Social Catalysts:

enhancing communication in mediated spaces

Kyratso G. Karahalios

S.B., Electrical Engineering, MIT.1995
M.Eng., Electrical Engineering and Computer Science, MIT. 1995
Masters in Media Arts and Science, MIT. 1997

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Doctorate of Philosophy in Media Arts and Sciences
at the Massachusetts Institute of Technology

September 2004

Author _____

Kyratso G. Karahalios
Program in Media Arts and Sciences
June 25, 2004

Certified by: _____

Judith S. Donath
Professor, Program in Media Arts and Sciences
Thesis Supervisor

Accepted by: _____

Andrew B. Lippman
Chairman, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Social Catalysts:

enhancing communication in mediated spaces

Kyratso G. Karahalios

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on June 24, 2004
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

Mediated communication between public spaces is a relatively new concept. One current example of this interaction is video conferencing among people within the same organization. Large scale video-conferencing walls have begun to appear in public or semi-public areas, such as workplace lobbies and kitchens. These connections provide a link via audio and/or video to another public space within the organization.

When placed in public or semi-public work spaces, they are often designed for casual encounters among people within that community. Thus far, communicating via these systems has not met expectations. Some drawbacks to such systems have been lack of privacy, gaze ambiguity, spatial incongruity, and fear of appearing too social in a work environment.

In this thesis we explore a different goal and approach to linking public spaces. We are not creating a substitute for face-to-face interaction, but rather new modes of conversational and physical interaction within this blended space. This is accomplished through the introduction of what we are defining as a social catalyst.

We address the need for designs best suited for linking public spaces and present a series of design criteria for incorporating mediated communication between public and semi-public spaces.

Thesis Supervisor: Judith S. Donath
Professor, Program in Media Arts and Sciences

Social Catalysts:

enhancing communication in mediated spaces

Kyratso G. Karahalios

Doctoral Dissertation Committee

Advisor: _____

Judith S. Donath
Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Thesis Reader: _____

V. Michael Bove
Principal Research Scientist
MIT Media Laboratory

Thesis Reader: _____

Chris Csikszenthmihályi
Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Acknowledgements

During my journey through MIT, I have met many people along the way who have directly or indirectly led me to pursue this work. I want to thank you all.

Starting from Judith Donath — my advisor and mentor. You have helped me find my groove and have always provided insightful perspective in areas of research and life. Mike Bove — who from my beginning in the Garden has always been encouraging. He has this amazing ability to answer every question no matter how obscure. You have always been a pleasure to work with, and I hope we continue to work together. Chris Csikszenthmihályi — Keith Umpteen — you introduced me to the world of ethnography and helped me relate it to my interests. You are an incredible inspiration. Andy Lippman — my introduction into the Media Lab and Garden. You gave me my start at the lab and have been there throughout providing guidance, support, and wit to the experience.

The Sociable Media Group has been my home away from home during the conception and execution of this thesis. I want to thank Fernanda, Lisa, Kelly, Hyun, Raffi, Joey. Especially, Fernanda, a long-time officemate and colleague has been extra supportive in every aspect possible. I value her friendship and her work. Expect great things from her!

Will Glesnes — for your kindness and talent. My sincere apologies for bugging you, and my deepest thanks.

I want to thank my early Garden colleagues. Henry Holtzman, who has helped me since my first day there. I am grateful. Klee Dienes who showed me what is possible. Stefan Agamanolis, who continues to be a good friend and colleague. Finally, Linda Peterson, Pat Solakoff, Brian Spires, Kevin Davis, Veronica, Linda, Derek, Gap, and Bhuwan for helping and smiling. Debra and Carol for keeping me company in the lab, Jesse and Mario — for keeping me company while working in the dorms.

Michael Naimark — working with you at Interval and knowing you has been very inspiring, valuable, and fun!

Paris — your love and unique style of vulcan, emotional support is priceless.

Finally, to my family for believing in me. My brother for sharing time with me at MIT and for grounding me, my parents for letting me leave home to pursue my curiosity — Mom for her almost stubborn believing that I could do anything. My father — who taught me to perservere and to find the humor in the journey — you have guided me and inspired me.

I dedicate this thesis to my father.

Μου έδωσες την αγάπη σου και την δυναμή σου,
Σε ευχαριστώ και σε αγαπώ.

Είσαι πάντα μαζί μου,
Κυράτσω

Table of Contents

CHAPTER 1

Introduction

Figure 1. The meter for social catalysts. My little dude.Through him, I met many labmates, sponsors, and strangers that I would have otherwise not talked to.

In this thesis, we are creating audio-video communication links between remote spaces for sociable and casual interaction. We first address many of the problems with existing audio-video links. We then create site-specific designs that complement these spaces both physically and socially. The main contributions of this work are the design principles for intelligent social interfaces that serve as catalysts to encourage new interactions between people within and between two spaces. We call these social catalysts.

The main idea of the social catalyst is to initiate and create mutual involvement for people to engage in conversation. For example, in a public space, it is not customary to initiate conversation with random strangers. However, there are events which act as catalysts that connect people who would not otherwise be communicating with each other. Such a catalyst may be an experience, a common object like a sculpture or map, or a dramatic event such as a street performer. The chess tables in Harvard Square's outdoor cafe are an interesting example. People flock to this public space for coffee, walks, etc. The chess players usually draw a small crowd from the masses. The act of the game then provides an ice-breaker and lowers the barriers to conversation.

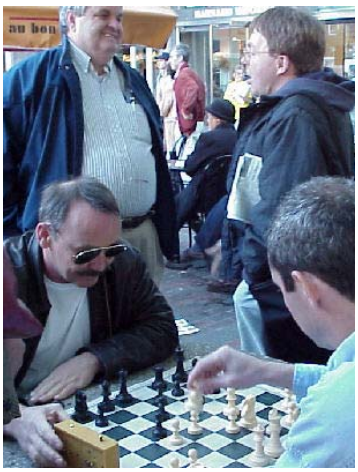


Figure 2. (top) Chess players and observers at Harvard Square. The game serves as a focal point to initiate intersecting among bystanders. (bottom) Kinetic, interactive, audio sculpture inside the Kendall subway station. Rotating levers at the inbound and outbound sides of the station cause the hammers to hit the chimes. The chimes connect to people at both sides of the station. Their timing and communication makes them sound.

Social Catalysts

"*A sign of a great place is triangulation. This is the process by which some external stimulus provides a linkage between people and prompts strangers to talk to each other as if they were not.*"

-William H. Whyte

Our hypothesis is that the creation of a social catalyst as an integral part of the environment will aid mediated communication between spaces by providing a spark to initiate conversation and the interest to sustain it.

The social catalysts of our installations extend Whyte's triangulation principle into the display and interface of the connected space. The form of our social catalyst is abstract, may defy physics, and allows one to interact in ways that are not possible in unmediated communication. It alters the space and communicative cues between the two spaces through the influence of the people in the space and their actions. One such social catalyst might be a connection where current conversation of the users appears as graffiti in the environment. This would allow the occupants to see they are affecting the space and might encourage them to alter it. While the possibilities are infinite, the challenge is determining which agents on the interface are effective as social catalysts and why.

In our linking of two spaces, we are augmenting the appearance of the familiar audio-video wall interface with stimuli that are initiated at either end of the connection. The wall is extended to be not only a display but an event in itself; the system becomes both medium and catalyst. This work further emphasizes the design of the interface as a complement to the space. We want the communication link and display to blend into the physicality and aesthetics of the space and to make the interactions sociable and intuitive.

Structure of this Thesis

We begin in Chapter 2 by examining a brief history of communication between remote spaces. We then discuss features of communication such

as social cues and feedback that are the fundamental elements necessary in communication channels between *connected* sociable spaces for people.

The following four chapters explore social catalysts along four different axes. These axes are:

- Transformation of Space through Interaction
- Abstraction for Visualizing Conversation
- Physical Manifestation of Virtual Presence
- Time and Motion for Blended Spaces

We explain each of these features though an example implementation. In Chapter 3, we transform the shared public space through interaction. We highlight examples of this in the *Telemurals* installation and describe how the social catalysts influenced this interaction. In *Telemurals*, participants can transform the representation of the abstracted space by altering their movement and speaking patterns over time. The design of this interaction was focused on initiating conversation and then enabling the interface to support sustained interaction.

In Chapter 4, we show the power of abstraction for visualizing conversation with the project, *Visiphone*. *Visiphone* is a communication object that connects two spaces via graphics and audio. The audio carried between both spaces directs the video component of the communication object; the graphics visually animate the conversational patterns such as turn taking, conversational dominance, rhythm of conversation, and history (time) between the two spaces. By moving away from traditional waveform audio representations and moving to simple graphical elements, this piece makes salient the social cues in audio conversation and makes the interface accessible to all.

Chapter 5 introduces the concept of a human-scale physical presence coexisting with a virtual medium. We illustrate this with the *Chit Chat Club* installation. *Chit Chat Club* incorporates the ease of use of online chat forums with the physicality and scale of the surrounding physical space in a cafe scenario. We revisit the concept of abstraction to enhance interaction as an alternative to photorealistic face-to-face interaction.

In Chapter 6, we use the elements of time and motion to blend, paint, and explore a shared space. We explain this with the project *Carousel*. *Carousel* is an exploration piece — time and motion transform both of the connected spaces. *Carousel* further shows us the importance of "time" by making the exploration periodic, and thus, infusing the element of seduction into the connected spaces.

Throughout these chapters, we see that the four features above are not mutually exclusive and there is much overlap. We revisit the concepts of transformation, abstraction, physicality and scale, and time and motion over and over. We discuss this in Chapter 7, the conclusion of the thesis.

CHAPTER 2

Bridging Space and Time in Communication

"*To talk to others who are unseen and far away is an experience which, before the telephone, occurred only in mythology. Gods, devils, and angels talked from the sky across the world, but not mere mortals.* "

- Ithiel de Sola Pool

A Brief History

"You've got mail.", The New Yorker

On an April day in 1746 at the grand convent of Carthusians in Paris, about two hundred monks arranged themselves in a long, snaking line. Each monk held one end of a twenty-five-foot iron wire in each hand, connecting him to his neighbor on either side. Together, the monks and their connecting wires formed a line over a mile long.

Once the line was complete, the abbé Jean-Antoine Nollet, a noted French Scientist, took a primitive electrical battery and, without warning, connected it to the line of monks — giving them a powerful electric shock.

(Standage 1998)

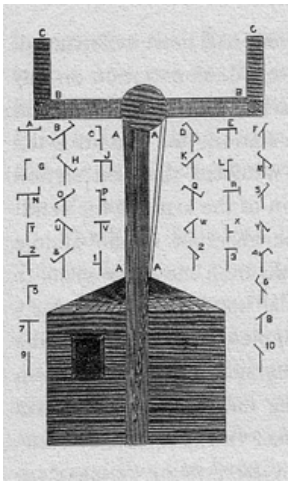


Figure 1. Chappe-style optical telegraph, showing arm positions corresponding to different letters. Mounted on the roof of a tower, the arms were controlled inside by an operator.

Why do this? Nollet was not in the practice of randomly shocking people with electricity. Rather this was one of many experiments to test the possibility, that a signal could be transmitted instantaneously over a mile of connected monks. Since this was the case, in theory, messages could be sent instantaneously over longer distances instantaneously, without the need of a messenger. This was a big deal.

Yet, sending signals over a long distance was unreliable. The medium for sending electricity was not fully understood, resulting in doubt as to whether such long-distance transmission was possible. It is much easier to believe something that can be seen, rather than that which cannot.

Visual cues such as smoke signals, sign languages, and optical signals were used to carry messages over visible distances for thousands of years. These communication systems were becoming intrinsically more complex. In 1793, Claude Chappe sent designs of an optical telegraph to the French National Convention. A trial was devised to construct three towers connecting in Belleville, Écouen, and Saint-Martin-du-Tertre covering approximately twenty miles. It took eleven minutes to send the message, "Danau has arrived here. He announced that the national convention has just authorized his committee of general security to put seals on the papers of the deputies" in one direction—nine minutes to send an equally dull message in the reverse direction. The committee was highly impressed.

This was more desirable than Chappe's previous signalling system that used a series of synchronized clocks and sent the din of clanging pots at expected intervals in accordance with a previously composed code system in the sense that it was not noisy.

Telegraph fever continued and some began to dare and test the limits of electricity once again. The biggest problem of the time was getting signals to travel across very long wires.

In 1832, Samuel F. B. Morse devised a bi-signal code that became dots and dashes and began his attempts at making a permanent record of these signals using the medium of electricity. Four years later, an Englishman, William Fothergille Cooke had a similar revelation and began his own experiments.

It was much harder selling the electrical telegraph to the public. It was obvious how the optical telegraph functioned. The electrical telegraph, in

Samuel Morse's message during the first major exhibition of the telegraph:

"What hath God wrought"

Transcripts from the first transcontinental telegraph link:

"CAN YOU RECEIVE ME?"

"PLEASE SAY IF YOU CAN READ THIS."

Alexander Graham Bell's fabled first words over the telephone:

"Mr. Watson, come here, I want you."

contrast, did not make intuitive movements, it rattled occasionally; the results appeared to be black magic or a hoax.

Despite monetary and government setbacks and claims to invention, the telegraph was slowly emerging in cities. An experimental line was set up between Baltimore and Washington. It wasn't regarded as useful. Occasionally it was used to "watch" chess matches between the best of each town.

A transcontinental line was the next step in testing the limits of the telegraph. On August 5, 1858, after the fourth attempt, the laying of 2,050 miles of cable across the atlantic was completed. On September 1st, it deteriorated.

In 1864, cable successfully connected India and Europe.In 1866, Europe and North America were again connected. This time, it became clear the messaging was not a hoax. The line finally became profitable. It inspired Wall Street, code makers and code-breakers, families, and even marriages over long distances.

Figure 2. Ad appearing in the Antioch Ledger, August 17, 1932. It emphasizes friends, fun, and sociability.

Research into the area of a harmonic telegraph to send several messages concurrently using vibrating reeds at both ends led instead to the invention of the telephone in 1876. The Bell Telephone company advertised their service in 1877. "No skilled operator is required; direct conversation may be had by speech without the intervention of a third person. The communication is much more rapid, the average number of words being transmitted by Morse Sounder being from fifteen to twenty per minute, by Telephone from one to two hundred." 30,000 telephones were being used by 1880 around the world.

The Telephone

Early advertisements for the telephone stressed that "NO SKILL WHATEVER is required in the use of this instrument" for one to use it successfully. It was marketed for use in businesses and originally was used primarily by men. Over time, the telephone appeared in the household and later advertisements read, "friends who are linked by telephone have good times". Today, the phone is used as much for socializing as for conducting business (Sproull and Kiesler) (Fischer 1992)

The telephone had many different social effects

Did the telephone, then "break up...the old practice of visiting"? That is too strong a conclusion. Telephoning probably changed visiting practices moderately during the first half of this century. People forewent some visits that they otherwise would have made, particularly unannounced drop-ins. (The elite practice of "calling" during designated at-home hours, printed card in hand, was probably in decline anyway.) Telephone users altered the character of other visits by telephoning ahead to arrange and confirm them. Finally, people probably made rendezvous they otherwise would not have, particularly appointments at public places, because they could telephone. Thus, telephoning perhaps had a limited effect on visits to another's home but made a greater difference in the ability to arrange an appointment outside the home. Furthermore, different people responded in various ways. Whether the sum total of face-to-face conversations with people outside the household declined because of the telephone we will never know, but it is much likelier that the total volume of social conversation increased notably. The telephone probably meant more talk of all kinds (Fisher 1992).

- it provided timely communication over great distances
- it reduced danger and isolation for those in remote areas
- it reduced the constraints of physical proximity
- it allowed for private conversations
- it gives one the power to intrude or make noise in a foreign home or space

The phone further allowed one to pay attention to different things, have different contacts, and different expectations. Sproull and Kiesler show us that like the telephone, computer mediated communication in some organizations is changing attention, social contact patterns, and interpersonal interactions.

Figure 3. Edison's Telephonoscope (transmits light as well as sound) George Du Maurier, Almanach Punch 1879. Although implementation of two-way audio-video connections emerged in the 20th century, speculation for such interfaces emerged alongside the telephone. Here we see a video window above the mantelpiece in a Victorian villa whereby the parents are communicating with their children. They are holding microphones to speak with them while they watch them on the video screen or telescope.

Arpanet

Packet-switching technologies of the 60's and early 70's made it easier to send messages to and from remote computers. Arpanet, starting in 1969, was created with the intent to enable scientist to remotely share resources such as specialized hardware, databases, and programs. This vision of Arpanet became a reality. What was not anticipated, and was quite a surprise, was that the most popular and most used feature of the Arpanet was electronic mail (Sproull and Kiesler). People could message each other if they were logged on to the same machine or they could leave messages that could be read later. Hence, the beginning of computer-mediated-communication (CMC).

Convergence

The telegraph, the telephone, Arpanet or as it is known today (2004) as the internet, have not changed our basic needs and drives. "Changes in attention, social contact, and interdependencies do not alter human nature of fundamental processes of society. People still fall in love, care about their bosses evaluations, and work for money." (Sproull and Kiesler 1991). Historian George Daniels generalizes this by saying:

> No single invention — and no group of them taken together in isolation from nontechnological elements — ever changed the direction in which a society was going... [Moreover,] the direction in which the society is going determines the nature of its technological innovations...
>
> Habits seem to grow out of other habits far more directly than they do out of gadgets.

The changes in technology, however, highlight our adaptability and our need for communication and interaction. The primary vision for these channels for interaction were work oriented. Nonetheless, people used these connections for casual, sociable interaction. The book, *Anecdotes of the Telegraph*, contains tales of weddings over the telegraph — taking traditional social mores and projecting them onto this new medium to create new ones. These new interactions began to further challenge the law to treat them as binding as if they were in the same geographical location. The telephone further projected the tendency to congregate even at a distance, it coined the term "call girl", and in many ways provided for a level class interaction ground.

