1995a). Robot teleoperation via www. In *International Conference on Robotics and Automation*. IEEE.
- [Goldberg et al., 1995b] Goldberg, K., Santarromana, J., Bekey, G., Gentner, S., Morris, R., Sutter, C., Wiegley, J., and Berger, E. (1995b). The telegarden. In *SIGGRAPH*.
- [Goldberg and Spaid, 1996] Goldberg, K. and Spaid, S. (1996). Telepistemology. *Wired*, 4:232.
- [Goldberg and Wallace, 1993] Goldberg, K. and Wallace, R. (1993). Data dentata. In *SIGGRAPH*.
- [Goodwin, 1981] Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

- [Goodwin, 1986a] Goodwin, C. (1986a). Gesture as a resource for the organization of mutual orientation. In *Seminotica*, volume 62, pages 29–49.
- [Goodwin, 1986b] Goodwin, C. (1986b). Gestures and co-participation in the activity of searching for a word. *Seminotica*, 62:51–75.
- [Gosling and McGilton, 1995] Gosling, J. and McGilton, H. (1995). The java language environment: A white paper. Technical report, Sun Microsystems.
- [Grange et al., 2000] Grange, S., Fong, T., and Baur, C. (2000). Effective vehicle teleoperation on the world wide web. In *IEEE International Conference on Robotics and Automation*.
- [Greaves, 1994] Greaves, A. (1994). State of the art in nuclear telerobotic: focus on the man/machine connection. In *Transactions of the American Nuclear Society*, volume 71, pages 520–521.
- [Green et al., 1995] Green, P., Hill, J., Jensen, J., and Shah, A. (1995). Telepresence surgery. In *IEEE Engineering in Medicine and Biology Magazine*.
- [Greenspan et al., 2000] Greenspan, S., Goldberg, D., Weimer, D., and Basso, A. (2000). Interpersonal trust and common ground in electronically mediated communication. In *ACM Conference on Computer-Supported Cooperative Work*.
- [Hall, 1966] Hall, E. (1966). *The Hidden Dimension*. Doubleday.
- [Hegel, 1979] Hegel, G. (1979). *The Phenomenology of the Spirit*. Oxford University. Translated by J. N. Findlay and A.V. Miller.

- [Hollan and Stornetta, 1992] Hollan, J. and Stornetta, S. (1992). Beyond being there. In *ACM SIGCHI*.
- [Human-Centered Computing, 1998] Human-Centered Computing (1998). Human-centered computing.
- [Hutchinson et al., 1996] Hutchinson, S. A., Hager, G. D., and Corke, P. I. (1996). A tutorial on visual servo control. *IEEE Transactions on Automatic Control*, 12(5):651–670.
- [iRobot, 2000] iRobot (2000). iRobot. 22 McGrath Highway, Somerville, MA 02143, (617) 629-0055, <http://www.irobot.com>.
- [Ishii and Kobayashi, 1992] Ishii, H. and Kobayashi, M. (1992). Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *ACM SIGCHI*.
- [Jouppi, 2000] Jouppi, N. P. (2000). eTravel: Mutually-immersive mobile telepresence project. <http://www.research.compaq.com/wrl/projects/eTravel/ET.html> (Accessed on 10 December 2000).
- [Kaminsky et al., 1999] Kaminsky, M., Dourish, P., Edwards, K., LaMarca, A., and Smith, M. S. I. (1999). Sweetpea: Software tools for programmable embodied agents. In *ACM SIGCHI*.
- [Kant, 1998] Kant, I. (1998). *Critique of Pure Reason*. Cambridge: Cambridge University Press. Translated by Paul Guyer, edited by Paul Guyer and Allen W. Wood.

- [Kaplan et al., 1996] Kaplan, A., Keshav, S., Schryer, N., and Venutolo, J. (1996). An internet accessible telepresence.
- [Kirsch and Kirsch, 1990] Kirsch, J. and Kirsch, R. (1990). Storing art images in intellegent computers. In *Leonardo*, volume 23, pages 99–106.
- [Knapp and Hall, 1972] Knapp, M. and Hall, J. (1972). *Nonverbal Communication in Human Interaction*. Harcourt Brace.
- [Koleva et al., 2000] Koleva, B., Schnädelbach, H., Benford, S., and Greenhalgh, C. (2000). Traversable interfaces between real and virtual worlds. In *ACM SIGCHI*, pages 233–240.
- [Kraut et al., 1994] Kraut, R., Rice, R., Cool, C., and Fish, R. (1994). Life and death of new technology: Task, utility and social influences of the use of a communication medium. In *ACM Conference on Computer-Supported Cooperative Work*.
- [Kuzuoka et al., 1994] Kuzuoka, H., Kosuge, T., and Tanaka, M. (1994). Gesturecam: A video communication system for sympathetic remote collaboration. In *ACM Conference on Computer-Supported Cooperative Work*, pages 35–43.
- [Kuzuoka et al., 2000] Kuzuoka, H., Oyama, S., Suzuki, K., Yamazaki, K., and Mitsuishi, M. (2000). Gestureman: A mobile robot that embodies a remote instructor’s actions. In *ACM Conference on Computer-Supported Cooperative Work*, pages 155–162.
- [Lanier, 2001] Lanier, J. (2001). Virtually there: Three-dimensional tele-immersion may eventually bring the world to your desk. *Scientific American*.

- [Laurel, 1990] Laurel, B. (1990). *The Art of Human Computer Interface Design*. Addison Wesley.
- [Laurel, 1993] Laurel, B. (1993). *Computers as Theatre*. Addison Wesley.
- [Li et al., 2000] Li, F., Gupta, A., Sanocki, E., wei He, L., and Rui, Y. (2000). Browsing digital video. In *ACM SIGCHI*, pages 169–176.
- [Long et al., 2000] Long, C., Landay, J., Rowe, L., and Michiels, J. (2000). Visual similarity of pen gestures. In *ACM SIGCHI*.
- [Lynch, 1991] Lynch, C. (1991). The technologies of electronic imaging. *Journal of the american society for informaion science*, pages 578–585.
- [Mann, 1996] Mann, S. (1996). ‘smart clothing’: Wearable multimedia and ‘personal imaging’ to restore the balance between people and their intelligent environments. In *Proceedings, ACM Multimedia 96*.
- [Mann, 1997] Mann, S. (1997). Smart clothing: The wearable computer and wearcam. *Personal Technologies*. Volume 1, Issue 1.
- [Marvin, 1988] Marvin, C. (1988). *When Old Technologies Were New: Thinking About Electronic Communicaion in the Late Nineteenth Century*. Oxford University Press.
- [McCanne and Jacobson, 1995] McCanne, S. and Jacobson, V. (1995). Vic: A flexible framework for packet video. In *ACM Multimedia*, pages 511–522.
- [McLuhan, 1963] McLuhan, M. (1963). *Understanding Media – The Extensions of Man*. MIT Press.