Some thought that these new connections such as the car, the telephone, etc. encouraged promiscuity and cut our moral fiber. That they destroyed

Figure 4. inTouch a communication object explores forms of interpersonal communication through touch. Force-feedback technology is employed to create the illusion that people, separated by distance, are interacting with a shared physical object. This is in some ways reminiscent of the telegraph.

space and shrank time. These concerns are rooted in a larger scope to the concept of modernity and the fear, uncertainty, and doubt associated with changes in social organization (Fisher). In *No Sense of Place*, Joshua Meyerowitz states that electronic media "lead to a nearly total dissociation of physical place and social 'place'. When we communicate through telephone, radio, television, or computer, where we are physically no longer determines where and who we are socially."

Yet, we strive to communicate using almost any medium capable of supporting communication cues. The most striking phenomena of these different media become not those that let people work more efficiently, but those that let people communicate and act in ways that were not possible before. This is what we are focusing on as we connect spaces in our interfaces. The internet is still in its infant stage, yet we are still moving forward to create enhanced computer mediated communication. The imagined cartoon in Figure 2 has passed several stages of implementation and testing. We have not yet found an acceptable interface for it. This thesis explores the relationship between the space and communication system to make this happen.

Situating this Dissertation

This work is situated in the study and design of sociable spaces that support interaction and the social cues we emit and perceive in them. We begin by looking at sociable spaces and follow a path that leads us to looking at communication cues and interaction in these spaces.

Looking around at sociable spaces from town squares to office lounges, it becomes apparent that some spaces attract people much more than others, and some remain consistently barren. Why is this? This question lies at the root of this research.

In our quest to create usable, sociable connected spaces, we begin by looking at sociology and urban planning literature and field studies of traditional public spaces. One of the most comprehensive studies on the social use and design of public spaces has been the work of William H. Whyte. We briefly describe his approach and his observations.

We then proceed to look at projects that have linked spaces for communication using audio and video. These fall into two main categories: telecommunication art and computer supported cooperative work. The telecommunication art projects focused more on connected cultural dispersion of the arts: people in disjoint locations performed concerts together, poetry readings were viewed from many different locations. Technologists approached the problem as how to enable people in disjoint spaces to work and collaborate on projects together. We will describe these projects and see how we can relate them to the design of sociable spaces.

This leads us to the fundamental core of this dissertation — social cues — how they are emitted and perceived — how they are transmitted in unmediated spaces, and how they should be transmitted in mediated spaces.

William Whyte: rediscovering the center

> *I am not, heaven forfend going on to argue for places of maximum gregariousness, social directors for plazas. Anomie would be preferable. What I am suggesting, simply, is that we make places friendlier. We know how. In both the design and management of spaces, there are many ways to make it much easier for people to mingle and meet.*
>
> -William H. Whyte

In 1969, William Whyte began a sixteen year observation study of the workability and use of public spaces within New York City and other cities (Whyte). Using time-lapse cameras, 35mm cameras, tele-photo lenses, and interviews, his group documented patterns of traffic and behavior in selected public spaces.

He observed people at street corners, hidden plazas, open plazas, building atriums, market places, alleyways, and mega structures. Within these spaces he documented climate, lighting, density of people, where they stood, sat, and walked, carrying capacity, and public events in the spaces.

One early hypothesis was that light was a key component in the desirability of a public space. He was not wrong - he was mostly surprised that sittable space was a such a more overwhelming factor. The observation that people tend to sit most where there are places to sit may not appear intellectually

ground breaking, and yet, it is so often overlooked! From his observations he concluded that there were seven key features that were necessary for designing a successful public space. These features are: sittable space, street, sun, food, water, trees, and triangulation. Subtler features such as change, personalization, and the importance of street corners have been further observed in frequented public spaces as well (Jacobs 1993).

Of these features, the most relevant and with respect mediated spaces and communication, however, is triangulation. Common triangulation examples are public events, sculptures, magicians. the "know-it-all", etc. This is because by creating a mediated space, we make the space and the interactions in the space the element of the triangulation. All the other Whyte features apply, but triangulation evolves and becomes a more malleable feature that does what triangulation in an unmediated space cannot. By extending this concept into a computer mediated environment, the triangulation now becomes much more abstract and transformable. The interface, and hence, the space, can defy physics — people can grow and shrink as they speak, the interface may spew questions and objects upwards to the people in the space, it may combine both spaces to create a surreal jointly cohabited space, and so on. This deviates from Whyte's triangulation examples that we have so far discussed, and we will from here on refer to these stimuli as social catalysts.

We will later see that this effect is missing in many current audio-video linked spaces. We have briefly addressed some of the physical parameters in designing sociable spaces. Now we will look at how spaces have been connected using audio and video and how these spaces are perceived.

Hole-in-Space

The early seventies showcased a growing number of artists creating telecommunication art. The pieces connected disjoint spaces using telephone lines as a network to send and receive audio, slow-scan images, and telefax messages. In 1977, the first live two-way audio-video satellite connection by artists, "Two-Way Demo" was presented. It connected the east coast and west coast via the US/Canadian Hermes CTS Satellite. Many of the early pieces focused on performance and cultural dispersion of art. I will briefly describe one of them in more detail, Hole-in-Space.

The communication sculpture, Hole-in-Space, was unveiled by video artists Kit Galloway and Sherrie Rabinowitz on November 13, 1980. Hole-in-

Space was a real-time audio and video connection between Lincoln Center in New York City and "The Broadway" department store at Century City in Los Angeles. A person walking past Lincoln Center would hear and see a life size television image of passersby in Century City. The interaction was reciprocal to those standing in Lincoln Center.

The exhibit was very public. It was visible to people outside on the street and sidewalk to create a bizarrely compelling corner. It was accessible to everyone walking in that vicinity and was simple to use for interaction. There were no distractions with multiple technical devices; in fact, the technology was very much hidden from the user.

The staging and timing of this event led to interesting public reaction. It was unannounced and lasted three nights. There was no signage, instruction, or description. Word of mouth provided the initial publicity for the installation. By the second night, people were arranging rendezvous at the site and by the third night there were sing-alongs and dancing.

There are many reasons why this project is interesting. The spontaneity of its appearance one night and its subsequent nightly intervals created a novel event and served as a catalyst. The images of the people onscreen were life-size, unlike the customary miniaturized figures we are accustomed to on television. This similarity in scale suggested interaction similar to face-to-face encounters. Care was taken in designing the space. The screens were integrated into the protected window panelling of the building so as not to appear like a television in a window. The aim was to have a seamless window opening into the other space.

Figure 5. (from left) First two images are snapshots from each portal of the Hole-in-Space installation. The third image is a Hole-in-Space portal at a distance. The display screen appears embedded in the structure of the building facade.

Figure 6. *Telematic Dreaming* by Paul Sermon. Two double beds are connected in different locations. One space is blue-screened while the other is dark. Both spaces have cameras. The bed in the dark space above shows the occupant and a projection of the remote participant in the other bed.

Hole-in-Space is one of the seminal works in telecommucation art. It let people look and speak out across the country, and it provided an unfamiliar power. At first glance, it was a stunning display. There were, however, some complaints from users at the opening. One of these was confusion with symmetry in the display. Others were gaze discrepancy, camera location, and difficulty communicating.

Attempts to create similar audio-video connections today have proved less effective. More work needs to be done to understand the attraction of Hole-in-Space. It only ran for three nights. Due to satellite transmission costs, it was not a twenty-four hour connection that would always be there. Its limited timing in the evenings may have been advantageous. Its setting and time in history most probably played a role. Lincoln center is a very public place where one often expects to see something new and out of the

Project	Description
"Hello" for "the Medium is the Medium Allan Kaprow (1969)	Closed circuit inputs from various areas of Boston: MIT, a hospital, a video-tape library, and the airport were connected with 5 cameras and 27 monitors. The piece combined the elements of randomness and chance with conversation. As people around Boston search to communicate with one another, Kaprow as director would randomly switch links during conversation.
VIDEOPLACE Myron Krueger (1970)	Extracted silhouettes and experimented with scale and alternative settings and scenarios.
Two Way Demo Send/Receive Satellite Network, Centre for New Art Activities, & Franklin St. Arts Centre (1977)	Three day audio-video experimental exchange between San Francisco and New York using satellite transmission.
"Pacific Rim - Slow Scan" Peacesat users group (1980)	Visual communication link between members of the Peacesat users group. There was audience participation at all colocations including Raratonga, Santa Cruz, Wellington, and Vancouver.
"The Conference on Artist's use of Telecommunications" (1980)	Conference was a live two-way video and audio network held at SF MOMA. Using the IP Sharp timesharing network, brought together people in 8 international cities to discuss ideas regarding satellite and slow-scan connections.
"Terminal Consciousness" Roy Ascott (1980)	8 artists in England, Wales, and America connected with portable computer terminals linked to each other and a data bank in California.
"Hole in Space" Galloway & Rabinowitz (1980)	Outdoor storefront audio-video connection between New York and L.A. Connection and performances viewed as an art piece.
"Double Entendre" Douglas Davis (1981)	Performed between the Whitney Museum, NY and the Centre Georges Pompidou, Paris. A performer was in each city. Pre-recorded clips were mixed into the live performance.
"Telesky" Eric Gidney (1982)	Artists simultaneously exchanged sky art images between Paddington Town Hall Sidney and MIT using slow-scan and telephone lines
"Electronic Cafe'84" (1984)	Cafes were connected in five remote locations. People could chat, leave messages, send poetry, and view performances.

TABLE 1. Sample of Telecommunications Art

ordinary; the same experiment done in 1980 may not be as novel or engaging in 2002.

Media Spaces

The original media spaces project was created by researchers at Xerox PARC and the University of Toronto. Miles of cable for audio and video were placed between Palo Alto and Portland to connect offices, conference rooms, and several public spaces within and between the two sites.

One force behind the creation of media spaces was maintaining the culture of collaboration in the labs when people were geographically separated. The goal was to find means to support cross-site work and maintain the necessary social connections (xerox, media spaces).

Media spaces were set up in common areas and offices. There were various media space displays. Some were television screens of various sizes. Others were desktop displays with various configurations (see Figure 7). Modifying access permissions to private spaces was possible using a desktop application.

The media space system differed from picture phones in several ways. The displays in the common areas tended to be active for the entire day unlike the typical phone call which is used in intervals. Media spaces included interfaces for awareness in the remote spaces. Sensors were attached to

Figure 7. (left) shows the original media spaces being used in a conference setting. (right) the Portholes implementation where several are shown concurrently.

doors as switches for office access, connections could be blocked, and video or audio could be turned off.

Many similar projects have evolved since the original media spaces (see Table 2) and now the term media space refers to any environment created using video, audio, and networked computers to support interaction between distributed groups of people. I will briefly outline one of these projects in more detail that focuses on linking public spaces within Microsoft Corporation to further discuss the concepts related to this thesis. I am referring to this project because it is one of the few which has conducted a user study on its system and its observations give clues for improving such interaction.

Linking public spaces at Microsoft. This project linked three kitchens within two microsoft buildings. The chosen kitchen locations were open to the employees and were frequented often throughout the day. They contained vending machines, free soda, coffee, and microwaves. They were open spaces with no doors.

Audio and video connections were made between the three spaces. The video was displayed on a large projection screen in each kitchen. The screen contained four windows: one for each video feed and one to attract users to the screen. The attractor's purpose was twofold: to lengthen kitchen visits and to provide a topic for conversation. CNN became the most common content in the attractor screen.

Figure 8. The Microsoft projection display. A porthole is shown for each kitchen. The fourth porthole is devoted to attracting the attention of those in the kitchen. A microphone and camera are placed on the refrigerator. Below the display, signage is placed for instructional use.

For privacy reasons, a large OFF was placed as one entered the kitchen. Enabling the button cut the transmission for twenty seconds. ON and OFF buttons were also available within the kitchen that could override the outer switch.

Initial reactions to the system led to a few quick changes. The OFF button duration lengthened to 180 seconds. The CNN audio was muted and closed captioning was used instead to mitigate the noise pollution in the space.

The audio proved to be the biggest technical challenge in the installation. Full-duplex audio best caters to informal conversation patterns. The audio quality and rate had to be maintained at multiple end points and this was difficult. There were echoes in the system due to the microphones capturing audio from the speakers. This is often easier to minimize in smaller

environments where there is less movement and the space can be explicitly controlled. In this situation, echo-cancelling speakers were tried and finally software echo-cancelling. This was and still is a difficult problem in many audio-video transmitting systems.

User experience was varied. The cameras were often disabled despite the privacy buttons. People felt their activity was compromised by the presence of cameras. There were issues of trust: were the camera and audio really off? Lack of reciprocity was also noted: people outside the camera's view could still see the other spaces. Camera placement proved to be another challenge. It is difficult to position a camera in the center of the screen, when there is light from a projector in its path. System observations noted that sustained conversations were rare. The reason for this is uncertain. It may

Project	Description
Picture Phone AT&T (1962)	Picture phone at the Seattle World's fair in 1962. They were presented in dispersed booths across the fair.
Media Spaces Xerox PARC (1985-1988)	Audio-video link between Palo Alto and Portland.
Telecollaboration US West (1989)	Audio-video link between Denver and Boulder, CO. Created call and lookaround functions and remotely controlled cameras.
VideoWindow Bellcore (1988)	Connected two public areas on different floors within one building with big screens.
Cruiser Bellcore (1990)	Examined concept of cruising past offices and seeing activity in each. There was reciprocity in that people could see who was cruising by their office.
Montage Sun (1994)	Similar to Cruiser. Added "peeking into offices" and shared calendars and email.
RAVE Rand Xerox EuroPARC (1993)	Connected every office and common area of three floors of the building. Provided modes such as glance video phone, background, sharing offices along with subtle audio cues to indicate state.
CAVECAT/Telepresence University of Toronto (1991)	Built using the RAVE framework. Added shared drawing. Included the Hydra and Telepresence studies.
Polyscope/Portholes Xerox PARC (1992)	Connection between PARC and EUROPARC. Many connections could be viewed at once in an image array to provide passive awareness.
Teamworkstation/Clearboard NTT (1991-1994)	Allowed for two remote people to share a drawing surface with a gaze awareness model.
Microsoft's Kitchen Links (2001)	Linked public kitchens between two buildings. Provided an additional common window for video content.
ICOM MIT Media Lab (2000)	Connects several areas in the Media Lab and Media Lab Europe to foster community awareness.

TABLE 2. Media Space Projects

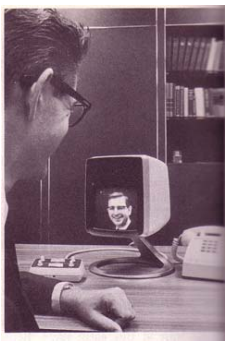


Figure 9. Photo from Bell Labs: The Picturephone: "A completely new video environment and life-style". It was not a marketing success

be the audio quality. It may be that there is little incentive to visit this space. Food was freely available before; perhaps a novel, tangible reward would evoke more interaction. There were also concerns of covert surveillance. Perhaps a later generation familiar with the ubiquitous use of cameras and camcorders will behave differently.

Still others responded with a sense of humor decorating the spaces with unrealistic creatures. A spoof of the audio-video connection was later placed in a men's room. This suggests that adding a playful or gamelike element to the system might make it less intimidating.

This interface provides very little in terms of a social catalyst. It is straight video and audio. Just as Hole-in-Space was engaging as a time-limited event, this may be interesting for the first hour. When it is on for days on end, it may lose it's appeal. A window into another space such as this is so distancing that it needs something to draw the occupants of the space towards it. The intended catalyst in the Microsoft system, the television feed, had the opposite of the intended effect. It was meant to provide content for conversation between the two spaces. Instead, it took the focus away from the conversation. What we need is a catalyst that exists within the focus and space of the interaction.

Moving Beyond Face-to-Face

In the examples that we have seen in the previous section, the interfaces and spaces attempted to recreate the feeling and spatial orientation of face-to-face interaction. This dissertation argues that we have been limiting ourselves by maintaining the metaphor of face to face interaction in computer mediated spaces. There is a lot to be learned from this metaphor, however, it has been taken to extremes.

Why mimic what is already perfect? The human face is the ideal interface. We have thousands of years of evolution in our favor for interpreting it. Since we are using a computing medium, let us take advantage of that and infuse into the channel information which is not evident on the human face.

Components of Communication

1. The *genre* or type of event (e.g. joke, story)
2. The *topic*, or referential focus
3. The *purpose* or *function*
4. The *setting*, including location, time of day, physical attributes of space, etc.
5. The *key*, or emotional tone of the event
6. The *participants*, including their age, sex, ethnicity, social status, etc.
7. The *message form*, including both vocal and nonvocal channels, and the nature of the code (which language, which variety)
8. The *message content*, or what is communicated about
9. The *act sequence*, or ordering of communicative/speech acts, including turn-taking and overlap phenomena
10. The *rules for interaction*, or what properties should be observed.
11. The *norms of interpretation*, including the common knowledge, the relevant cultural presuppositions, or shared understandings, which allow particular inferences to be drawn about what is to be taken literally, what discounted etc.

(Saville-Troike 2003)

Social cues and their medium

In their keynote ACM paper, 'Beyond Being There', Holland and Stornetta describe a few characteristics of information rich channels in face-to-face interaction:

- cue variety
- feedback
- message personalization
- simultaneously being reminded of a need to talk to someone
- having a communication channel
- turn-taking, repair, stylized openings

There are various channels and back channels for communication where these features pass. As with face-to-face interaction, in computer-mediated-communication, we have various communication channels. We believe the above features mentioned by Holland and Stornetta are just as crucial in computer-mediated-communication as in face-to-face interaction. We should, therefore, not alter the cues, but the interface for transmitting and manifesting these cues. At issue is making the channels capable of not only transmitting letters of codes, but also social information about the interaction. This approach we suggest is the following:

- Explore which social cues to transmit between mediated spaces
- Sense these cues (gaze, agreement, etc.)
- Visualize the cues, social patterns, and feedback
- Incorporate the cues and media space into catalysts for interaction

This leads us now to evaluation of computer-mediated-communication. Computer-mediated-communication is fundamentally different from face-to-face communication. To date, we have been evaluating them using the same metrics. In keeping in mind, that these two mediums are different, we should also keep in mind, that we should begin developing different metrics for evaluating them.