- [Merleau-Ponty, 1992] Merleau-Ponty, M. (1992). *Phenomenology of Perception*. Routledge.
- [Merz et al., 1997] Merz, M., Froitzheim, K., Schulthess, P., and Wolf, H. (1997). Iterative transmission of media streams. In *ACM Multimedia*, pages 283–290.
- [Milgram, 1974] Milgram, S. (1974). *Obedience to Authority*. Harper and Row.
- [Milgram, 1977] Milgram, S. (1977). *The Individual in a Social World: Essays and Experiments*. McGraw-Hill, Inc.
- [Minsky, 1980] Minsky, M. (1980). Telepresence. *Omni*, 2(9).
- [Mirtich and Canny, 1995] Mirtich, B. and Canny, J. (1995). Impulse-based simulation of rigid bodies. In *Symposium on Interactive 3D Graphics*, New York. ACM Press.
- [Mitra et al., 1996] Mitra, Honda, Y., Matsuda, K., Bell, G., Yu, C., and Marrin, C. (1996). Moving worlds: behaviors for VRML. In *SGI Technical Report*.
- [Moravec, 1988] Moravec, H. (1988). *Mind Children. The Future of Robot and Human Intelligence*. Harvard University Press.
- [Mosher, 1964] Mosher, R. S. (1964). Industrial manipulators. *Scientific American*, 211(4):88–96.
- [Norman, 1992] Norman, D. (1992). *Turn Signals Are the Facial Expressions of Automobiles*. New York: Addison-Wesley.
- [Ntuen et al., 1993] Ntuen, C., Park, E., and Kimm, S. (1993). A blackboard architecture for human-machine interface in mining teleoperation. In *Human Computer Interaction*, volume 1, pages 50–55.

- [O'Malley et al., 1996] O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., and Bruce, V. (1996). Comparison of face-to-face and video mediated interaction. *Interacting with Computers*, 8(2):177–192.
- [Ontario Telepresence Group, 1990] Ontario Telepresence Group (1990). (Accessed on 10 December 2000) <http://www.dgp.toronto.to.edu/tp/>.
- [Open H323 Project, 2000] Open H323 Project (2000). Open h323 project.
- [Packard, 1996] Packard, D. (1996). *The HP Way: How Bill Hewlett and I Built Our Company*. Harperbusiness.
- [Pauline and Paulos, 1997a] Pauline, M. and Paulos, E. (1997a). Further explorations in lethal experimentation. ZKM — Center for Art and Media Technology, Karlsruhe, Germany.
- [Pauline and Paulos, 1997b] Pauline, M. and Paulos, E. (1997b). Increasing the latent period in a system of remote destructibility. InterCommunication Center (ICC/NTT), Tokyo, Japan.
- [Paulos, 1999] Paulos, E. (1999). Workshop: Current challenges in internet robotics. In *IEEE International Conference on Robotics and Automation*.
- [Paulos and Canny, 1995] Paulos, E. and Canny, J. (1995). A world wide web telerobotic remote environment browser. In *Fourth International World Wide Web Conference*.
- [Paulos and Canny, 1996a] Paulos, E. and Canny, J. (1996a). Delivering real reality to



- the world wide web via telerobotics. In *IEEE International Conference on Robotics and Automation*.
- [Paulos and Canny, 1996b] Paulos, E. and Canny, J. (1996b). Space browsers: A tool for ubiquitous tele-embodiment. In *SIGGRAPH Digital Bayou Visual Proceedings*, page 83.
- [Paulos and Canny, 1997] Paulos, E. and Canny, J. (1997). Ubiquitous tele-embodiment: Applications and implications. *International Journal of Human-Computer Studies/Knowledge Acquisition*. Special Issue on Innovative Applications of the World Wide Web.
- [Paulos and Canny, 1998] Paulos, E. and Canny, J. (1998). Prop: Personal roving presence. In *ACM SIGCHI*.
- [Paulos and Canny, 2000] Paulos, E. and Canny, J. (2000). Tele-embodiment and shattered presence: Reconstructing the body for online interaction. *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*.
- [Paulos et al., 2001] Paulos, E., Goldberg, K., and Pescovitz, D. (2001). Tele-actors: Tools for collaborative telepresence. Technical report, University of California, Berkeley (CSD-01-1135).
- [Paulos et al., 1996a] Paulos, E., Pauline, M., Canny, J., and Donith, J. (1996a). Legal tender. In *SIGGRAPH Bridge Art Show Visual Proceedings*, pages 43–44.
- [Paulos et al., 1996b] Paulos, E., Pauline, M., Canny, J., and Donith, J. (1996b). Legal tender. In *DEAF: Dutch Electronics Arts Festival*.

- [PeopleBot, 2000] PeopleBot (2000). ActivMedia Robotics, 44 Concord Street, Peterborough, NH 03458, (603) 924-9100, <http://www.activrobots.com>.
- [Pesce et al., 1994] Pesce, M., Kennard, P., and Parisi, A. (1994). Cyberspace. In *First International WWW Conference*.
- [Reeves and Nass, 1996] Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge Univ Press.
- [Rhodes, 2000] Rhodes, B. (2000). *Just-In-Time Information Retrieval*. PhD thesis, MIT Media Laboratory.
- [Rice, 1992] Rice, R. (1992). Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. *Organization Science*, 3:475–500.
- [Richardson et al., 1994] Richardson, A., Montello, D., and Hegarty, M. (1994). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory and Cognition*, 27(4):741–750.
- [Ringley, 1996] Ringley, J. (1996). Jennicam. <http://www.jennicam.org> (Accessed on 10 December 2000).
- [Rocco, 1998] Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *ACM SIGCHI*, pages 496–502.
- [Rostand, 1923] Rostand, E. (1923). *Cyrano de Bergerac*. The Modern Library.