From Social Cues to Social Catalysts

So where do we go from here? Our approach in creating intuitive and useful communication spaces is to extract social cues from the connected spaces and transmit them to both local and remote spaces through visualization. The act of instantaneously visualizing these social cues provides feedback and the ability to alter the combined space by interacting within in. The composition of space and interaction interface with incorporated social cues provides a social catalyst to support and encourage interaction.

Social Cues in Interaction

CHANNEL

CODE		Vocal	Nonvocal
	Verbal	Spoken language	Written language (Deaf) Sign Language Whistle/drum languages Morse code
	Nonverbal	Para linguistic and prosodic features Laughter	Silence Kinesics Proxemics Eye behavior Pictures and cartoons

Figure 10. Four-way distinction of message form. (Saville-Troike 2003)

Social cues[1] are perceived by our senses and provide for framing a context about our space, situation, and our interactions. For example, if someone speaks to you in a dialect uncommon in that area, one might assume they did not originate from that place. If one starts screaming loudly, and they are bleeding, it is logical to assume they are in pain.

Conversation cues can be even subtler. Oftentimes, we don't even realize that we perceive them. If two people are having a conversation, and one speaks while raising their pitch at the end of their phrase, one might

assume he or she asked a question. This would then be a cue for the second person to begin speaking and perhaps address that particular question. Another example is sentence repair. As two people are speaking, if one notices the other grimace in disapproval, they may compete their sentence in a different manner already anticipating the opinion of the other person. The cues for turn-taking in conversation are so ingrained in us, that if there is a lull in the conversation, we infer that something is amiss.

Interpreting these social cues helps situate us by providing additional contextual information. From them we can gleam perceptions of who is important, what is legitimate, and who is prestigious (Sproull and Kiesler 1991).

Social cues are transmitted on several different channels. Some of the less obvious are referred to as back channels. These allow for self-disclosure, impression formation, and cultural inferences. These channels are more easily processed when one is in the proximity of others. Some of these back channels are lost as people become more geographically distant. Our goal with mediated interfaces is not to recreate these back channels, but to create a new depiction — perhaps, not as traditional — of the back channels and the resulting interaction.

The Effect of Visualizing Social Cues

People send cues through their presence and actions and receive similar cues from the other end of the channel. There are effective and ineffective manners in which to present these social cues and channels. Our approach

1. Hauser makes a technical distinction between *cues* and *signals*. "Cues, like signals, represent potential sources of information. Cues, however, differ from signals in two important ways. First, cues tend to be permanently ON, whereas signals are more plastic and can be in and ON and OFF state. As a result, signals, but not cues, are produced in response to socioecologically relevant and temporally varying changes in environment. Second, cues typically correspond to an individual's or species' phenotype, and their expression carries no immediate extra cost. Signals, although individual- and species-specific, are associated with significant costs of expression." In this work, I am not making this distinction and am grouping signals and cues into the same category.

is not to interpret the meaning of these cues or weighted sum of cues, but to create a visual language that represents the cues, and allows people to interpret them as they will. We don't want to distort any facts or present erroneous information. Imagine what damage a 'lie detector' would have on a negotiation if it was always portraying false positives. The interpretations may change with use over time, and may be perceived differently by different cultures.

The ability to sense presence and to transform a space by altering the visuals is a strong catalyst to using that space as a medium for interaction. Similarly, a common language of social cues transmitted properly can stimulate conversation to create a common ground and interaction.

Summary

I assert that we have a basic need for social interaction. We have the ability to adapt and massage communication channels to fit our need. The telephone is a perfect example of this. It was considered a business tool — frivolous for use in the home — and yet its social uses prevailed.

Computer-mediated-communication is a larger and less organized domain for study. Text-based channels such as email, IRC (Internet Relay Chat), and Usenet have their audiences and have created their own language and etiquette protocols for interaction. Video-mediated-communication has thus far not met expectations. This may be due to technology limitations. Research in the specific area of choosing video-mediated interaction in lieu of face-to-face suggests this is not the case (Holland and Stornetta 1992).

The goal then has been wrong. Perhaps, we should not be literally mapping computer-mediated systems to face-to-face interaction. These "old-school" systems disregard the social setting they inhabit, they rely on social cues that are expected in physical face-to-face interaction, and ignore our need for social interaction, instead focusing on task-oriented interaction. In short, they filter away social catalysts that attract and suggest interaction. With this work, we suggest a means for adding this missing element into mediated communication.

This dissertation expands social spaces to include computer mediated social spaces. It addresses what features are necessary to sustain a healthy connected sociable space across two remote spaces. We have looked at some features that help create physical spaces through the work of William Whyte, and specifically stressed the use of a social catalyst. We then addressed some of the advantages and limitations of audio-video linked spaces from the perspective of telecommunications art and media spaces. Finally, we examined some of the features that are necessary for sustaining interaction over a communication channel.

With this work, we aim to take what we see as advantages in current computer-mediated-spaces, such as the low barrier, and low risk in text-based systems such as email and Usenet, and combine it with the affordances of physical space. In sum, we want go from interfaces that look like this,

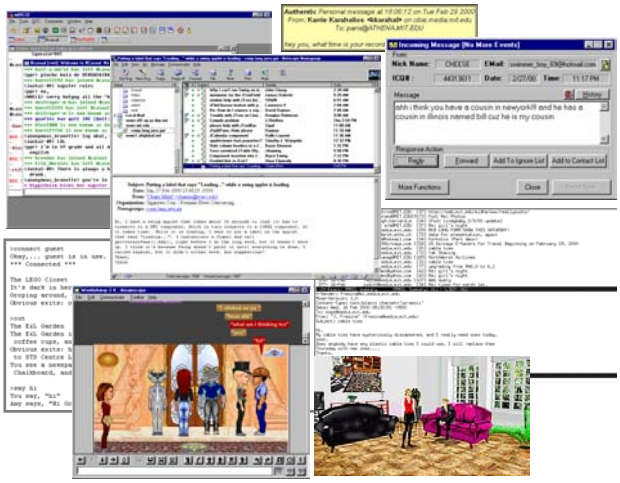


Figure 11. (clockwise from top left) Snapshots of an *IRC* reader, *Usenet* reader, *ICQ* client, email reader, a 3-dimensional virtual world, the two and a half-dimensional world, *Avaterra*, and, the MUD, *MOOSE Crossing*.

and create interfaces that look like this:

Figure 12. (clockwise form top left) Snapshots of *AgoraPhone*, *Visiphone*, *Telemurals*, the Orlando avatar seat of *ChitChatClub*, and the Slim avatar seat with guests in the *ChitChatClub*.

This thesis work bridges the space between the computer mediated communication and the study of communication channels to create a connected space for casual, sociable conversation.

CHAPTER 3

Transformation of Space through Interaction

This chapter focuses on transformation of space — meaning that the participants alter the space by simply being there, moving, or interacting. To explore and discuss these concepts, we will do so in the context of the project, *Telemurals*, which was designed with these features in mind.

Telemurals is a project that connects two lobbies of graduate student dormitories on the MIT campus. One of the dormitories is the newest graduate dormitory on campus, Sindey-Pacific, and the other is the oldest graduate dormitory on campus, Ashdown. Sidney-Pacific approached us with a public art fund to create a space that would encourage students to gather. The dorm was new and students were not taking advantage of the available public space. Similarly, Ashdown was undergoing renovations and wanted to experiment with gathering spaces. Both dormitories were open to the idea of connecting two public spaces within the dorm, and thus, a two-year collaboration was born.

The Process of Creating Telemurals

We began the process of creating *Telemurals* by meeting with the dorm committees of the two dormitories. After several meetings, we decided to connect the

two dormitories during the respective study-breaks of both dormitories on Wednesday and Thursday.

The next step was to choose a location in each dormitory. We chose locations that were in the path of traffic and were in public spaces within the dorm, yet did not attract passersby to stay and talk to people in that space.

Already we had differences between the two dormitories — in style and in occupant personality. Ashdown was older. It used to be a hotel and had a very traditional feel with wood accents and marble staircases. Sidney-Pacific was new and the decor was spartan. In Ashdown, the more established of the dorms, students stopped and talked to us while we were installing cameras. In Sidney-Pacific, very few people talked to us or to each other. This was a very large dorm that was not filled to capacity with very few students knowing each other.

Installing Cameras, Projectors, Microphones, and Speakers

We installed cameras and projectors in the two chosen locations of the respective dormitories. The first test captured audio and video from one dorm, projected them to the other dorm and vice versa.[1] Signage was posted at all possible entrances to the spaces with cameras to notify the inhabitants of the cameras and microphones in the space.

During initial immersed observations and birds-eye view observations, it was noted that there was a tendency for students to walk around the space as opposed to approaching the camera. We had not reproduced the *Hole-in-Space* scenario (see Chapter 2, p. 21). This avoidance behavior observed in both dormitories, however, was consistent with the behavior of participants in the *Microsoft Virtual Kitchen* installation (see Chapter 2, p. 26). This connection might be viewed as an intrusion to people's privacy. Both the Microsoft and our installation were in public spaces, yet they were indoors where people might expect more privacy.

1. Predating the installation process was a two-month interaction with COUHES (Committee On the Use of Humans as Experimental Subjects). We were required to put up signs whenever cameras and microphones were present.

Figure 1. (left) Scott McCloud describes cartooning as a form of amplification through simplification. (right) Snapshot from the movie *Waking Life*. The rendering styles varied throughout the movie; they varied to reflect changes is intensity and content. Note: both of these renderings were make with human assistance. It is far more challenging to create such abstractions using an automated system.

We then delved into the world of cartooning and decided to use it as a means to transform people through abstraction. Our hypothesis was that if people are not entirely recognizable, they might not fear they are losing their privacy.

Transformation and Abstraction

> *When we **abstract** an image through cartooning, we're not so much **eliminating** details as we are **focusing** on **specific details**. By **stripping down** an image to its essential **"meaning"**, and artist can **amplify** that meaning in a way that realistic art **can't**.*
>
> *-Scott McCloud (see Figure 1)*

Inspired by Scott McCloud (see Figure 1), I wrote a real-time algorithm to render everything within the field of view of the camera into a cartoon. The result was as if observing a moving cartoon of the other space (see Figure 2).

During observations, people were moving closer to the camera than in the previous installation. The space was amusing, playful, and, I believe, did not have the surveillance-like effect of the photorealistic video connection. The simpler representation of the cartoon world helps us to better identify with ourselves. This is one of the primary fascinations of cartoons we have when we are young (McCloud 1993).

Figure 2. First abstracted rendering. The effect was of having a moving comic of the connected space. The lower left image is in one space and the upper right image is in the other.

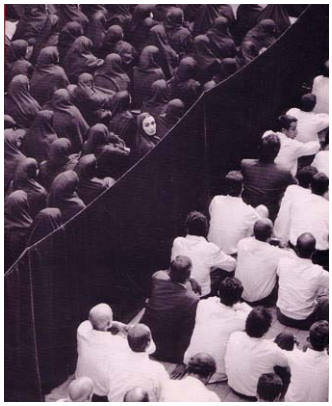


Figure 3. *Fervor*, two-screen projection film by Shirin Neshat. The screens were placed side-by-side. Each loop captured one facet of a narrative from its own angle — two loops, side-by-side, are happening simultaneously. We see a "his" and a "hers" space. "Befitting a narrative of romantic attraction, *Fervor's* adjacent (rather than opposing) screens allow viewers to imagine the would-be suitors as occupying almost — but not yet — shared social space." (Horrigan 2000) This artistic piece is almost the inverse of *Telemurals*. The inhabitants are not in distant spaces, but they are separated. The schism in perspective is represented by the two screens and the curtain separating the two spaces. They want to get communicate, but cannot.

The feedback we received from the second iteration was *far* more favorable than with the straight video. It was playful, and not only drew people into the space, but also encouraged them to spend some time there. This interface did, however spawn some new requests. The students wanted to see themselves as well as the participants at the other end.

In retrospect, this made sense. When one has a caricature drawn of them, the desire to see the resulting representation is strong. This connection was rendering inhabitants in a different manner that caused a reaction, but they did not know how they appeared at the other end or how their actions altered their appearance.

One solution to this request would be to use a picture-in-picture scenario as is done with many video-conferencing systems. We felt that this would emphasize the remoteness of the spaces as opposed to bringing them together. This led us to discover the importance of the framing of the images in our projections.

"The frame as we know it today developed during the Renaissance from the façade-like construction of the lintels and pilasters that surrounded the altarpieces. As pictorial space emancipated itself from the wall and created deep vistas, a clear visual distinction became necessary between the physical space of the room and the world of the picture. This world came to be conceived as boundless — not only in depth, but also laterally — so that the edges of the picture designated the end of the composition, but not the end of the represented space. The frame was thought of as a window, through which the observer peeped into an outer world, confined by the opening of the peephole but unbounded in itself." (Arnheim 1974)

The Window Frame Hypothesis

In video-mediated communication, framing the remote space, whether as one image or as two frames (e.g. picture-in-picture or side-by-side) further emphasizes the distance between the two spaces and makes the viewer feel confined by that space. This changes the space of the interaction and makes things appear even more remote.

Our approach to mitigate this effect is to blend the two spaces together as was done in the Reflection of Presence project (see Figure 4) and to eliminate the structure of the frame. The silhouettes of both *Telemural* spaces are blended into one single image. We eliminate the frame by projecting onto a wall of the space and not onto a projection screen or a physically bounded structure. The background of the projection is black such that no boundary can be seen — only silhouettes of people.

Telemurals

Figure 5. Snapshot of *Telemural* projection. The orange silhouettes are inhabitants of that local space. The red silhouettes are inhabitants of the remote connected space.

Figure 4. Snapshot of *The Reflection of Presence* project at the MIT Media Lab. In this audio-video connection, three remote participants appear on the same frame. The interface featured customizable backgrounds and conversational dominance through the visual prominence of the speaker.

Figure 5 is a snapshot of the blended *Telemurals*. Silhouettes of the participants in the local space are rendered in orange. The participants at the

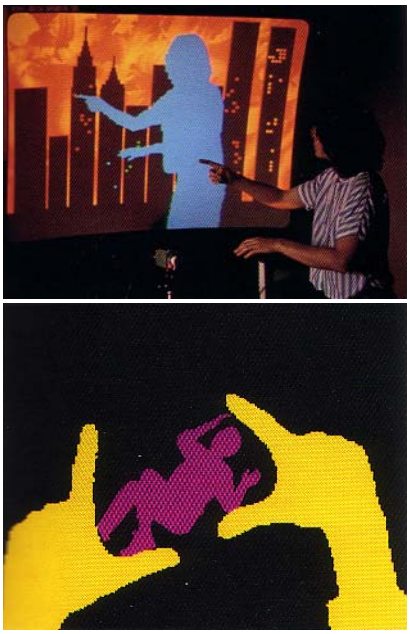


Figure 6. (top) shot from Kreuger's *Painting the Town.* User interacts with New York skyline. (bottom) *VideoDesk* participant captured image of *VideoPlace* participant.

remote end are rendered in red. When they overlap, that region becomes yellow. The aim of this cartoon-like rendering is to transmit certain cues such as number of participants and activity level while preserving privacy by not initially revealing too many social cues about the identity of the participants.

The renderings of the participants are now silhouettes. Much thought has been given to the design of the renderings in *Telemurals*. We wanted to maintain the benefits of video in their simplest form. Adding video to a communication channel improves the capacity for showing understanding, attention, forecasting responses, and expressing attitudes (Isaacs and Tang 1993). A simple nodding of the head can express agreement or disagreement in a conversation. Gestures can convey concepts that aren't easily expressed in words; they can express non-rational emotions, non-verbal experiences.

Yet these cues are not always properly transmitted. There may be dropped frames, audio glitches. Lack of synchronicity between image and audio can influence perceptions and trust of the speaker at the other end. Other challenges include equipment placement. For example, camera placement has long been a reason of ambiguous eye gaze in audio-video links. A large camera offset gives the impression that the person you are speaking to is constantly looking elsewhere. The abstraction in *Telemurals* has the added

Figure 7. *Telemurals* in the settings of the two dormitory spaces.

Figure 8. Body Movies by Rafael Lozano-Hemmer. This projection piece transforms public spaces with 400 to 1800 square meters of interactive projections. Photos are taken and stored of the inhabitants of the location of the public installation. As people walk by, their shadows from light sources project their shadows on the public building. The stored portraits then appear inside the shadows of the passers-by. The scale of the shadows changes as people move closer and further from the light sources.

benefit of mitigating the perception of offset eye gaze and the effects of unsynchronized audio and video.

With *Telemurals*, we are creating an environment where rendered video maintains subtle cues of expression such as posture and hand motion, yet also enhances other cues. For example, changes in volume can alter the style of the rendered video. By adding another layer of abstraction into the video stream, we can enhance cues in a manner that is not possible in straight video streams.

Incorporating Features from Online and Physical spaces

The creation of an abstract environment in *Telemurals* has some similarities to virtual spaces in online worlds. Examples of fantasy environments for conversation can already be seen in online graphical communities. One essentially puts on a mask and enters an electronic communication space. This element of mediation removes many of the social barriers by allowing one to safely enter a stimulating world. Just as this lack of concrete identity may aid in suddenly and continually confronting new people in virtual environments, it may help in our physical-virtual hybrid.