- [Rubine, 1991] Rubine, D. (1991). Specifying gestures by example. *Computer Graphics in SIGGRAPH*, pages 329–337.
- [Rumi, 1997] Rumi, J. (1997). *The Essential Rumi*. Castle Books.
- [S. Whittaker and O’Connail, 1997] S. Whittaker, S. and O’Connail, B. (1997). The role of vision in face-to-face and mediated communication. In Finn, K. and adn S. Wilbur, A. S., editors, *Video-Mediated Communication*, pages 23–49. Hillsdale, NJ: Lawrence Erlbaum Associates.
- [Saucy and Mondada, 1998] Saucy, P. and Mondada, F. (1998). Khep on the web. <http://www.remotobot.net> (Accessed on 10 December 2000).
- [Scheinman, 1987] Scheinman, V. (1987). Robotworld: A multiple robot vision guided assembly system. In *Proceedings of the 4th International Symposium on Robotics Research*.
- [Schiano et al., 2000] Schiano, D., Ehrlich, S., Rahardja, K., and Sheridan, K. (2000). Face to interface: Facial affect in (hu)man and machine. In *ACM SIGCHI*, pages 193–200.
- [Sellen et al., 1992] Sellen, A., Buxton, B., Arnott, J., Bauersbeld, P., Bennet, J., and Lynch, G. (1992). Using spatial cues to improve videoconferencing. In *ACM SIGCHI*, pages 651–652.
- [Sheridan, 1992] Sheridan, T. B. (1992). *Telerobotics, Automation and Human Supervisory Control*. MIT Press.
- [Shibata and Tanie, 2001] Shibata, T. and Tanie, K. (2001). Mental commit robot. In *IEEE International Conference on Robotics and Automation*.

- [Short et al., 1976] Short, J., Williams, E., and Christie, H. (1976). *The Social Psychology of Telecommunications*. Chicester: Wiley.
- [Siegwart, 1998] Siegwart, R. (1998). Workshop on web robots. In *Intelligent Robots and Systems (IROS)*.
- [Siegwart and Saucy, 1999] Siegwart, R. and Saucy, P. (1999). Interacting mobile robots on the web. In *IEEE International Conference on Robotics and Automation*.
- [Simmons, 1998] Simmons, R. (1998). Xavier: An autonomous mobile robot on the web. In *IEEE/RSJ International Conference on Robots and Systems (IROS)*.
- [Starner, 1999] Starner, T. (1999). *Wearable Computing and Contextual Awareness*. PhD thesis, MIT Media Laboratory.
- [Starner and Pentland, 1995] Starner, T. and Pentland, A. (1995). Visual recognition of american sign language using hidden markov models. In *International Workshop on Automatic Face and Gesture Recognition (IWAAGR)*.
- [Stein, 1998] Stein, M. (1998). Painting on the world wide web: The pumapaint project. In *Intelligent Robots and Systems (IROS)*. <http://yugo.mme.wilkes.edu/villanov/>.
- [Stephenson, 1992] Stephenson, N. (1992). *Snow Crash*. Bantam Books.
- [Steuer, 1992] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 4.
- [Taylor and Trevelyan, 1995] Taylor, K. and Trevelyan, J. (1995). Australia's telerobot on the web. In *International Symposium On Industrial Robots*.

- [Tesla, 1983] Tesla, N. (1983). *My Inventions: The Autobiography of Nikola Tesla*. Hart Brothers.
- [Tissandier, 1884] Tissandier, G. (1884). *Application de L'Electricite: A La Navigation Aeriennne*. Imprimerie Jules Tremblay.
- [Tomovic, 1969] Tomovic, R. (1969). On man-machine control. *Automatica*, 5.
- [Trojan Room Coffee Machine, 1994] Trojan Room Coffee Machine (1994). Trojan room coffee machine. University of Cambridge Computer Laboratory.
- [Turing, 1950] Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.
- [Turkle, 1997] Turkle, S. (1997). *Life on the Screen: Identity in the Age of the Internet*. Simon and Schuster.
- [Turkle, 1998] Turkle, S. (1998). Cyborg babies: From technosex to technotots. *Cyborg Babies and Cy-Dough-Plasm: Ideas about Life in teh Culture of Simulation*.
- [Wallace, 1996] Wallace, R. (1996). A notebook active vision system. *IU Workshop*.
- [Weisbin and Lavery, 1994] Weisbin, C. and Lavery, D. (1994). NASA rover and telerobotics technology program. In *IEEE Conference on Robotics and Automation Magazine*.
- [Weiser, 1991] Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, pages 94–110.
- [Willeke et al., 2001] Willeke, T., Kunz, C., and Nourbakhsh, I. (2001). The history of the mobot museum robot series: An evolutionary study. In *FLAIRS*.

[Williams, 1977] Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84:963–976.

[Yamaashi et al., 1996] Yamaashi, K., Cooperstock, J. R., Narine, T., and Buxton, W. (1996). Beating the limitations of camera-monitor mediated telepresence with extra eyes. In *ACM SIGCHI*.

## APPENDIX A

# PROF HARDWARE INVENTORY

This appendix contains a listing of most of the hardware included in the current PROF construction. There is a significant amount of hardware customization that is not listed. This list is always in flux and will almost certainly be out of date by the time this dissertation is published. Nevertheless, some attempt has been made to enumerate the basic elements of a PROF system.

<b>Item</b>	<b>Vendor</b>	<b>Approx Cost in 2001 USD</b>
Robot Base (Scout II)	Nomadics	\$2500
Single Board Computer (PCM-9574-00A2)	Advantec	\$545
Computer Wiring Kit (PCM-10586-4A00)	Advantec	\$35
Intel Pentium III 600 MHz CPU	various	\$229
128 MB RAM in SO DIMM 144 PC100	various	\$130
Toshiba 6GB hard drive / 2.5 inch EIDE internal	various	\$159
Logitech iTouch wireless keyboard and mouse	various	\$88
WaveLan 802.11b PC 11 Mb/s network PC-Card	Lucent	\$200
WaveLan Range Extender Antenna	Lucent	\$70
Analogue 12.1 inch XGA TFT color display screen	various	\$349
Canon VC-C4 camera	Canon	\$1049
Videum VO PCI video capture card	Winnov	\$149
<b>Total</b>		<b>\$5503</b>





## APPENDIX B

# PRoP SOFTWARE INSTALLATION

This appendix contains information on how to install and configure the software on the actual PRoP. It takes about three hours to perform this entire install. A PRoP user does not perform this install. This is only the installation for the PRoP server software onboard the PRoP.

1. BIOS<sup>1</sup> settings for serial ports should be as follows:

COM Port	Address	Interrupt
COM1	3F8	4
COM2	2F8	3
COM3	3E8	10
COM4	2E8	5

Disable the parallel port entirely in the BIOS.

2. Windows 2000 Workstation CD. Format and Perform basic install on NTFS partition.
3. Choose custom networking options

---

<sup>1</sup>BIOS stands for Basic Input Output System

4. Perform Windows Update for service patch on network using wired LAN
5. Copy 2000 Workstation CD I386 to local disk to allow access after removal of CD-ROM drive
6. Download and install WinZip from [www.download.com](http://www.download.com) to allow unzipping of various programs and drivers
7. Download and install VGA drivers from [www.advantech.com](http://www.advantech.com). More recently, you can also use Windows Update to receive these newer drivers.
8. Download and install audio drivers from [www.esstech.com](http://www.esstech.com). The chipset to install is the ESS1989. Test the setup with Microsoft NetMeeting Audio Tuning Wizard. Set the microphone advanced properties to 20dB gain. This setting is found under the advanced audio properties for the microphone under playback, not recording as expected.
9. Download WaveLan driver from [www.wavelan.com](http://www.wavelan.com). This package contains the drivers and the Client Manager, both of which should be installed.
10. Insert WaveLan wireless network card and insure that the driver installed is from Lucent, not Microsoft.
11. Disable the onboard LAN in BIOS once WaveLan is online.
12. Download capture driver from [www.winnov.com](http://www.winnov.com)
13. Insert capture card into PCI slot then run install program and re-boot.

14. Disable capture server from icon tray and check for Vidum updates using configuration tool. Reboot after updates installed.
15. Download and install jdk1.3 from `java.sun.com`.
16. Download Java Comm package. Move the `dll` file to the `bin` sub-directory of the Java JDK install. Move the `jar` and `properties` files to the `lib` sub-directory.
17. Add `c:\jdk1.3\bin` to the `PATH` environment variable.
18. Add `.` and `c:\jdk1.3\lib\comm.jar` to the `CLASSPATH` environment variable.
19. Install the web server from Control Panel ⇒ Install Programs ⇒ Windows IIS Services
20. Turn off FTP and SMTP in Sys Admin ⇒ Tools ⇒ IIS
21. Change home directory from Sys Admin Tools ⇒ Personal Web Manager to be set to `[desktop]\prop\client`
22. Make sure to set security properties to allow web server to access pages by setting `[desktop]\prop\client` to Everyone/Readable
23. Insure `index.htm` is a valid default document name in the web server settings.
24. Install software into client directory. Copy all client software from its location on the `www.prop.org` web server to `[desktop]\prop\client`. The most recent software may not always be in this location if under development. The most recent copy can always be found with Eric Paulos (`pauloscs.berkeley.edu`).
25. Install software into server directory. Copy all software from its location on the `www.prop.org` web server to `[desktop]\prop\server`.

26. Make shortcut for `java Server` and setup shortcut to autostart when the system boots
27. Make an Emergency Repair Disk in case of problems or disk crash
28. Insure that the VC-C4 camera is set to use full range and 19.2kb/s communication speed. Also, place VC-C4 into VC-C4 communication mode. All of these setting must be done with the VC-C4 camera remote control.

The time for this installation can be saved by performing it once and making multiple copies of the disk image onto other hard drives.

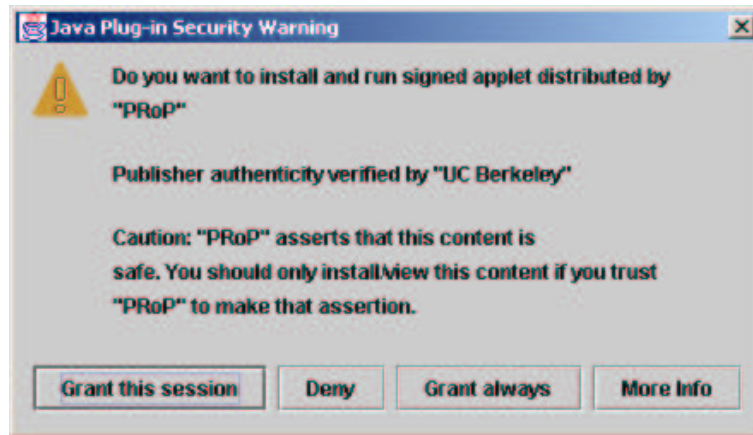
## APPENDIX C

# PROF USER SETUP

This appendix contains information on requirements and setup for a user to prepare to interface to a remote PROF. This information can also be found on the web at the following

URL: <http://www.prop.org/howto>

1. Check that the computer has a full-duplex audio card, headphones (or speakers) and a microphone
2. A video capture card and a camera are also required
3. Install a joystick with at least two buttons and a hat (sometimes also called a point-of-view) controller
4. Check that the internet connection bandwidth is at least 384 Mb/s both up and downstream
5. Check that the most recent version of Internet Explorer (5.0 or later) is installed
6. Check that the most recent version of NetMeeting (3.0 or later) is installed



**Figure C.1:** Java security warning dialog to grant permission to applet

7. Launch NetMeeting and run the Audio Tuning Wizard to check correct audio functionality
8. Check for correct Java Runtime Environment Plug-in 1.3. This can be done by going to the URL at the beginning of this appendix which contains a sample applet to perform this Java test.
9. The PRoP is controlled by the joystick. However, Java does not currently support direct access to the joystick. Therefore, the joystick library for the java code needs to be installed to read the joystick and thus control the PRoP.
10. Download the security certificate from the URL at the beginning of this appendix and install it using the corresponding batch file
11. Download the `joystick.dll` from the URL at the beginning of this appendix and save it in `C:\WINNT\SYSTEM32` (Windows NT and 2000) or `C:\WINDOWS` (Windows 98)
12. When connecting to the PRoP with the control Java applet a window similar to

Figure C.1 will appear. By granting permission the Java applet will be able to read the joystick.

13. Connect to the P<sub>RoP</sub> through a URL provided for the particular session. This URL will be different depending on the access but can always be found through a central location service.





## APPENDIX D

# CREATE AND USE JAVA CERTIFICATES

This appendix contains information on how to generate a security certificate and sign Java code. Signed Java code executing as an applet can request to perform actions not typically permitted by the security mechanisms for Java applets. For the P<sub>RoP</sub> this is mainly to allow the Java code to access the joystick. Java security comes in so many different forms that each browser, and Java version has its own method of handling security. After a great deal of effort the solution presented in this appendix works quite well.

The steps described in this appendix create a self-signed applet. This is useful for testing purposes and for use by small groups. However, because the generated certificate cannot be verified by the embedded certificate authorities in most browsers, it is not recommended for full public release. It is unreasonable to expect typical users to import and trust a self-signed applet from an unknown developer. For final release a “real” certificate issued by an authority like VeriSign or Thawte should be used. Currently, a code signing certificate costs approximately \$700 USD.