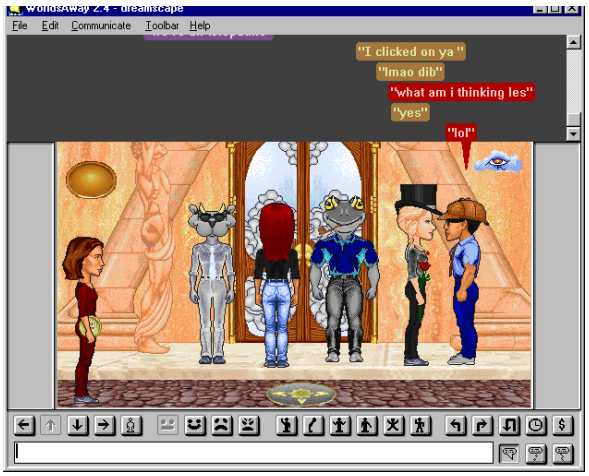


Figure 9. Snapshot from *Avaterra* online community: people communicate as a in this fantastical world.

Figure 10. The first pass at the fading catalyst. The silhouettes would slowly fade to black-and-white photorealistic images the more people participated in the space.

In the snapshot of the virtual world (see Figure 9), the user happens to be the avatar on the left of the window. They can see how they appear with respect to the other participants and change their appearance at will.

Introducing Social Catalysts for Interaction

This rendering of the spaces brought people close to the cameras and close to the inhabitants of the remote space. The interactions, however, were very mundane. People waved back and forth, said, "hello", "what time is it there?", "where are you?", etc. The visual was good enough to be an ice-breaker for interaction, but it was not good enough to sustain interaction.

Fading

We began to explore ways to maintain interest in the space after the initial encounter. It could be that for more intimate interaction, people wanted to see more than just a silhouette of a person. The idea was to reward people for staying in the space by showing them more of the space the more they participated. Therefore, as people moved about the space and spoke, the silhouettes, would slowly fade to a black and white photorealistic representations (see Figure 10).

Reactions to this catalyst were not very favorable. Participants found the floating heads and disconnected limbs of the figures disconcerting and eerie. Several different fading algorithms were implemented until we discovered one that participants were content with and that increased interaction times in the blended space. In this new fading, participants are seen only as bare outlines if they are not interacting (see Figure 11). The more

Figure 11. *Telemurals* rendering over time. The more people become involved (i.e. the more they talked or moved in the space), the more their image begins to resemble a filled in cartoon.

they move and talk, the more their silhouette fills in and begins to resemble our first abstracted cartoon rendering in Figure 2. This prompts the participants to move closer into the space to see. If conversation stops, the images fade back to their silhouette rendering. In this manner, the participants can choose their own level of commitment in this shared space (Jacobs 1993).

Audio Graffiti

The first three weeks during which the installation was up, we did not get much speech across the audio channel. This may be because the microphone did not resemble a traditional microphone, and therefore, participants were not aware of this channel of communication.

We first placed a sign with an arrow next to the microphone so that people would be aware of its presence. We then provided the system with some intelligence to modify its space according speech patterns. When people speak, their audio is passed through a speech recognition algorithm. The algorithm returns text of the closest matching words in its dictionary. This text is then rendered on the shared wall of the two spaces. This audio graffiti is meant to be a catalyst for speaking.

The combination of the signage and the graffiti increased audio interaction five-fold. Exaggerated gestures decreased, probably because people realized it was not their only channel for interaction.

It should be noted that the speech recognition in this open space was far from perfect. The resulting words resembled the basic sounds of the audio. This was usually enough to let people know they were the ones affecting

the space. It also provided the unexpected, yet added benefit of providing some comic relief.

Evaluation Methodology

This work combines the disciplines of technology, communication, and design. Evaluation of this work is therefore threefold.

Engineering

We evaluate if the system functions. Does it work? That is, does it transmit audio and video? Is the sound quality acceptable? Is the video quality and speed acceptable? Are the interface and networks reliable?

Ethnography

We observe and evaluate how the people use the system. The field for this observation study is the semi-public space within the two chosen dormitories. The participants are graduate students who live in the respective dormitory and their friends. We are primarily interested in seeing, (1) how people use *Telemurals*, (2) if the catalysts attract them, and (3) how we can improve the system.

We performed three different types of observations:

- Observation while immersed in the environment
- Observation from mounted camera video
- Observation from abstract blended video

The footage from these tapes was used to annotate patterns of use for this study and were then discarded. We were interested in observing:

- How long people speak using *Telemurals*
- The number of people using the system at any one time

- The number of people present but not interacting
- The number of unique users (if possible)
- The number of repeat users (if possible)
- The number of times and the duration that people use *Telemurals* in one space only
- Repeated patterns of interaction: gestures, kicks, jumps, screams

These are factors that we believe are indicative of levels of interaction. However, one must always be open to the unexpected and attempt to find other underlying patterns as well in studying the social catalysts.

Design

We evaluate if the system interface is well-thought-through, coherent, and innovative. This was in the form of a studio critique. Professors from various architecture and design departments and research scientists have been invited and have volunteered to participate in a series of critiques.

Discussion

Technical

As an engineering project, *Telemurals* works. It runs on the school network and typically uses less than 1MB of bandwidth with audio latency varying from 500ms to 1 second depending on network usage. The networking audio and image libraries are all written in C over UDP, and we use the Intel OpenCV library for image segmentation.

The video was reliable, the audio had acceptable lag, and the system ran continuously for over three months. The one technical challenge that could use improvement is the audio quality. Using just one microphone does not cover the intended space and the acoustics of each space play a huge role. We are experimenting with microphone arrays and with physical objects that one interacts with that contain the microphone.

Telemurals was evolving throughout its construction and connected installation period. We experimented with several different renderings of people at each end, we changed the fading algorithm, changed the hours of operation, and changed the *Telemural* wall site at one of the dorms. These changes were made according to suggestions and critiques of the residents of the two dormitories and professors within the lab throughout a five month period.

Social: comparisons and contrasts

Time-schedule, social events, signage, interface, trust, site selection, and a changing environment proved to influence population mass at the *Telemurals* sites. The motion of people, ambient noise, and the speech-to-text graffiti created from the users' own words kept people at the site.

Hours of operation

The *Telemurals* observation took place in May and June of 2003. During the initial two weeks, Telemurals ran for two hours each Wednesday and Sunday night in conjunction with a coffee hour/study break.We had requests from both spaces to increase the hours of the connection. *Telemurals* then ran every night for two hours over a two week period and then ran continuously, twenty-four hours a day.

We discovered we had a larger population of use per hour and longer linked interaction times when *Telemurals* was up for shorter intervals of time (two hours, two days a week and two hours every night vs. twenty-four hours a day). We believe it became more of an event - something that should not be missed. Nevertheless, we continued getting requests to run it continuously.

Figure 12. A snapshot of the day of the wedding.

During the twenty-four hour a day use, there was an increase in the time the system was used at one end and not the other compared to the shorter time intervals of use. Many more people used it as a mirror when they thought they were alone.

Events

Dormitory events such as meetings and social hours attracted large crowds to the *Telemurals*. Sometimes it was for comic relief; other times it was because of the quantity of people. One person at the *Telemural*, whether at the local or remote end, tended to attract more people. A wedding party

Figure 13. Three by three foot signs placed at entrances of both sites:

Telemural Connection in Progress
This is an audio and video connection between Ashdown and Sidney Pacific semi-public spaces.

Video is sent over as you see it in the Telemural here. This transmission is not stored or save.

The camera here is used for an ethnography study. Patterns will be observed and footage will be destroyed.

For more information, please see:

The folder contained detailed information about contacting medical services and the person responsible for the installation if problems arise.

proved to be the most interactive period, with children repeatedly running back and forth across the wall. Food associated with these events also attracted people. Moving food in the field of view of the mural provided an object for interaction and made the *Telemural* a popular spot.

Signage

Signage was placed in the entry ways of both spaces to describe what was being transmitted, where it is being sent, and to inform people of the presence of the camera, microphone, and the ubiquitous link. Of over twelve hundred people living in both dorms, we had one complaint asking that we shut off the microphones between both spaces. This person felt the system was eavesdropping on them as they waited for the elevator. The abstracted images were not a concern in this case.

The link signage was more problematic. People were not entirely convinced there was a link connecting the two dormitories. This may be because there was no one at the other end at that time or because the interface was unfamiliar. There were several instances of students arranging meetings at each mural and calling each other on cell phones to verify the connection.

Others didn't read the signage and thought it was simply an abstracted mirror. People were often confused when another silhouette appeared and sometimes left that space of the screen so as not to be in the way.

It was three weeks before we noticed a significant amount of interaction across the link. Prior to that point, people were gesturing and moving, but very few people were speaking across the link.

We added more specific signage labelling the microphone and speaker. The presence of speech at either end of the mural increased five-fold. The speech recognition then became more apparent. It became a positive feedback loop whereby people kept speaking and staying in the space until what they said was recognized adequately.

Ideally, we wanted to avoid all instructional signage and let the interface guide the interaction. This did not work. Perhaps as the interface becomes more commonplace, we could remove the labels.

Interface

The interface went through several iterative design phases. We first began with a connection that linked two spaces with straight video and audio. After a period of observation, we wanted to make the wall more approachable. People were not attracted by it; in fact, some people deliberately avoided it.

We wrote an edge-detection algorithm to render the video at both ends so that the video appeared as a real-time moving comic (see Figure 2). This effect was described as "fun" by students. It added a layer of abstraction and increased participation in the link compared to the straight video. We received many requests to provide this filter as a "mirror" with which to look at oneself as well as a link.

People wanted to see how they were presented at the other end. Providing a picture-in-picture image or segmenting the screen into two rectangles appeared to emphasize that the spaces were remote and distinct. To create a shared space, we blended the two spaces in the same scale and form. When people saw themselves as well, they preferred more abstracted renderings and silhouettes to photorealistic images in order to protect their privacy.

We now had our mirror and our link. The mirror helped make the wall interactive even when no one was at the other end, sometimes even attracting someone to the space. Simple silhouettes were enough to depict the presence of a person or persons at the other end and a sense of activity in the space. This provided an environment for safely starting interaction.

We had many interactions between the spaces, however, they were not sustained for a significant period of time. The next step was to provide a motivation for staying in the space. We had the ice-breaker; now we needed people to use the channel in more interesting ways than simply saying "Hello" and "Where are you?".

We altered the image over time to become more detailed, approaching the detail of our cartoon rendering, the longer a person moved or made noise in either space. By doing so, the length of discussions increased slightly and people spent more time in the shared space.

Figure 14. (left) Original site after projecting surface was painted green. We placed an easel with a canvas in the location for project that evening. (right) Final location on the opposite side of the room.

Site Selection

The two participating dormitories have an interesting history. A good number of the inhabitants of the newer dormitory had previously lived in the other. This meant that some of the students had a higher likelihood of knowing someone at the other end. We don't know to what extent this played a role, but we know there already existed a social connection. Oftentimes, students arranged meeting times and spots at their respective *Telemural* either as a "rest-area" or a social break.

One day we arrived in one of the dormitories to find that our white projection wall had been painted dark olive green. We improvised and put up an easel with a white canvas and projected onto this. This proved more frustrating than useful.

During the time when the *Telemural* ran for two hours at a time, or when no one was at either end, the screen looked bare and abandoned. When one sees a large canvas on an easel, they expect to see something on this canvas. We moved our site to a different location near the elevators and projected onto the white wall. The new space provided more of a surprise, better mural visibility because the surroundings were not as bright, and more time to interact while one waited for the elevator. The empty wall was not as disturbing when it was white because that was its default state -

and it provided this surprising reward - when one did see something there that was not in the norm. The wall also eliminated the distancing effect of framing that is common with projection screens and picture frames.

Observations

There were many trade-offs among the three different observation methods that we used. When observing and taking notes while immersed in the environment, we originally thought participants would be self-conscious about being watched. On the contrary, we found that having a person present at either end is a big attractor, even if it is a note-taker. People would sometimes come over just because someone was watching the wall.

As with the immersed observation, we expected the live video captured from the camera mounted on the wall would cause people to behave differently. In this case, we believe it did. People were not very comfortable with the idea of being taped.

The final observation method involved capturing the abstracted video from *Telemurals*. This was the easiest because the image was reciprocal from both sides and did not cause any objections from the occupants of the dormitories. This did not provide data outside the camera field of view, as did the immersed observation, but it did not upset anyone and even encouraged some to perform in front of the camera as if hoping to be watched.

By combining these three methods, we hope to have gained a better understanding of the use and interactions in the space and with *Telemurals*.

Design

Telemurals has been critiqued by three professors while in use at each installation site. It was noted that the abstraction not only enhances certain social cues such as gesture, but also mitigates the confusion associated with gaze and audio-video synchronicity in teleconferencing systems. The interface was described as “evocative and fascinating” by one design professor.

This design evaluation is far from complete. We need to hold many more design critiques in order to state a meaningful evaluation on the design front.

Privacy

Privacy control was one of the major forces behind the design of *Telemurals*. The privacy gained through abstracting people in the setting allowed for participation with less risk and without a covert feeling of surveillance. Out of over 1,700 users, we had one submitted complaint about privacy. A student was appalled that a private conversation near the elevator might have been heard in a dorm over a mile away. It was not the visual that was disturbing, but the audio. This suggests that perhaps we should consider filtering the audio to resemble garbled chatter as we abstract the video to silhouettes, and then moving towards clear audio the more people interact.

There are also serious privacy issues in observing such an interface. We recorded some video of the interactions and of the people. If straight video directly from the camera was captured, there was a large sign saying this might be happening. Abstracted video was captured at random times every day. All of the audio and video captured in the *Telemurals* interface was annotated, analyzed, and then destroyed.

Summary

The transformations in this project — the abstraction of person, the blending of silhouettes, the graffiti conversation, and the fading from abstract to detailed — provide social catalysts for the experience. This new wall generated by filtering creates an icebreaker, a common ground for interaction, and an object for experimentation.

Further elements that contributed to the sociability of the *Telemurals* space that we will revisit again in later examples of this dissertation are the abstraction, the human scale of the interface, and the adoption of language.

Abstraction helped us mitigate privacy concerns as well as create a world that is more elastic that those of unmediated environments. The human-scale interface made it possible for the display to occupy a large wall of a room and blend in with the passersby. Participants would sometimes dance together remotely and perform kicks onto their remote companions. This

helped users negotiate space and proximity within the space and between their remote companions.

The ultimate goal of *Telemurals* is to create new styles of gestural movement and speech interaction by providing a common language across the two spaces. The pursuit of this theme threads throughout the remainder this work.

Future Directions

Engineering

We would like to improve the sound recording in each space by providing several microphones in an array to capture audio evenly so that people don't crowd around the one microphone. We found that those near the microphone spoke more than those further away.

Design

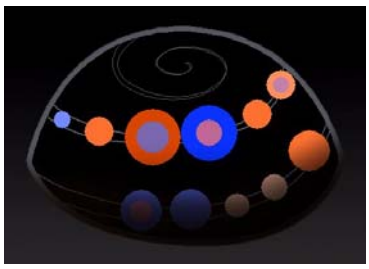
The renderings could be even more evocative of the activity in the space. For example, if someone yells or speaks loudly, their scale would increase and tower over the other participants.

Ethnography

We would like to create a *Telemurals* link between another pair of sites. This would provide a comparison among a number of sites, how public they are, as well as a study over a longer period of time. An outdoor mural would be exciting and more representative of a public space.

CHAPTER 4

Abstraction for Visualizing Conversation



In this chapter, we take an audio conversation between two spaces and render it abstractly to create a visualization of the conversation that highlights certain cues that might not be readily apparent in an audio-only channel.

To illustrate these concepts, we designed and implemented a communication object called, *Visiphone*. We will first describe this project, and then address how abstraction and time affect its function as a communication object and as a catalyst for interaction.

Visiphone

Visiphone is a communication object that bridges the distance between two physically separate spaces. We began with the notion of building a virtual portal between two spaces, one that would allow the inhabitants of the two separate spaces to communicate easily and to be intuitively aware of each other's presence.

Previous Work

There have been a number of "media space" projects that connect geographically distinct locales with some combination of audio and video (Bly and Irwin 1993, Buxton 1992, Dourish and Bly 1992, Ishii et al 1992), as well as studies of the relative affordances of audio, video, and other media (Isaacs and Tang 1993). Smith and Hudson's work on low disturbance audio found that audio, even when filtered to be incomprehensible (for privacy, in their application) provided a good sense of awareness of the presence and activity of others (Smith and Hudson 1995). The *Somewire* project is one of the most relevant, since it was designed to foster casual interactions among colleagues (Singer et al 1999). Here, Singer et. al experimented with a number of visual interfaces in conjunction with an audio-only media space. They found that control over such features as localization or other attributes was not needed, but that information that supplemented users' knowledge of the social aspects of the space, such as awareness of the presence of others, was quite useful.

However, most "media space" work has been done in the context of work environments, which differ significantly from the home in many regards. While most studies of technology for the home have tended to focus on labor-saving devices and home automation, some useful ethnographic studies have examined the importance of communication in the domestic sphere and the types of technology that support it (Venkatesh 1996); it is becoming increasingly clear that communication is one of the most valued uses of computer technology in the home.

Design Process

One approach would have been to establish a live connection, such as the "video windows" (media spaces — see Chapter 2) that have been installed in venues ranging from night clubs to research offices. Yet, while they provide a high degree of communicativity and awareness, these connections can also feel intrusive and awkward: some researchers have noted that users of these systems, especially, in more private domains, eventually just keep the video turned off (Jancke et al). Another approach would have been an audio only connection, which is less intrusive and still allows the participants to talk informally and easily. Yet audio alone has several drawbacks, especially in terms of awareness. It is difficult, especially in a noisy environment to know if one's voice has carried or if others are speaking at the other

end. Furthermore, lone periods of silence make it easy to forget the device, which then takes on a quality of covert surveillance.

Our approach is to connect the spaces aurally and then to visibly render the sound flowing between them. *Visiphone's* graphics express the dynamics of the conversations originating at both locations, thus providing visual feedback that one's voice has carried sufficiently and indicating the presence of those on the other end. It portrays the existence of the connection even in moments of silence, thus removing the surveillance-like aura of the audio-only system.