1. The applet needs to run in the plugin, as only the plugin is platform and browser independent. And without this independence, it makes no sense to use Java.
2. Create Java code for the applet as usual. It is not necessary to set any permissions or use any security managers within the code.
3. Install JDK 1.3. The path for use of the following commands is [jdk 1.3 path]\bin.  
The password for the keystore is any password.
4. Generate key:

```
keytool -genkey -keyalg rsa -alias PRoP}
Enter keystore password: *****
What is your first and last name?
[Unknown]: Eric Paulos
What is the name of your organizational unit?
[Unknown]: PRoP
What is the name of your organization?
[Unknown]: PRoP
What is the name of your City or Locality?
[Unknown]: Berkeley
What is the name of your State or Province?
[Unknown]: CA
What is the two-letter country code for this unit?
[Unknown]: US
Is CN=Eric Paulos, OU=PRoP, O=PRoP, L=Berkeley, ST=CA, C=US
correct?
[no]: yes
```

(wait...)

```
Enter key password for PRoP
(RETURN if same as keystore password):
```

(press [enter])

5. Export key:

```
keytool -export -alias PRoP -file prop.crt
```

```
Enter keystore password: *****
Certificate stored in file prop.crt
```

6. Create JAR file from collection of class files

```
jar cvf PRoP.jar *.class
```

7. Sign JAR file:

```
jarsigner PRoP.jar PRoP
Enter Passphrase for keystore: *****
```

8. Verify Signing:

```
jarsigner -verify -verbose -certs PRoP.jar
```

```

          1988 Thu Jun 07 16:42:00 PDT 2001 META-INF/MANIFEST.MF
          2041 Thu Jun 07 16:42:00 PDT 2001 META-INF/PROP.SF
           953 Thu Jun 07 16:42:00 PDT 2001 META-INF/PROP.RSA
             0 Thu Jun 07 16:41:58 PDT 2001 META-INF/
smk      597 Fri Jun 01 13:49:20 PDT 2001 Base.class
smk     6644 Fri Jun 01 13:49:20 PDT 2001 Base_Nomadics.class
smk     7903 Fri Jun 01 13:49:20 PDT 2001 Base_Pioneer2AT.class
smk     5730 Fri Jun 01 13:49:20 PDT 2001 Client.class
smk     1689 Fri Jun 01 13:49:20 PDT 2001 ClientBaseData.class
smk     1302 Fri Jun 01 13:49:20 PDT 2001 ClientHeadData.class
smk     1596 Fri Jun 01 13:49:20 PDT 2001 ClientSonar.class
smk     1320 Fri Jun 01 13:49:20 PDT 2001 ClientStatus.class
smk     5613 Fri Jun 01 13:49:22 PDT 2001 FTPClient.class
smk     4122 Fri Jun 01 13:49:22 PDT 2001 FTPControlSocket.class
smk       399 Fri Jun 01 13:49:22 PDT 2001 FTPException.class
smk       585 Fri Jun 01 13:49:22 PDT 2001 FTPTransferType.class
smk     3009 Fri Jun 01 13:49:20 PDT 2001 Hand.class
smk       576 Fri Jun 01 13:49:20 PDT 2001 Head.class
smk     5913 Fri Jun 01 13:49:20 PDT 2001 Head_VCC3.class
smk     6222 Fri Jun 01 13:49:20 PDT 2001 Head_VCC4.class
smk     1034 Fri Jun 01 13:49:20 PDT 2001 Joystick.class
smk     3155 Fri Jun 01 13:49:20 PDT 2001 Locate.class
smk     4517 Fri Jun 01 13:49:20 PDT 2001 Net.class
```

```
smk      156 Fri Jun 01 13:49:20 PDT 2001 NetListener.class
smk     2701 Fri Jun 01 13:49:22 PDT 2001 NetSmoother.class
smk     7088 Fri Jun 01 13:49:20 PDT 2001 PRoP.class
smk      270 Fri Jun 01 13:49:20 PDT 2001 PRoPElement.class
smk     3023 Fri Jun 01 13:49:20 PDT 2001 Serial.class
smk      161 Fri Jun 01 13:49:20 PDT 2001 SerialListener.class
smk      660 Fri Jun 01 13:49:20 PDT 2001 Server.class
smk      292 Fri Jun 01 13:49:22 PDT 2001 TimeStampedData.class
smk    36864 Mon Aug 28 12:57:42 PDT 2000 joystick.dll
```

```
s = signature was verified
m = entry is listed in manifest
k = at least one certificate was found in keystore
i = at least one certificate was found in identity scope
```

jar verified.