For a piece such as *Visiphone*, form is function: it must be attractive and intriguing enough to claim a central place in a kitchen or living room. The dome shape makes it readable from any location; it's placement on the pedestal puts it into the category of sculptural object (as well as concealing the projector). The graphics themselves are designed to convey a sense of rhythm and activity - to visibly represent the connection between the two spaces.

Design Goals

Our goal in designing *Visiphone* was to make a device that would enable casual communication between to distant common spaces. Unlike and office environment, in which users are generally seated at a desk, people at home or in common spaces move about freely. We located our design scenarios in this multitasking environment, in which people are engaged in several activities at once and where conversations often ebb and flow as attention is temporarily diverted to other tasks.

One of the key differences between conversations that are conducted face to face and those on the telephone is that the latter tend to be more continuous. In person, a conversation can start and stop, pausing when a topic is exhausted or other tasks beckon; on the telephone, long silences are awkward and uncomfortable - and often signal serious disagreement between the parties. To support more casual conversation, our design would need to help overcome the requirement that the conversation be the unceasingly forefront activity. We identified several possible causes to get telephone conversation's center stage position: cost and the phone's audio-only indication of presence.

Cost: We began with the premise that Internet telephony will soon make it feasible for people to maintain long-distance, long duration audio con-

nections without the anxiety of mounting toll charges that accompany today's long distance phone calls. The first *Visiphone* prototype used the *MBone* to send over the network.[1] Currently, all that is required is an internet connection (of reasonable bandwidth).

Audio-only indication of presence: In a phone call, silence is often met with the response, "Hello? Hello? Are you still there?" When speech ceases, there is the sense that the connection has been lost. We hypothesize that people feel compelled to maintain continuous conversation on the telephone, in part, to maintain the perception of a live connection. *Visiphone's* interface is in motion as long as there is a connection; our intent here was to provide a separate, intuitive indicator of connectivity that would relieve the audio channel of this responsibility.

An open long term audio connection alone between two spaces would make it possible to walk around a room and talk to a distant friend at will. Yet such a system would also have an unfortunate surveillance-like quality: it would be easy to forget that one's space was actually a portal to another space and that all of one's comments were being heard elsewhere. A visible interface that indicates the existence of the live connection alleviates this problem, serving as an ongoing reminder of the audio link.

Another function of the visible interface is to serve as a focus of attention. Anyone who has attended a meeting at which some participants are present via speaker-phone knows the phenomenon by which the little black object on the table becomes a stand-in for the person: one addresses one's remarks to it, one looks at it when it is "speaking". We focus on the object because it is difficult to converse with a disembodied voice - if an object is associated with that disembodied voice, it becomes the representation for the connection and person. With the traditional speaker-phone, the physical object provides no information, nor is it designed to evoke a person, conversation, connection, etc. With *Visiphone* we sought to make an object that is abstract enough to be suitable to all users, contexts, and topics, yet that also reflects the rhythm of the ongoing conversation.

1.A later version of *Visiphone* was implemented using Java Media Streams API.

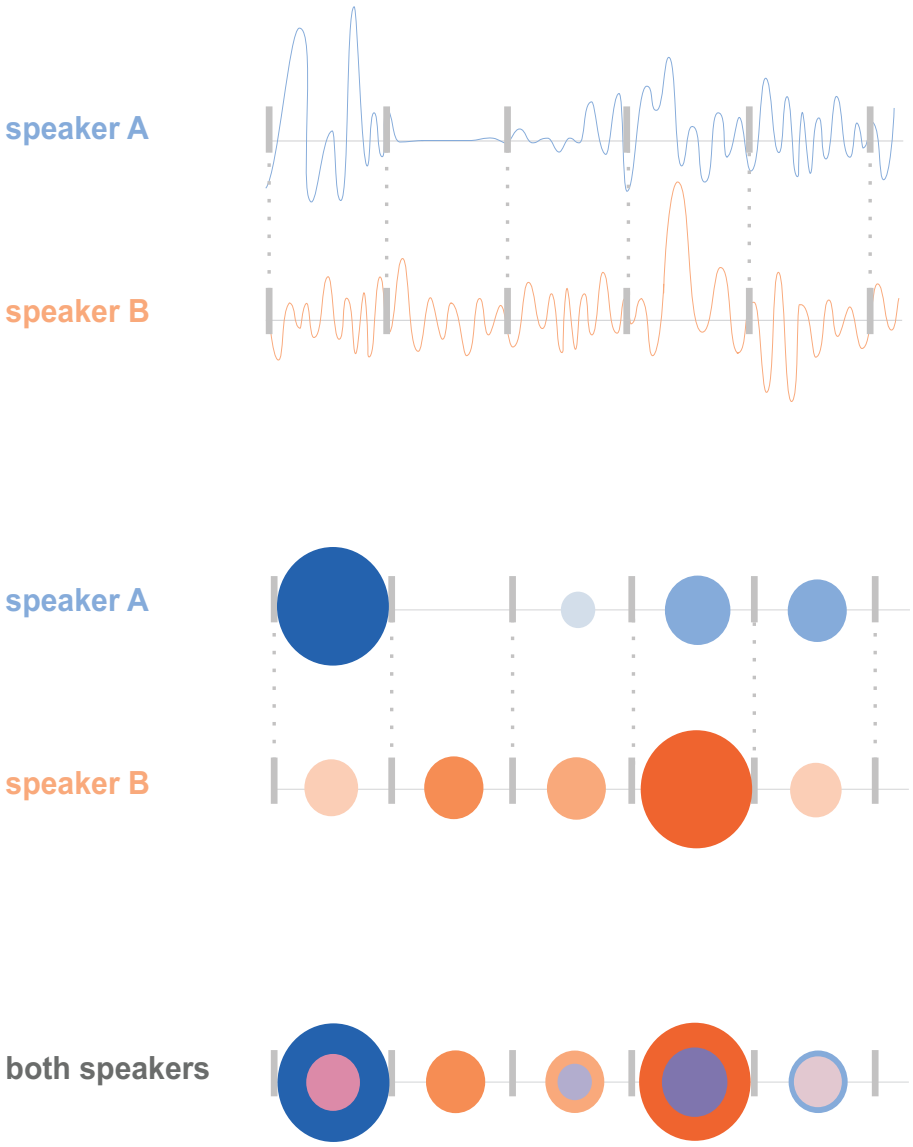


Figure 1. Process of audio renderings: (top pair) a pair of traditional audio representations. (middle pair) the power amplitude of the audio is mapped to the radius of a circle and to the saturation level of that circle. (bottom row) the renderings from the second pair are blended in HSV space to produce a single blended chain of circles.

Two further considerations influenced our design. The system needed to be easy to use: no complex controls, no custom installation. And, it needed to be a decorative object, one that people would welcome as a featured object in the home.

The Design

Our final design for the *Visiphone* system consists of two stations connected via the internet. Each station has a dome on which the visualization is projected (see Figure 5). When a live connection exists, the dome displays a continuous moving spiral of circles. The central dot represents the present moment. If it is a small gray dot, there is no sound going between the two spaces.

When sound is originating locally, the current circle is orange; when sound originates at the remote location the circle is blue. The size of the circle is proportionate to the volume of the audio.If sound is coming from both locations, the colors are shown as concentric, blended circles (see Figure 1).

The dots spiral outward from the center, so the display shows the history of the last half minute or so of conversational rhythm (see Figure 2).

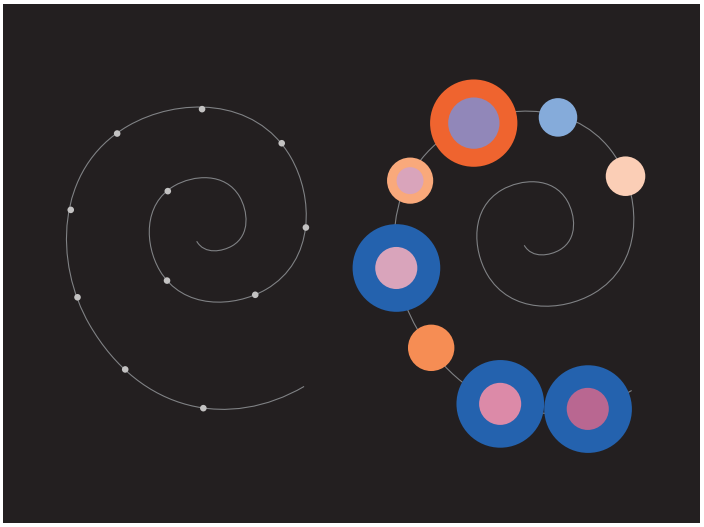


Figure 2. (left) *Visiphone* spiral connecting two silent locations (right) *Visiphone* spiral connecting two locations with emanating sound.

Visiphone's display is a translucent dome sitting on a pedestal (see Figure 3). The graphics are projected into the dome from below. The dome shape makes it an interface in the round: one can view it from any side. This is essential for an object meant to create a connection between two inhabited, real-world spaces in which people move about. The design of the dome itself is also a key element in this multimodal interface and its size, location, and appearance influence its use and its ability to portray a sense of awareness and continuous connection in the space.

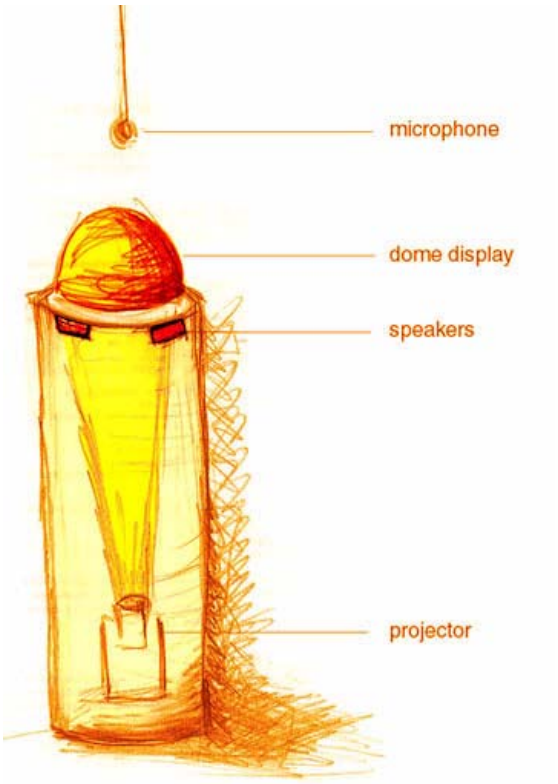


Figure 3. *Visiphone* installation setup.

Figure 5. *Visiphone* in use at two remote locations.

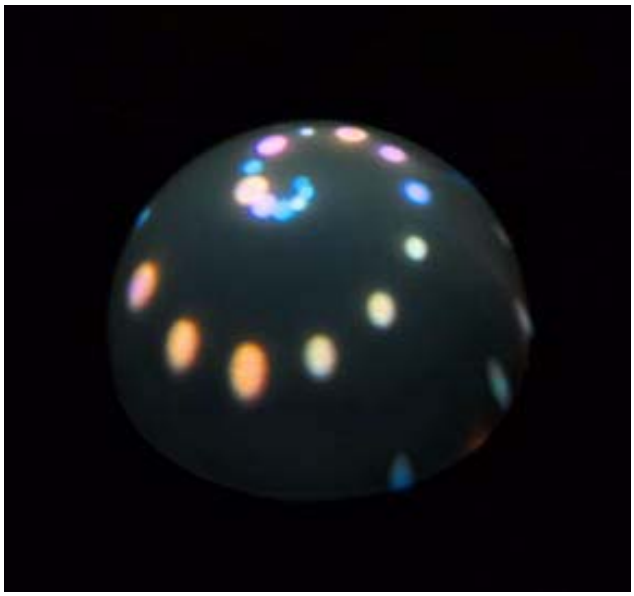


Figure 4. The illuminated *Visiphone* dome.

Figure 6. A *Visiphone* station in use at the SIGGRAPH '99 Emerging Technologies exhibit.

User Response

Although we have not conducted formal user studies of *Visiphone*, it has been on display in public environments[2] and used by hundreds of people. People have been quite enthusiastic about it. The form of the display has proved to be quite important - the spiralling dots are often described as mesmerizing and this aesthetic appeal is an intrinsic part of its value. One of the more surprising comments, but one we heard repeatedly, was that people thought it would make a good "therapy" tool; they were interested in the way the dots could show patterns of interruptions and of individual conversational dominance.

2. *Visiphone* was shown in the Emerging Technologies exhibit at SIGGRAPH '99 and has been a featured demonstration at the MIT Media Lab

It still remains to test the system in its intended environment, a domestic space. In real use, it would need to compete with the telephone for clarity of sound (our current implementation is rather noisy and has some lag). And it remains to be seen in practice how well it accomplishes its stated goal, which is to support casual conversation in an active environment.

Summary

Visiphone does not add any new data that is not already part of the audio channel. It simply re-presents a subset of the data (source and volume, in temporal slices) in another medium. Although from an information theoretic standpoint the visual data is redundant, from a user's standpoint this reification of the audio serves several purposes. Using an audio-only speaker phone to provide continuous, long-term connection has several drawbacks: in a noisy environment, it is difficult to know whether one's voice has carried over to the remote end or to know to pay attention to new voices emerging from the phone; long periods of silence make it easy to forget the device, which then takes on the unwanted quality of unobtrusive surveillance. By making the audio visible, *Visiphone* turns the speaker-phone into a portal between spaces. The existence of the connection is recognizable even during moments of silence and the dynamics and inflections of the conversation are made salient by the abstract visualization. The sudden appearance of vivid dots draws one's attention, even in ambient noise and initially masked sound. Seeing the color change on the display is a simple but effective mechanism for noting whether one's words can be heard at the other end. *Visiphone's* graphics do not portray the technical aspects of sound (as do, for example, the audio renderings seen in sound-mixing boards); it's purpose is rather to enhance awareness.

Visiphone as Social Catalyst

There are several features that act as social catalysts in *Visiphone*. As I discuss the differences in the *Visiphone* designs, I will highlight the social catalysts were effective.

Figure 7. Various *Visiphone* sizes and shapes: (a) this is a 3 inch diameter dome interface. It rests on a 14 inch high pyramid pedestal that houses a tiny projector. (b) this is the 8 inch diameter table-top dome (c) this is a 12 inch diameter angled two-dimensional *Visiphone*.

Scale

Altering the scale changes everything. Scale dictated public and private uses. The three inch diameter dome in Figure 7(a) was suited to private conversations and was often cupped in the palm of the hand. People tried to rotate the dome to go back in time. The eight inch version (Figure 7(b)) was usually surrounded by several people; it was of a good size to rest both palms on it and many people did just that. The twelve inch display was usually viewed at a distance, and from one general direction.

The smaller dome encouraged people to move in closer to have a conversation. This *Visiphone* was treated as an intimate object that encouraged intimate conversation. The 8 inch dome encouraged conversation not only between the two locations, but also around the respective local tables as well. The largest *Visiphone*, was more of a public showpiece. People asked what it did, but it was more of an abstract display.

Form

The form of these different interfaces suggested that they be used in a different manner. For example, the dome shape encouraged people to draw nearer to the display and around the display. There was a great tendency to grasp the dome one's hands — oftentimes to try and move the dots in this manner.

The flat angled display was intended to be viewed from a distance as well as close-up. It was designed for a larger auditorium setting where people faced the moving display. This interface was not attractive to users. Perhaps it was its distant placing so that it could be viewed by many or perhaps it was the flatness of the interface.

The dome interface was a more effective social catalyst. It was visible from most angles, and therefore attracted people around it at all angles. This encouraged local interaction as well as remote interaction.

Motion

The motion of the object was not only a calming, aesthetic visual, but also provided a sense of assurance within the spaces. One knew what to expect in terms of connectivity, privacy, and conversational turn-taking. It was also described as mesmerizing which kept people near the dome for a good amount of time.

Previous *Visiphone* interfaces that were not in continuous motion or were in predictable discreet motions were often misunderstood and, therefore, ignored. A static interface of an ongoing conversation is equally meaningless.

Abstraction

Abstracting the traditional audio wave patterns, into the simple graphical element of the circle allows one to think of audio from a different perspective. The circle unites divorced the technical connotations from audio channel and for a new frame of reference to the conversations.

We are familiar with other forms of volume representation whether they be bar chart visualizations such as in Quicktime™ or waveforms as in many sound recorders. Although these representations are readable, they have a

technical connotation and their representations are fleeting. The string of beads visualization in *Visiphone* separates itself from these traditional technical representations. Its simple graphical elements are easy to comprehend and, therefore, append to each other to add a historical context to the audio stream. The color and the discrete units display salient features such as turn-taking and interruptions far more clearly than would a continuous waveform representation of two audio streams.

Time

The time element implies a history element. The size of the *Visiphone* object dictated the amount of history displayed. A large display, showed longer interaction times. In this display, conversational dominance would be striking. A shorter history implied more immediacy in the conversation, yet also kept one staring for the next circle to appear. Changes were more evident and pronounced.

Revelation

The major catalyst in *Visiphone* is revelation. *Visiphone* allows you to see things you know but may not realize that you know. This ability to perceive a wide spectrum of conversational patterns from the one cue of volume shows the power of this mediated connection.[3] As previously mentioned, we are not adding information to the audio stream. Just using what is already there, we are highlighting patterns that are not obvious in an audio-only mode or using more traditional representations of audio.[4]

3. In one version, we included pitch-tracking. This added feature did not prove as effective as the reinforced visualization of volume in the cycle of a conversation.
4. We had an uncommonly large number of requests from marriage counselors with regard to *Visiphone*. This audience was not our intended audience, but it does show that these patterns may be useful. There were also requests for implementing *Visiphone* as a means to visualize negotiations during meetings.