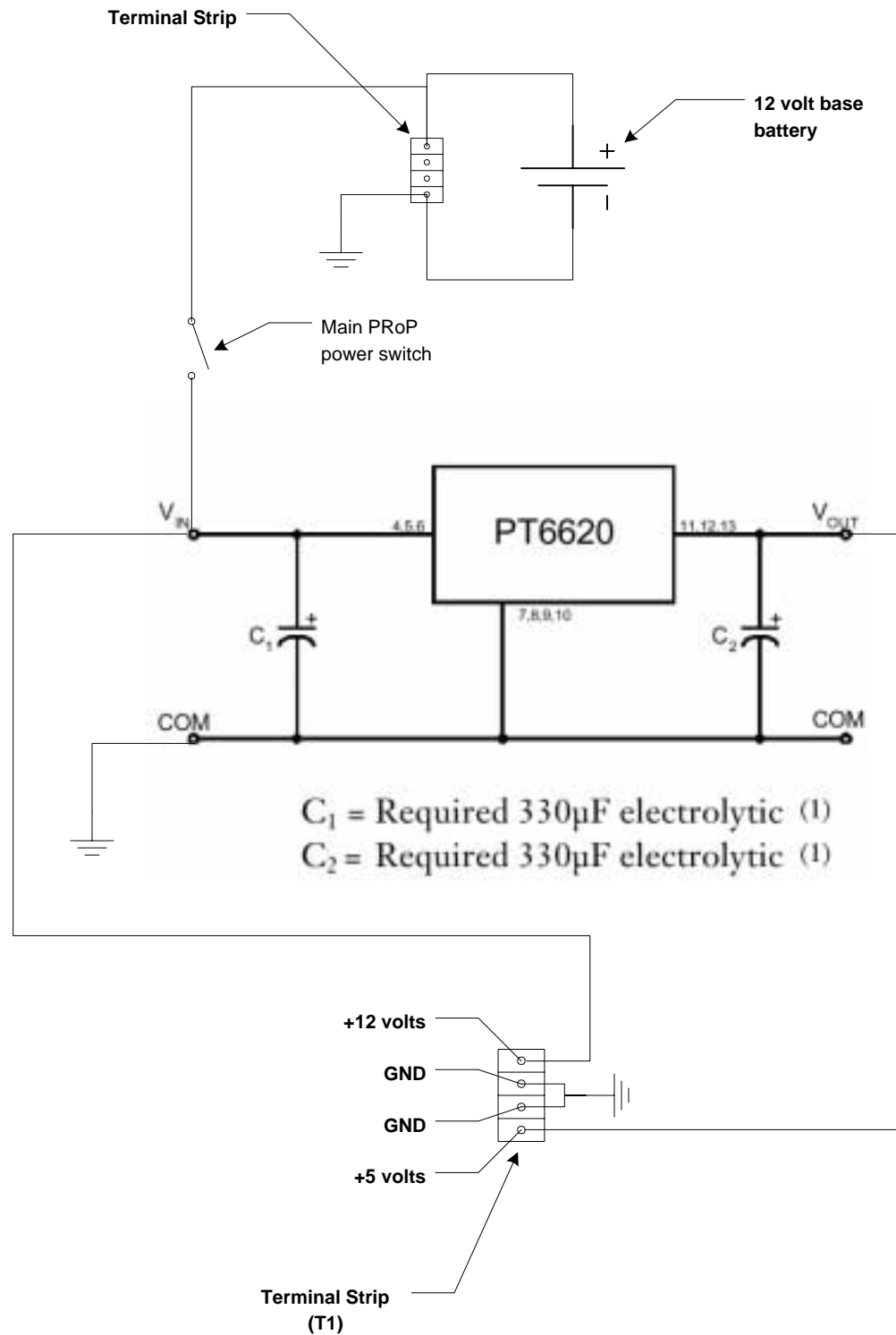
9. Create HTML-File for use of the Applet by the Sun Plugin 1.3 using HTML Converter

Version 1.3 from Sun. This will allow use of the signed Java applet.

## APPENDIX E

# PROP POWER WIRING DIAGRAM

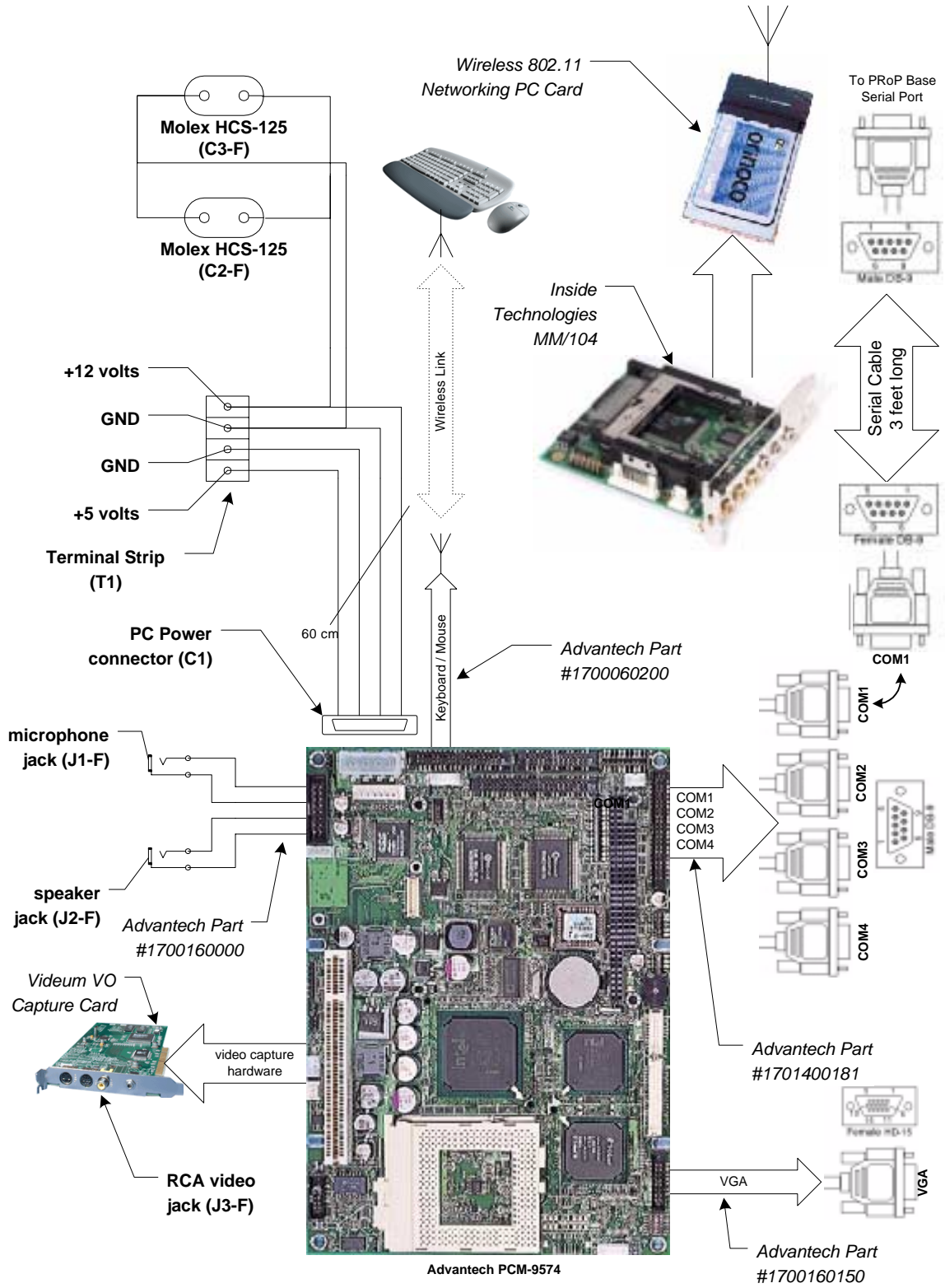
# PRoP Power Wiring Diagram



## APPENDIX F

# PROP BASE WIRING DIAGRAM

# PRoP Base Wiring Diagram

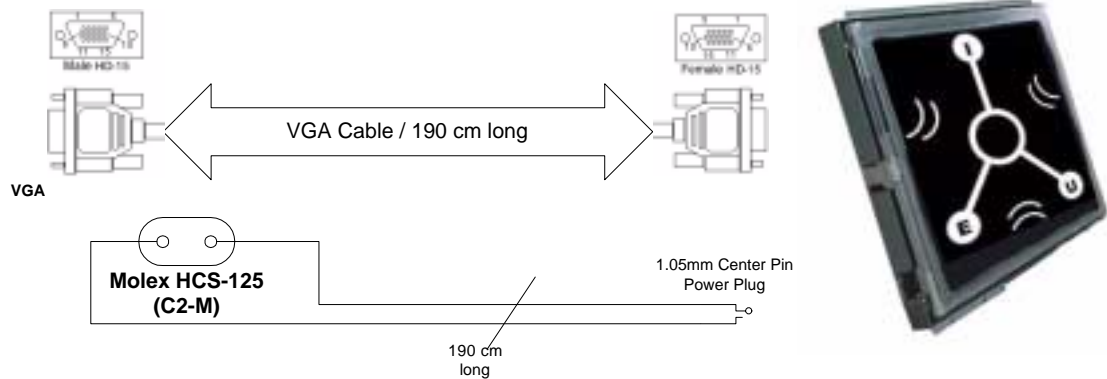
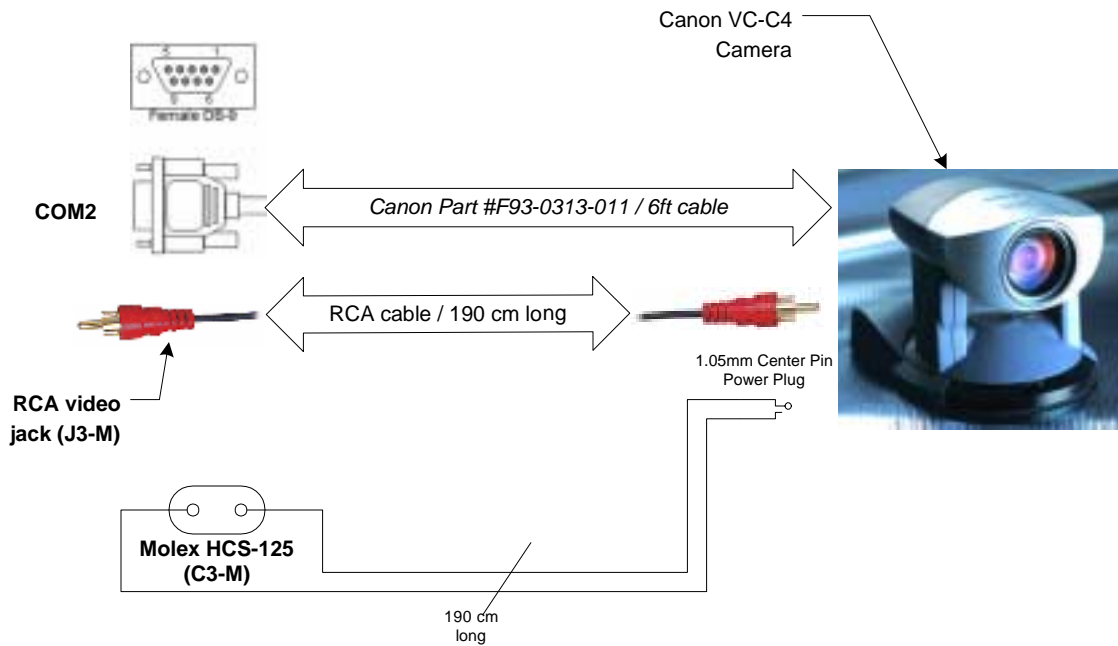
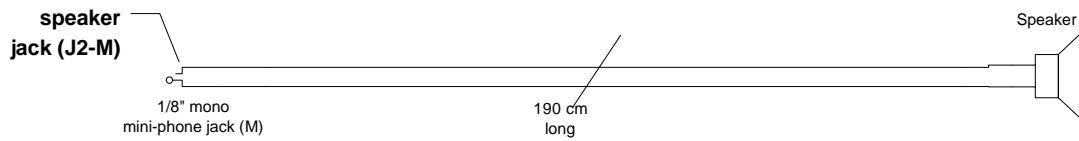
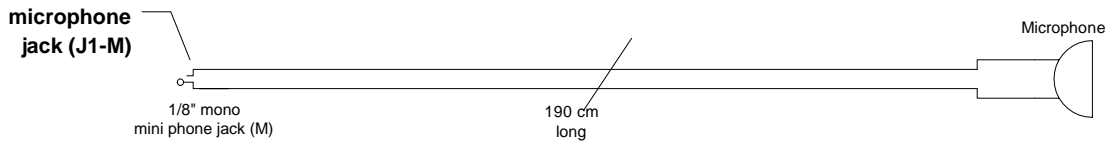