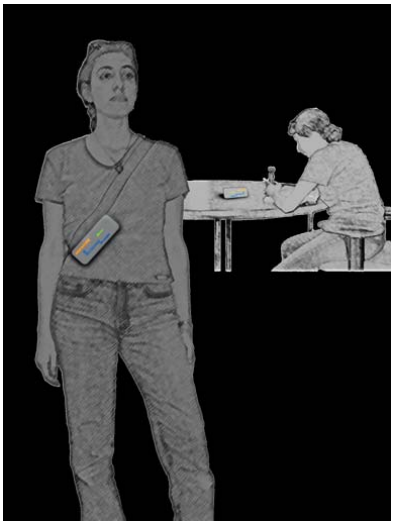


Figure 8. A mobile derivative of *Visiphone* has been developed for mobile phones. This supports multiple users — Each user is represented by a colored snake of beads. The more one talks, the longer the snake becomes until it cuts itself off. Snakes participating in turn-taking congregate together. Lone snakes lurk in the periphery. This version combines the game-like quality of the arcade game Centipede with speech influenced movement.

CHAPTER 5

Physical Embodiment of Virtual Presence

Cafes function very well as informal public gathering places. One can enjoy the company of others or be quite comfortable alone. And they are great places to sit and watch people.

The online world also functions as a public gathering place. As in the cafe, conversation is one of the primary activities - but with some striking differences. Online, conversing with strangers is quite common and there are few barriers to such interactions, while in the real world such encounters are less common and occur couched in complex social rituals. In the online world, one is fundamentally alone: although there are many others virtually present, one's sense of their presence is minimal. In the real world cafe, the number of people is fewer, but their presence is far greater.

These two worlds come together in the *ChitChatClub*. It is a real cafe, with real tables, real coffee and pastries. Yet the customers seated round the tables may be present physically or virtually -- some of the chairs are ordinary seats, accommodating the human form; others are seats for avatars, equipped with monitors and network connections

In this chapter, thru *ChitChatClub*, we explore the physical embodiment of virtual presence.

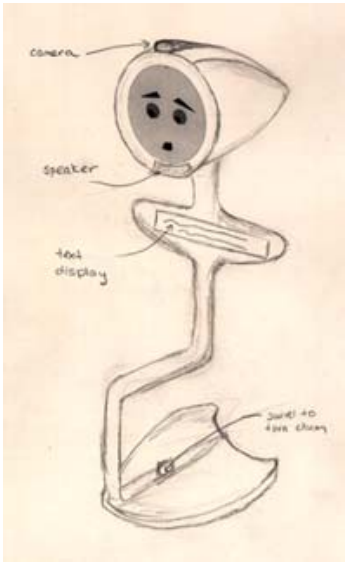


Figure 1. (top right) Artistic rendering of the *ChitChatClub*. A cafe space is populated by people who walked into the cafe, and who remotely connected to it through the conduit of the "avatar" seats at the tables.

(bottom) Early rendering of an "avatar" seat.

Figure 2. Cafes introducing computers into the environment often separate their space in two — as if an invisible wall separates those who are there to sit and chat and those who make it a workspace. We wanted to use technology to foster interaction. Hiding the computer from view aids us in this task.

The Cafe Scenario

ChitChatClub is an experiment in bringing people together in a mixed physical and virtual environment. Online chatrooms and real world cafes are both venues for social interaction, but with significant differences, e.g. the participants' knowledge of each other's expressions and identity and the more governing introductions, turn-taking, etc. Our goal is to create, thru careful design of the physical environment and computer interface, a place that gracefully combines these two cultures (see Figure 1); the analysis of how well this space actually functions will further our understanding of social interaction, both online and in person.

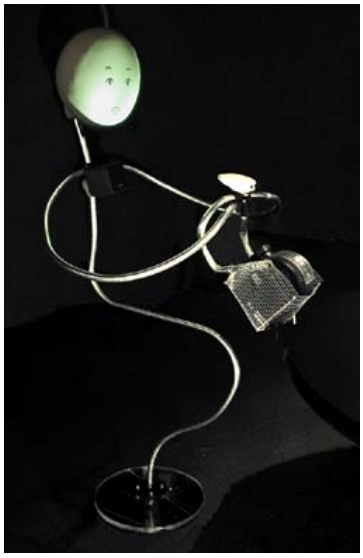


Figure 3. Slim. Our first *ChitChatClub* "avatar" seat.

The First ChitChatClub Installation

Physical Avatar

The physical "avatar" seat was designed to be of human scale. The idea communicate with an interface that was at the same head level as opposed to a monitor with similar proxemic codes as sitting face to face. We designed the "avatar" seat to be anthropomorphic to a degree, but not so anthropomorphic that one would expect human movement and human expression.

The body frame is meant resemble a relaxed figure of human proportion. The head resting on the frame has some curvature, and is painted white so that it makes a good projection surface. A projector hangs in a wire basket beneath the crossed "hands" of the "avatar seat", or in this case, Slim. This projector is aligned to project moving faces onto Slim's head. Above Slim's the crossed "hands", rests a camera facing away from Slim. This camera captures video of Slim's companions in the physical *ChitChatClub*.

Figure 4. (left) Physical *ChitChatClub* (right) Remote user

The Local Space

The local space of the *ChitChatClub* is a cafe setting. The "avatar" seat, "sits" at a table, and people entering the cafe, may choose to sit near him or her. Figure 5, shows a picture of one table in a *ChitChatClub*.

The Remote Interface

The online site is a portal for the remote visitors to enter the *ChitChatClub*. Using the *ChitChatClub* website, the remote cafe attendees can create an appearance for their visit, they can choose where to sit, and they can con-

Figure 5. Two people sitting on each side of Slim, our "avatar" seat.

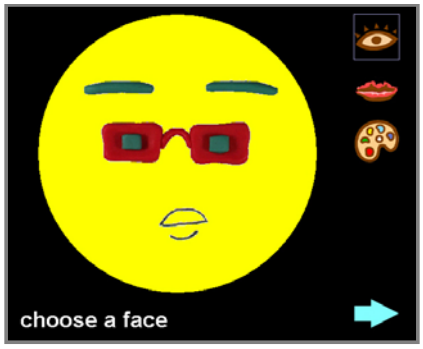


Figure 6. Alternate view of face-creating interfac with some claymation components.

verse with the with the cafe's local participants using either audio or text. While they are conversing with the visitors of the physical cafe, they see an abstracted video representation of their table companions from the point of view of the avatar seat they occupy. Although the remote visitors may type as well as speak to converse, they always hear what the cafe's physical participants are saying.

Getting Started

When an online participant first encounters the online *ChitChatClub* interface, they are presented with an interface for creating an avatar face for their respective avatar seat (see Figure 7). Using this interface, the visitor can customize the appearance of their avatar seat's face. They may choose face shape, eyes, lips, as well as the color of each feature from a selection of handrawn, claymation, and cartoon facial components. The avatar faces resemble cut-out animations.

After the visitor creates their appearance, they are ready to proceed to the *ChitChatClub* entrance. Here they see a graphical birds-eye view of the layout of the physical space (see Figure 7). This representation shows the location of the tables, chairs, and avatar seat. It also shows which seats - regular and avatar - are occupied. The participant uses this image to select which avatar chair to occupy.

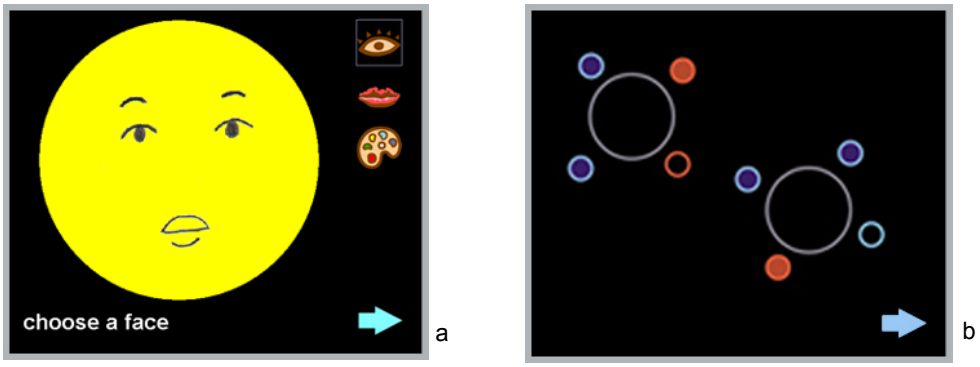


Figure 7. (a) The face selection interface.The user can cycle through different eyes, lips, and face color by clicking on the representative icons on the right. (a) Birds-eye view of the cafe.Blue represents walk-in visitors and red represents kiosk visitors respectively. A vacant seat is depicted by an outlined circle; an occupied seat is depicted by an opaque circle.

Figure 8. Communication interface with different orientation.

The Communication Interface

Once a seat is selected, a two-way audio and video connection is established. In the cafe, the remote participant's avatar face appears in the chosen chair. At the remote location, the participant sees a live, processed image of the cafe as seen from that chair (see Figure 9) and can hear, see, and participate in the conversation at that table.

While connected, remote visitors can communicate by talking or by typing and can momentarily change their avatar's expression to appear happy, bored, disgusted, sad, or angry. Conversation from the cafe to the online participant is still received in the form of audio and processed video.

This interface was ultimately flawed. The remote users would spend a large portion of their time selecting facial animations; this deterred from the conversation. They also continued to click on the facial expressions so as not to appear inattentive. Users also wanted to direct the gaze of their avatar themselves as apposed to negotiating with the visitors to the physical cafe for rotation of gaze. We concluded that for these reasons, we had to make facial expression selection easier and effort-free, and allow the remote user to see how they appear at the physical cafe end.

Figure 9. The remote interface. The top shows a live video feed of the physical space. Below the user can choose from five expressions. The physical avatar face animates to that expression and then back to neutral.

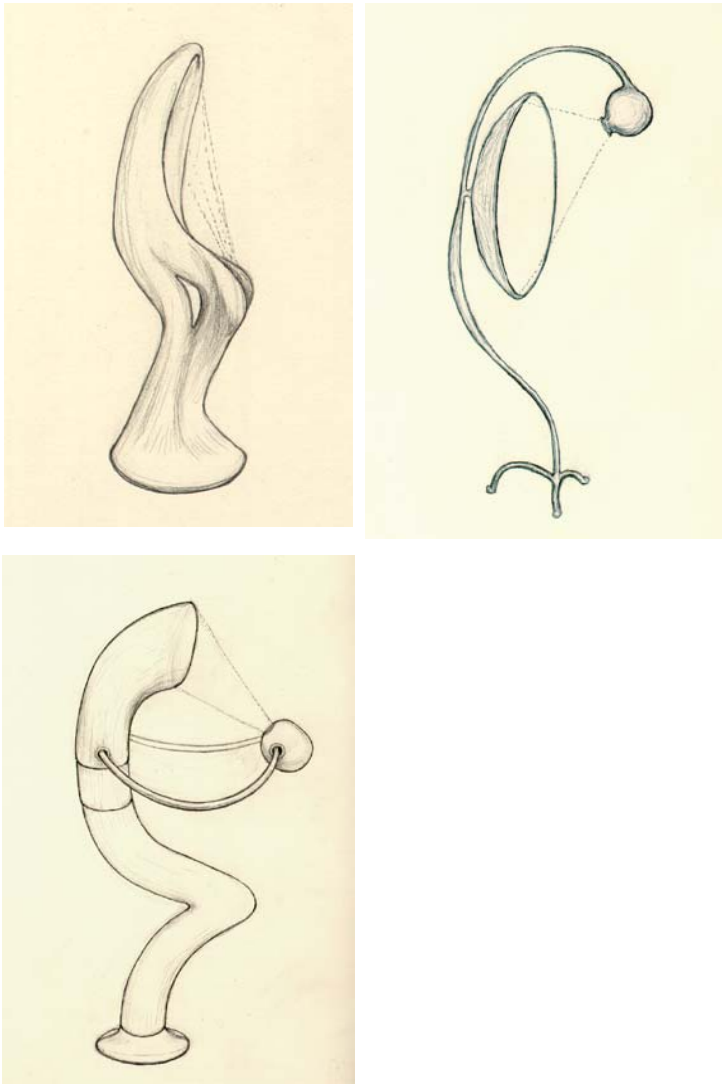


Figure 10. Design sketches for second "avatar" seat. The top two designs were for experimentation with more abstract social cues. The bottom design, housed a motor for controlling gaze. We chose the last design.

Figure 11. Building the new chair.

The Second ChitChatClub Installation

Physical Avatar

We began to design a new "avatar" seat to complement the needs of our new proposed remote interface. Allowing the remote user to control their gaze into the space was a priority. The new seat, therefore, would require a redesign with a motor.

The new chair would also possess an anthropomorphic form, yet would not be a human sculpture. The final design and implementation of the second "avatar" seat, Orlando is shown below in Figure 12.

Figure 12. Orlando. The robotic "avatar" seat.

The Remote Interface

Entering the virtual component of the cafe is the same as in the previous version. What has change in the second version, is the addition of a set of facial components, the communication interface, and the "avatar" seat.

The Communication Interface

In the new remote communication interface, the remote participant could view the expression that was being projected onto their avatar seat head and control the motorized gaze of their avatar seat (see Figure 13). Novel to this interface was a new, comprehensive expression palette in the form of a wheel.

Expression Wheel

The expression wheel is designed to be a simple, intuitive interface for performing facial expressions. Facial expressions composed of componential elements (Smith 1997) such as the lowering of eyebrows and the raising of lip corners are mapped in smooth transitions around the circle clockwise. (see Figure 14). Higher intensities of these expressions are at the perimeter of the circle and blend to more neutral expressions towards the center of the circle. To make it easier on the user, there was no clicking involved. One simply had to mouse over, the area of the expression.

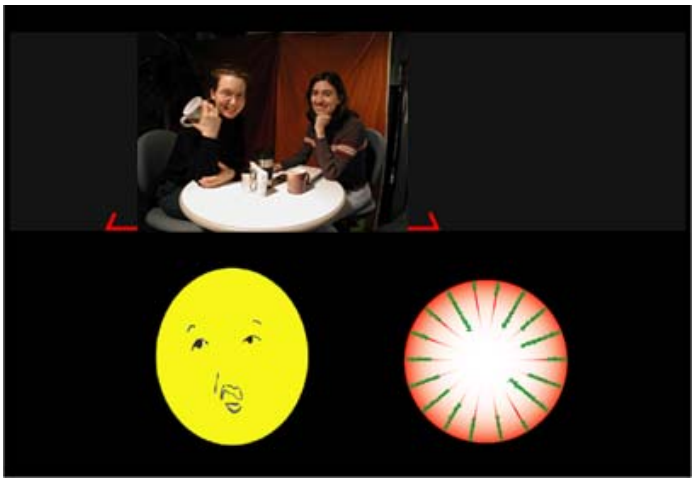


Figure 13. Remote interface with visible remote expressions, gaze control, and new expression wheel.

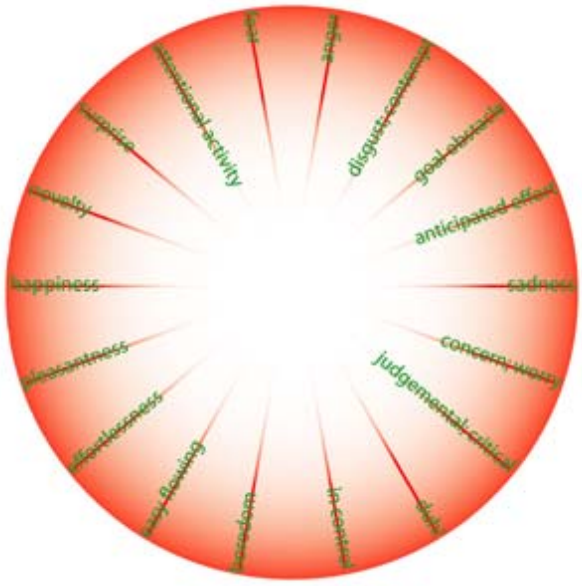


Figure 14. The continuous expression wheel.

Even with this new continuous wheel, users continued to focus on the expressions and not on the conversation. To shift the attention to the interaction while maintaining expression cues during the conversation, we decided to abstract the representation of the expression wheel and make the expression selection somewhat autonomous.

We do this by tracking the pitch of the remote user in real-time and using simple heuristics to alter facial expression. One example of this is the correlation of rise in pitch to rise in eyebrows. Although, the expression selection is semi-autonomous, the user always has the ability to override the system and select their own expression independently.

Figure 15. Abstracted continuous emotion wheel. This expression wheel maintains the continuum of the previous wheel and also allows for a level of automation in expression selection.

Figure 16. The *ChitChatClub* communication interface. This implementation incorporates the abstracted expression wheel while maintaining gaze control and remote expression visibility.

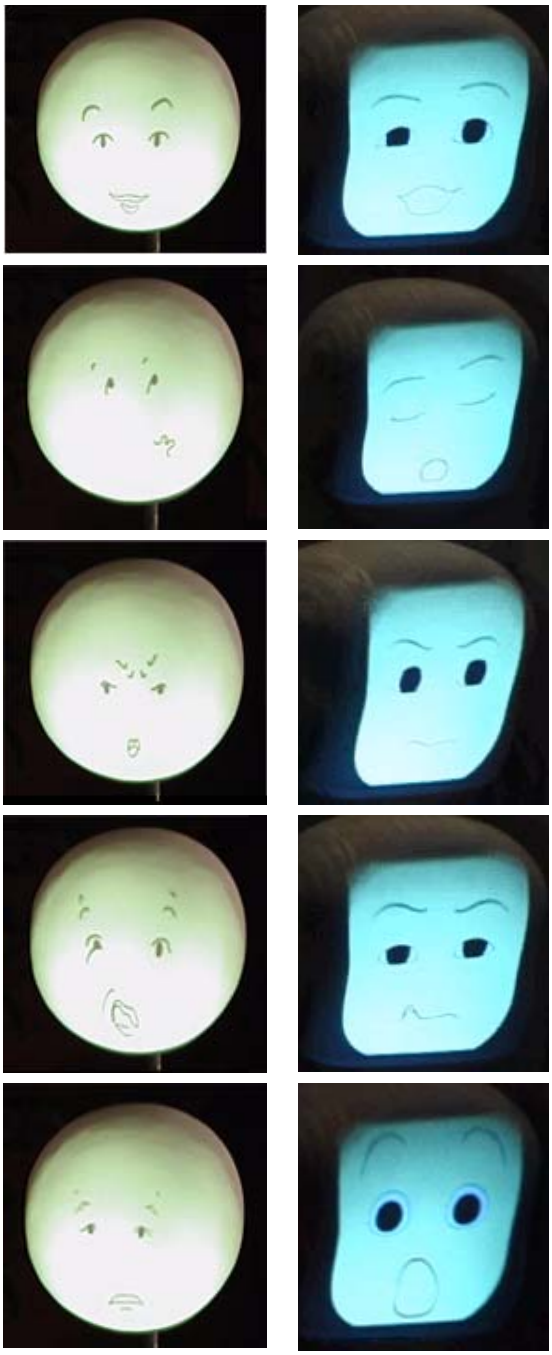


Figure 17. Expression projections on avatar seats. (left column) *Slim* expressions. (from top to bottom) happy, bored, angry, "duh", sad. (right column) selected *Orlando* expressions. (from top to bottom) happy, sleepy, angry, disgusted, surprised.

The abstracted expression wheel allowed the remote user to focus more of their attention on the conversation and less on moving the mouse. There was, however, a trade-off in the expressiveness of the faces.

With the automated expression selection, we were cautious not to deduce a false expression. Hence, the arc of the expressions did not always reach the extremes. In retrospect, it was the extreme, cartoon-like animations that provided more of a catalyst for interaction then the subdued ones.

Figure 18. The *Hydra* four-way teleconferencing system.

Scale in ChitChatClub

ChitChatClub was designed through several iterations. Care was taken to make the avatar seats human scale. If the seat is bigger and looks down on the person, it is intimidating; if it is much smaller, it is often ignored. This way, the remote participant occupied a similar space as the physical participants.

The seat was made to look anthropomorphic. There was a head, a seated body and arms. We did not want it to look so human that participants would expect human attributes, but we also wanted it to be accepted as an interesting seated visitor. The second avatar seat was motorized so the remote user could direct the gaze. This offered more control to the remote user and a focus for attention to the local users at the physical *ChitChatClub.*

Summary

Physicality and human-scale was the central social catalyst of *ChitChatClub*. The physicality makes for interaction far different from what happens while staring at a computer screen. In contrast to the *Hydra* system pictured in Figure 18, the human-scale in *ChitChatClub* provides for gestural behaviors at eye-level and not to several chess-piece-like screens. The similarity in scale blends the physical and virtual worlds together to emphasize togetherness versus remoteness.

A final note on *ChitChatClub*: Although the installation emphasizes togetherness, there is an asymmetry in the public and the private. The local

physical *ChitChatClub* occupants perceive more of the social catalysts and the physicality, although they only see an abstracted representation of the remote cafe-goer. In contrast, the remote user sees a fuller view of the participants albeit at a smaller scale and in a physically remote setting.

CHAPTER 6

Temporality in Space

Time, said St. Augustine, is a threefold present: the present as we experience it; the past as present memory; the future as present expectation. Hopi Indians, who thought of themselves as caretakers of the planet, used only present tense in their language: past was indicated as "present manifested," and the future was signified by "present manifesting." [1] Until approximately 800 B.C., few cultures thought in terms of past or future: all experience was synthesized in the present. It seems that practically everyone but the contemporary man has intuitively understood the space-time continuum. (Youngbood 1970)

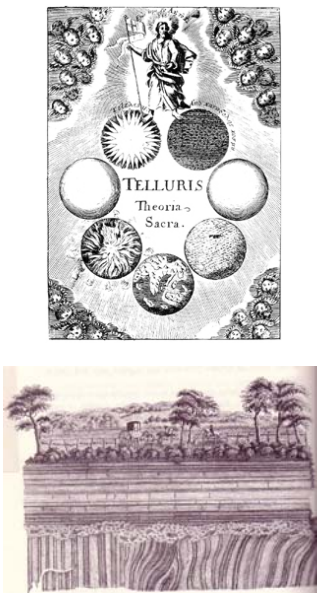


Figure 1. (top) Thomas Burnett's *Telluris*. The illustration of time as a cycle.(bottom) John Clerk's engraving of Hutton's inconformity. Time as linear layers.

The notion of time in communication usually takes two forms — time as a cycle and time as a linear archive. We can envision time as a cycle by looking at the second hand on a traditional clock or by describing the rotation of the earth about the sun. ALternatively, we can envision time as an archive of history. The engraving at the bottom of Figure 1 depicts layers of time of a specific place while the present resides above.

1. Whorf, Benjamin. *Language, Thought and Reality.* Cambridge, MA: Massachusetts Institute of Technology, Publications Office. 1956.

Figure 2. Mixing/Switching control room used at WGBH-TV, used to handle switching for *Hello*. Boston, Massachusetts.

Both of these notions of time have an impact on communication (Ong 2002). Talk is ephemeral — what we say right now may be forgotten immediately or remembered differently by different people. Letters are inherently archived (assuming they are kept).

The same applies in mediated communication. The nature of interaction is affected based on where along the time axis a communication interface lies (Karahalios 2001). Email and Usenet, like letters, are inherently archived. Speech may be archived in a mediated environment, and whether or not it is seriously affects how people use this mediated speech channel.

In terms of history and the linking of spaces and viewing over time, security cameras are probably the closest medium for doing this although the intent is not communication between these spaces. Playing with this phenomena and the elements of randomness and chance, Allen Kaprow created his *Hello* installation (see Figure 2) in Boston for "The Medium is the Message". He interconnected 4 Boston area locations — M.I.T, a hospital, an educational videotape library, and Boston Airport — with 5 TV cameras and 27 monitors. He viewed this as a sociological conduit. When people started talking to a remote person in the library, they might be randomly switched and end up talking to someone in a hospital (Youngblood 1970).

Figure 3. Artifacts of the Presence Era visualization.

A different approach was taken at the MIT Media Laboratory for the Artifacts of the Presence Era project. This project did not connect two spaces, but offered a poetic view more commonly associated with surveillance footage. A snapshot of this can be seen in Figure 3. Time lapse photos were layered using an anthropological metaphor. The sound wave of an interval

near the taking of the photo carved its shape into the photo. The etchings then layer and accumulate upon each other. Older layers became more and more compressed over time.

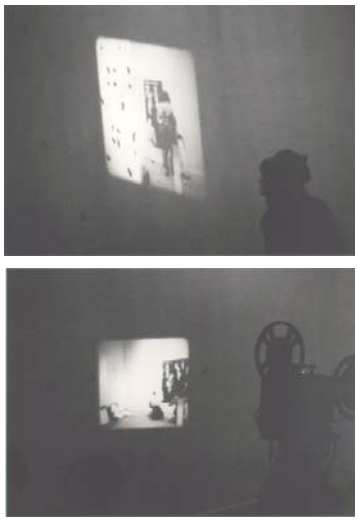


Figure 4. Snapshots from the original *Finch College Project* in 1969.

Carousel

Carousel takes the traditional static media space setup and moves the camera and display about the space. This installation connects two spaces using moving cameras and projectors. The display is all the upright wall space of each respective room. As the projector moves about the space, it paints a sliver of a view through a window of the remote location at that space in time.

Carousel was inspired by Robert Morris's, *The Finch College Project*. In his installation, a film camera was placed on a platform that rotated at one revolution per minute. One side of the room had a life-size black and white panorama of an audience. On the opposite side, a crew of people were constructing and deconstructing a wall of mirrored tiles. This sequence of activities in the room was recorded by the camera. The room was then emptied leaving only dots from the mirror grid alignment. The projector was placed on a platform rotating at the same rate as the film camera and the film of the audience appearing and disappearing as the mirror wall was build and unbuilt, was projected onto the walls of the same room [15].

Carousel uses the visual technique of the Finch College Project to create a communication link between two public spaces. The center of each room has a rotating podium. A camera, microphone, speakers, and a projector sit on each podium. The podiums rotate slowly at the same rate. As they rotate, the images captured as the camera rotates in one room are projected around the space in the other room and vice versa. The result is a moving snapshot it time along the periphery of the room as people, both physically present and projected, are moving about the space.

Carousel is an exploration piece. People go to public spaces to see and be seen. By offering a piece of the whole, with *Carousel*, we hope to entice the viewer into moving about and observing the whole of the space. As people move about each space, they may draw on the walls and leave puzzle

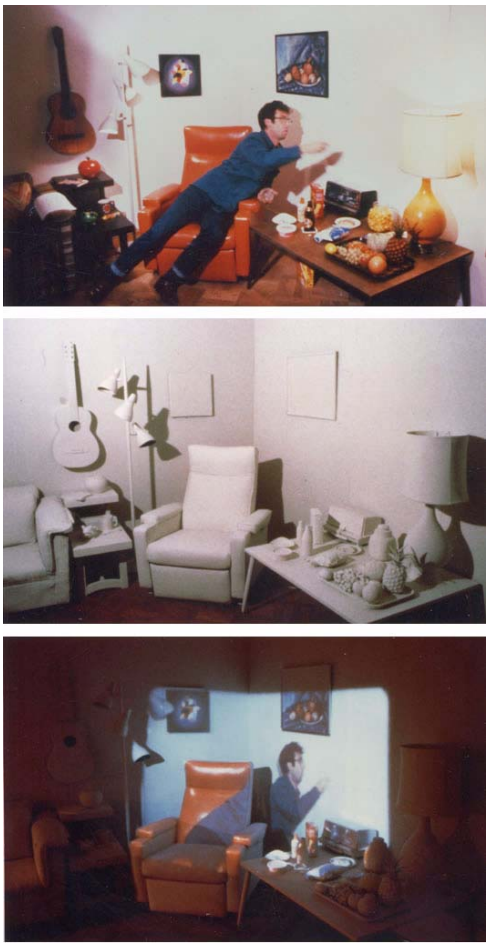


Figure 5. Three snapshots from Michael Naimark's *Displacements*. (top) A film camera was set upon a podium rotating at a steady velocity. With this camera, we filmed two people moving about a room in a 360 degree view.

(middle) The room was then painted white in its entirety.

(bottom) A projector was now placed on the same pedestal that the camera was placed. This pedestal turned at the same rate as the original camera and projected the footage onto the white three-dimensional canvas.

The alignment between the projected image and the furniture recreated the original room. The people took on a "ghostly" effect as they roamed this new space for they were two-dimensional objects moving about this canvas.

pieces. The artifacts and the people from the two spaces blend, resulting in a cooperative, recursive graffiti created by a blended crowd of people both present and remotely located.

The catalyst in this piece is the act of constructing the whole of a new space and crowd from fragments of two spaces. The ideal setting for *Carousel* would be in a museum or gallery. This piece would not work in a generic space. It is less about conversing and more about feeling space and people. People need to have the time to see and explore.

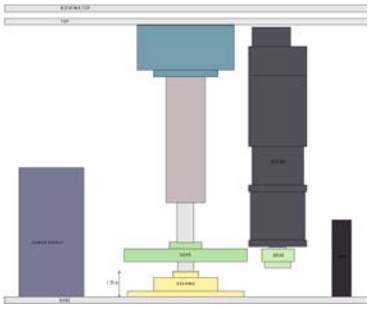


Figure 6. *Carousel* turntable layout as viewed from the side.

Construction

Figure 6 shows the basic layout in the construction of the *Carousel* turntables. For more detailed viewing of the individual components of the turntable, please refer to Appendix B.

Prototype

Figure 8 illustrates the first *Carousel* prototype in one space. It was installed between two conference spaces within the Media Laboratory during several sponsor meetings. During these events, we had to refine the synchroniza-

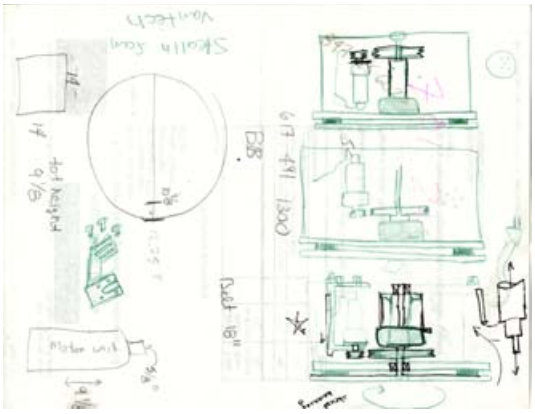


Figure 7. Original *Carousel* turntable design.

Figure 8. *Carousel* Prototype. (left) room A. (middle) room B. (right) room B a moment later.

tion of the turntables, for they tended to drift. We did this by switching from analog components for controlling the motors, to using motor control boards.

Motion and Time as Catalysts

Motion is the strongest visual appeal to attention.

-Rudolf Arnheim

Happenings and Time

Motion and time create the social catalyst in *Carousel*. As in a movie or a play, as we watch, the dramatic conflict unfolds and the narration is revealed to us. It creates in us a feeling of suspense which may not be part of the situation being observed, but a result of the passing of time. In *Carousel*, this is amplified by the movement of the focus of interest.

It is in the experience of passing time in *Carousel* that we understand our local space, the remote space, and the people moving about in both spaces. The result is a moving painting, ever-changing around us.

> "Something very similar is required for true understanding of a symphony, a film, or a dance. At any particular moment we may not know what will come next, but we must not dismiss from our consciousness what we have heard of seen before. The work grows step by step into a whole, and as we accompany its progress we must constantly hark back to what has disappeared from direct perception by ear or eye, but survives in memory. The past as such is never available to the mind. The percepts and feelings, not only of yesterday but of a second ago, are gone. They survive only to the extent that within us they have left remnants, i.e. memory traces. Whatever the nature of these traces in the brain, they certainly persist in spatial simultaneity, influence one another, and are modified by new arrivals." (Arnheim 1974)

Speed

The speed of the rotating turntables affects the experience of the space of *Carousel*. There is a speed at which the inhabitants of the space become comfortable exploring. Too fast — and the space has a dizzying effect. Too slow — and people leave before they realize anything is happening.

We began running the installation with one revolution completing in 60 seconds. This final period of rotation that we used was 176 seconds. Slowing down the angular velocity provided for a smoother flow of the visual and was more comfortable for the inhabitants in the space.

The Installation

The turntables were modified so that they could withstand a greater weight while keeping the same constant rotation. Images using these modified turntables can be seen in Figure 9.

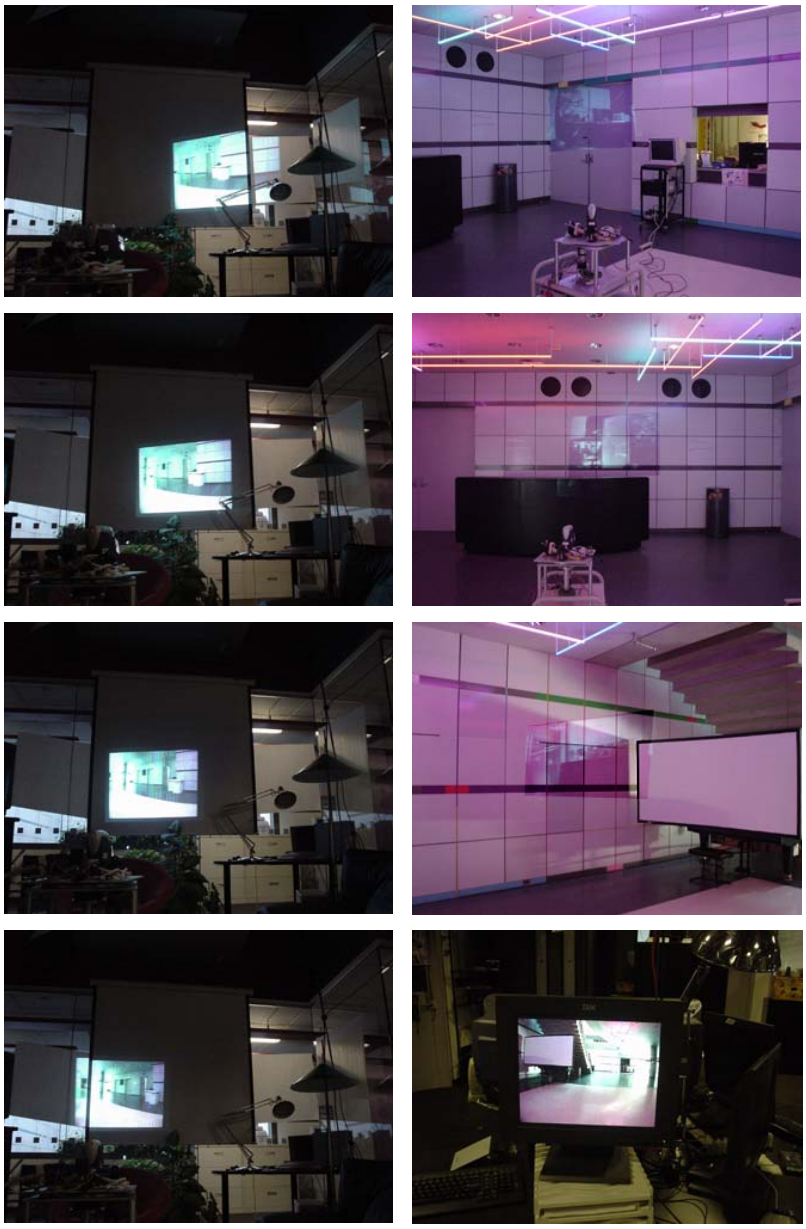


Figure 9. Snapshots from calibrated turntables. (left column) top to bottom: views in space as turntable rotates counter clockwise. (right column) top 3 images: views in larger lobby space as turntable rotates at the same rate and direction as the remote turntable. Note: the proportions and dimensions of the connected spaces are different. At larger distances from the projector, objects locations distort. Given, identical spaces, we achieve space alignment.

There are a few suggested technical improvements, we would recommend for *Carousel*. Among these are:

- Using cameras and projectors with the same field of view and focal lengths.
- Presenting the piece in two remote rooms where portions of the room are of similar size and proportion.