## APPENDIX G

# PROP BODY WIRING DIAGRAM

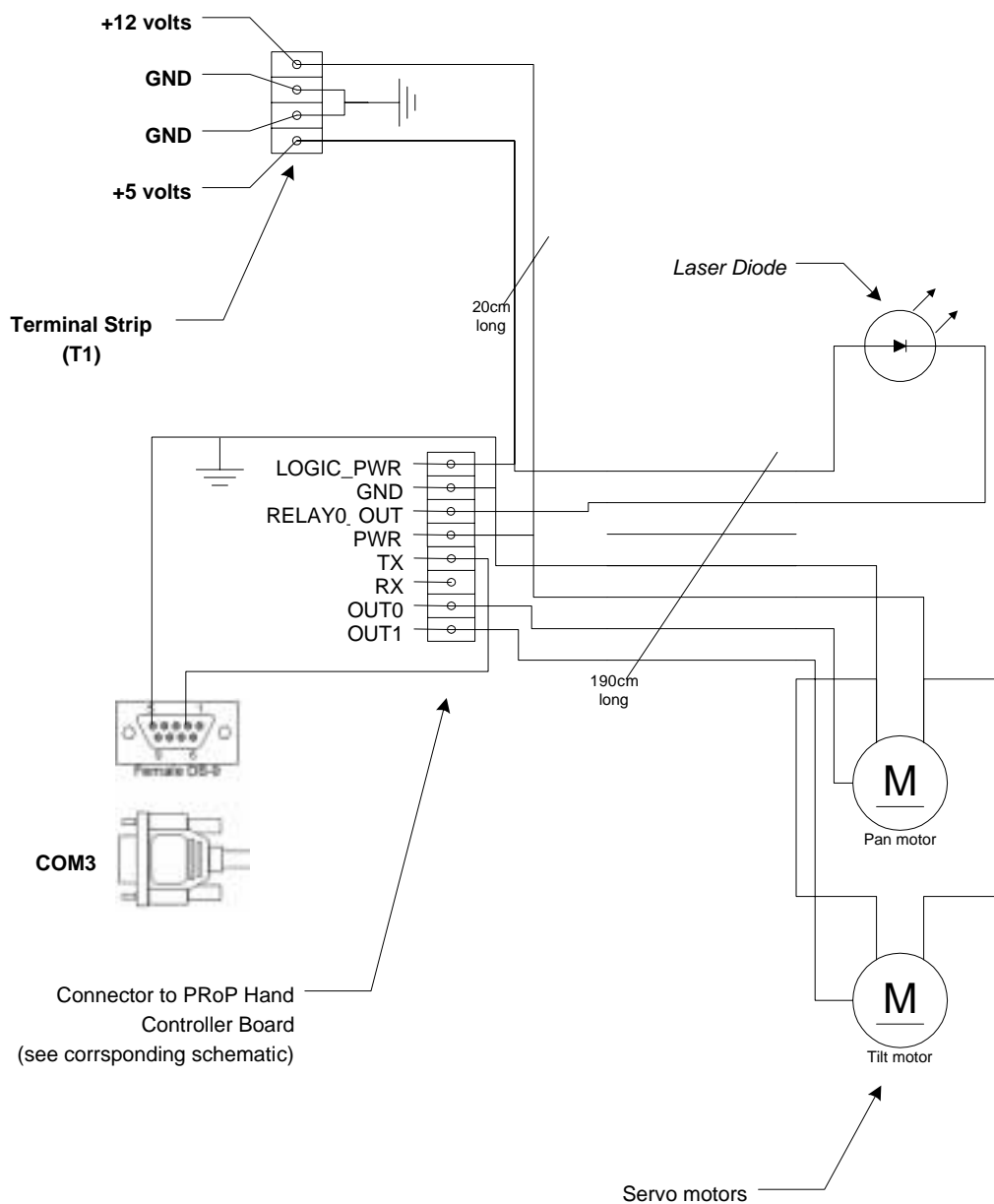
# PRoP Body Wiring Diagram



## APPENDIX H

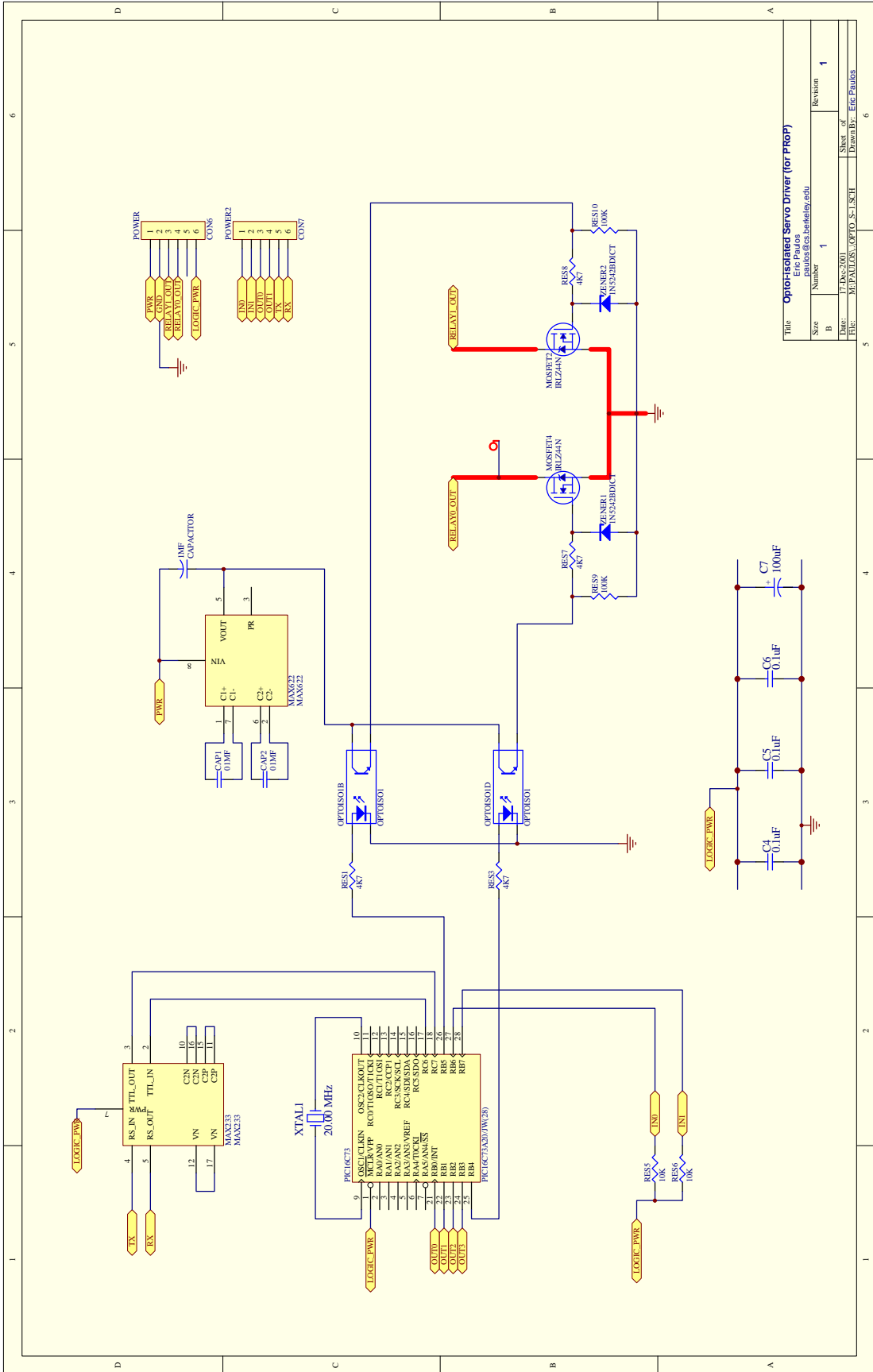
# PROP HAND WIRING DIAGRAM

# PRoP Hand Wiring Diagram



## APPENDIX I

# PROP HAND DRIVER CIRCUIT



Title: **Opto-Relay Servo Driver (for ProSP)**  
 Eric Paulos  
 paulos@cs.hawaii.edu

Size	Number	Revision
B	1	1

Date: 7-18-2001  
 File: MIPALOSLOPTD\_S-1.SCH  
 Sheet of: 6  
 Drawn By: Eric Paulos