The Interaction

Carousel is an installation about sensory perception in a linked hybrid space. The design and implementation of the engineering components of *Carousel* are complete. The next step is conduct a formal study of the use of the space given different public settings and spatial configurations.

Current interaction using *Carousel* has been observed only during the building and documentation process of the installation. At first glance, people asked if the projection on the wall was a direct feed from the camera directly opposite it on the turntable. The two turntables were set up in lab spaces that had similar decor — couches, computers, etc. and the difference in the spaces was difficult to perceive.

Later, *Carousel* was installed between two distinctly different spaces between the lab — one was a lab space with couches, computers, and a table; the other space was a lower atrium lobby with a very open feel and distinct tiling on the walls that made it obvious to differentiate the first space from the second.

The users of *Carousel* have so far been MIT Media Lab students. People often sit on the couches of the lab space for long periods of time while the installation is running. Oftentimes, they run down to the lower atrium to see what is happening there as well. In the cases, where people meet, they initially feel compelled to walk around and maintain alignment with the camera, before settling down in the space and looking around the whole space. Initial observations suggest, that privacy is not as big a concern as with static cameras. The turntable was designed to be a very "obvious" mechanical device, so that people always know where the camera is. It is

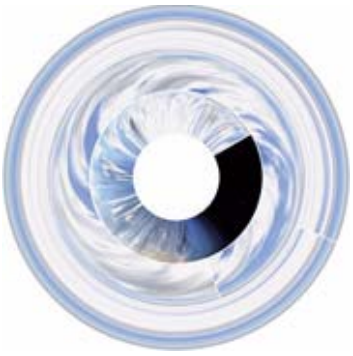


Figure 10. *Last Clock* by Ross Cooper and Jussi Ängeslevä. Instead of displaying the current moment, *Last Clock* shows a time-sliced buffer of one minute, one hour, and 12 hours in the trails of its hands. The clock displays the rhythm and history of the space.

reminiscent of kinetic art constructions and draws people to stare at the inner workings of the moving gears and motor.

We need to conduct interviews and more rigorous observations to understand how people perceive this space — whether they think of it as the local space they are in or a different hybrid space.

Summary

Future experiments with *Carousel* involve manipulating the architectural structure of the surrounding space and conducting observation studies. In the current implementation (as described in Figure 9), the images of the two spaces sync when the walls are at similar distances from the camera and in similar proportions. At other times, they often drift to create a merged dynamic space.

The next version will have two spaces where slivers of wall will match in orientation and distance from the turntable and the remainder of the space will be freeform. People will be able to move behind and around the slivers of fixed matching space and use the remainder of the space as it was used before.

Looking back to the earlier discussion on cyclical time, *Carousel* is reminiscent of a spatial clock. It rotates at a constant rate and shows us changes in lighting and presence during these rotations.

Last Clock in Figure 10 is a visualization of one space over time. Like *Artifacts of the Presence Era*, this is conceptually a linear model of time, although it is represented as circular bands emanating from the center.

In this line of thought, an experimental extension of *Carousel* would be to combine the notion of linear and cyclical time over a long duration into one immersive installation — a *historical carousel* — where one can interact with others in a remote space while perceiving a sense of the history of that space. This would significantly alter the sensory perception of the space. The question, obviously, is how the space would be altered and if it would affect the serendipitous nature of the interface as it is now.

CHAPTER 7

Conclusion

We began this study by looking at examples of mediated communication — old and new. We then looked at features of communication that provide information to both face-to-face and mediated interaction. This study culminated in incorporating this information — the social cues, feedback, etc. — into the space, interface, and language of our computer-mediated spaces to create a catalytic space.

In this dissertation, we explored the creation of social catalysts along four different axes:

- Transformation of Space through Interaction
- Abstraction for Visualizing Conversation
- Physical Manifestation of Virtual Presence
- Time and Motion for Blended Spaces

It may have seemed arbitrary earlier why we had chosen these four parameters to frame our study. The reason for their selection is because they are all evident in our project landscape as we study social catalyst, with one being the predominant feature of each. They are the principle components our social catalyst matrix.

Figure 1. The four features used to explore our social catalyst. They became Chapters 3,4,5, and 6.

Creation of Hybrid Spaces

One of the major inspirations for creating these installations is the blending of the online and the physical. From the online world, we wanted to maintain the ease of interaction with strangers, the ease of entry into public spaces, and the low risk associated with interaction. From the physical world, we wanted to maintain the affordances (navigation, sense of touch, visual field, etc.) of the physical space surrounding us, the presence of physical, tangible forms, and the sensory perception of space.

In creating these hybrid worlds, we no longer interact as we do in an online-only space or in a physical-only space. We have created a new style of space that brings people together in unfamiliar and serendipitous ways.

Why is this important? Connections exist all around us and given the market for technology, they will continue to increase. There is already too much information to comprehend in online public spaces — too many people to take notice of in physical public spaces. That is fine, and that works. When we combine them, the connections multiply exponentially.

With the four principles we address in this study: transformation, abstraction, physicality, and motion, we create social catalysts that allow one to navigate this hybrid space and focus on aspects of interaction that we have designed instead of all the surrounding signals. We provide a focus for attention, a focus for presence, and a recreation space in a social link that would otherwise be undecipherable. The transformation highlights changes in the interaction —the abstraction eliminates *extra* information — the physicality provides for an object to look at and manipulate — motion draws our attention and guides us about the space.

The Role of Recreation and Play in the Interface

All of these projects were designed with catalysts for interaction. The social catalysts often inspired play and recreation on top of interaction. These interfaces were designed with social interaction in mind above completing work-oriented tasks.

Telemurals differed from the direct audio-video connections we see in *Media Spaces* and similar teleconferencing systems in style of interface and in style of interaction. In work-oriented tasks involving teleconferencing systems, social banter and chit chat are essential in establishing a common framework and trust. Current video-conferencing interfaces don't facilitate this. *Telemurals* facilitates these social mores. The transformative silhouette interface designed in *Telemurals* engaged people to gesture and talk in ways that are not customary in straight audio-video connections.

Visiphone encouraged people to talk and to perform in ways to create intended patterns on the *Visiphone* dome. In *ChitChatClub*, play was involved in creating the faces and a exploring a distant space for the remote user. The local users had a social catalyst in the avatar seat to link them and a new communication game in interacting with the avatar face and body. The remote users controlled their exploration of a new space.

Finally, the *Carousel* space bridged together the sensory experience from two spaces through the use of a kinetic sculpture. The sculpture was a focal point in each space that bound the two spaces together. It encouraged people to walk/run about the space to see the revealing portions, to follow the gaze of the turntable, and to linger in a space expecting the unexpected. The social catalysts of these installations attracted people to the common spaces and created interaction that was not previously possible.

Closing Thoughts

We can consider these interfaces as public interrogations of self and communication. I want to stress the need to make these installations public and available so as to provide this revelation from social catalysts.

Communication Spaces as Public Art

Public sculpture has some social function. It has moved from large scale, outdoor, site-specific sculpture into sculpture with social content. In the process, it has annexed a new territory for sculpture that extends the field for social experience. (#9)

Public sculpture is not artistic creation alone but rather social and cultural productions based upon concrete needs. (#11)

Public sculpture is a cooperative production. There are others besides the artist who are responsible for the work. To give all the credit to the individual artist is misleading and untrue. (#12)

We enter public sculpture not as a thing between four walls in a spatial sense but as a tool for activity. (#15)

Public sculpture depends upon some interplay with the public based upon some shared assumptions. (#22)

Public sculpture should not intimidate, assault or control the public. It should enhance a given space. (#24)

MANIFESTO
Siah Armajani, Reading Spaces

Appendix A — Telemurals Data

Week 2[1]

Observations from Abstracted Video (122 people)

	Look	Wave	Encounter	Talk	Play
2hr/night	.98	.45	.07	.02 / 0.0	.44 / .05

Observations from Immersive Observations (41 people)

	Look	Wave	Encounter	Talk	Play
2hr/night	1.0	.73	.12	.10 / .02	.66 / .07

Observations from Photorealistic Observations (29 people)

	Look	Wave	Encounter	Talk	Play
2hr/night	.97	.48	not known	.03 / not known	.24 / not known

Week 4

Observations from Abstracted Video (43 people, series of study breaks at one end)

	Look	Wave	Encounter	Talk	Play
2hr/night	.93	.81	.44	.16 / .12	.40 / .30

Observations from Immersive Observations (61 people, includes night of wedding)

	Look	Wave	Encounter	Talk	Play
2hr/night	1.0	1.0	.32	.30 / .16	.31 / .28

1. In a table entry of the form '**a** / **b**', **a** is the percentage of people who spoke to someone at the other end. **b** is the percentage of the people who spoke for over a minute.

Observations from Photorealistic Observations (13 people)

	Look	Wave	Encounter	Talk	Play
2hr/night	.85	.70	not known	.08 / not known	.08 / not known

Week 6

Observations from Abstracted Video (136 people)

	Look	Wave	Encounter	Talk	Play
24hrs/7 nights	.88	.80	.15	.15 / .04	.15 / .09

Observations from Immersive Observations (71 people)

	Look	Wave	Encounter	Talk	Play
24hrs/7 nights	.94	.45	.11	.10 / .01	.11 / .06

Observations from Photorealistic Observations (52 people)

	Look	Wave	Encounter	Talk	Play
24hrs/7 nights	.62	.29	not known	.10 / not known	.08 / not known

Week 8

Observations from Abstracted Video (271 people)

	Look	Wave	Encounter	Talk	Play
24hrs/7 nights	.84	.61	.21	.31 / .12	.36 / .19

Observations from Immersive Observations (69 people)

	Look	Wave	Encounter	Talk	Play
24hrs/7 nights	.90	.80	.45	.29 / .22	.42 / .22

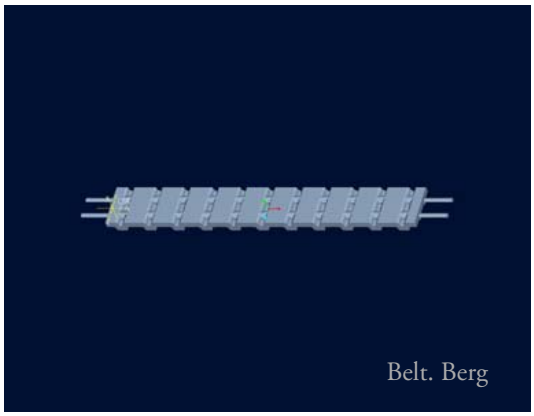
Observations from Photorealistic Observations (32 people)

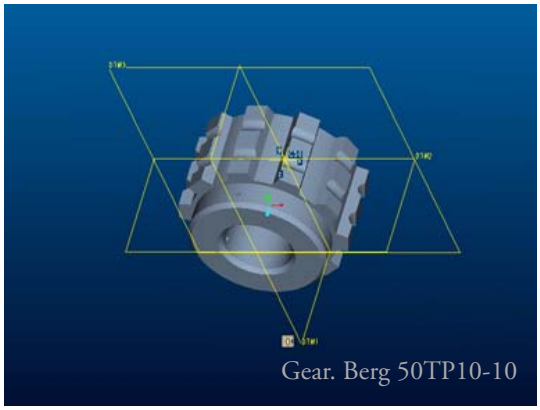
	Look	Wave	Encounter	Talk	Play
24hrs/7 nights	.84	.63	not known	.56 / not known	.25 / not known

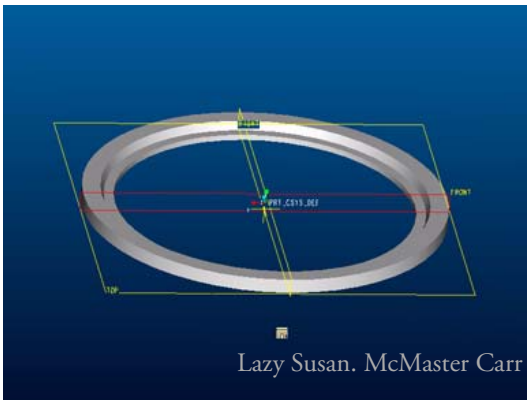
Comments

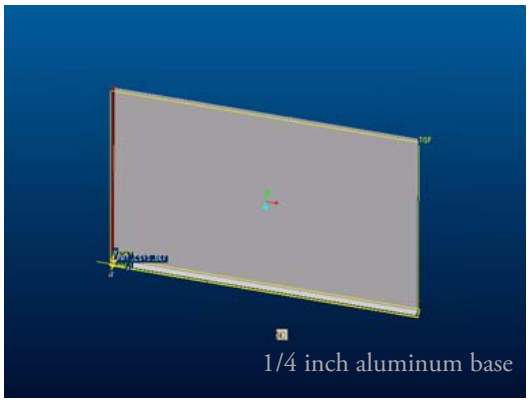
- Observations were run at random times while the system was running.
- Look: passerby glanced at the *Telemural*
- Wave: passerby gestured or waved hands at the *Telemural*
- Encounter: passerby was at one end of the *Telemural* when someone approached the *Telemural* at the other end
- Talk: passerby spoke into the Telemural
- Play: passerby showed signs of play — kicked, jumped, danced, exaggerated gesture, spun around, etc.
- In a table entry of the form 'a / b', a is the percentage of people who spoke to someone at the other end. b is the percentage of the people who spoke for over a minute.

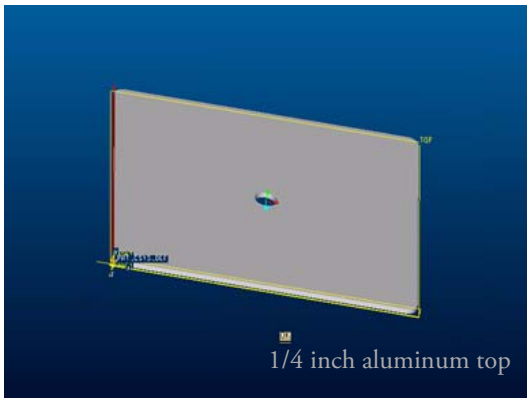
Appendix B — Modeled Carousel Components

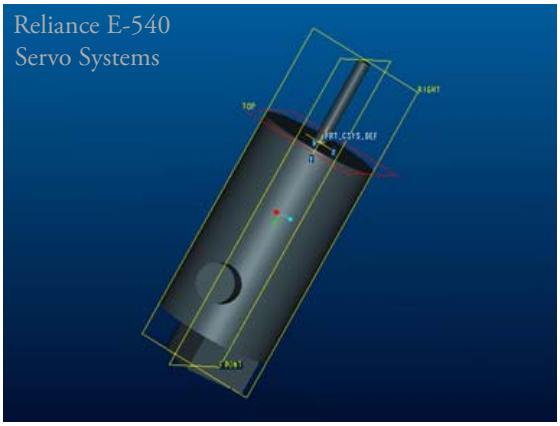


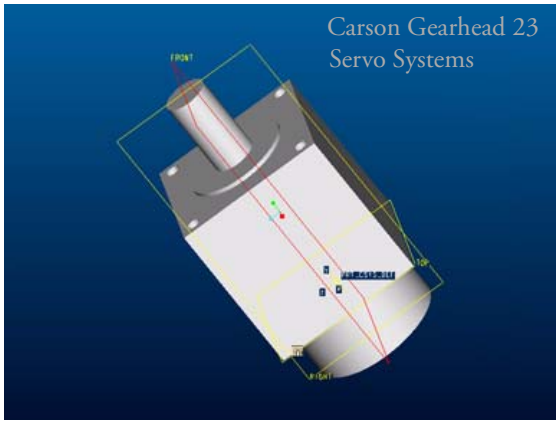


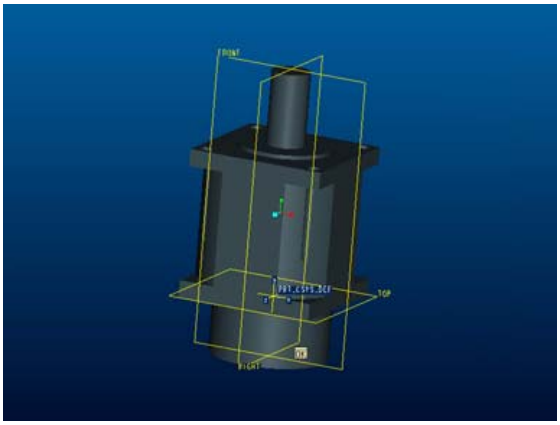


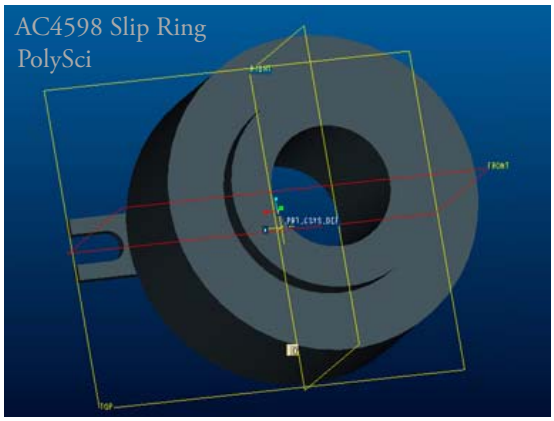


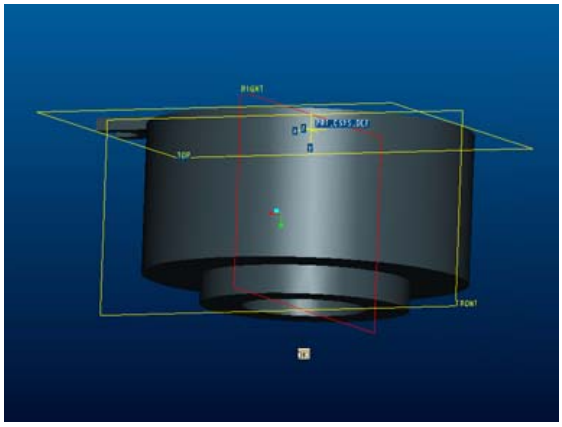


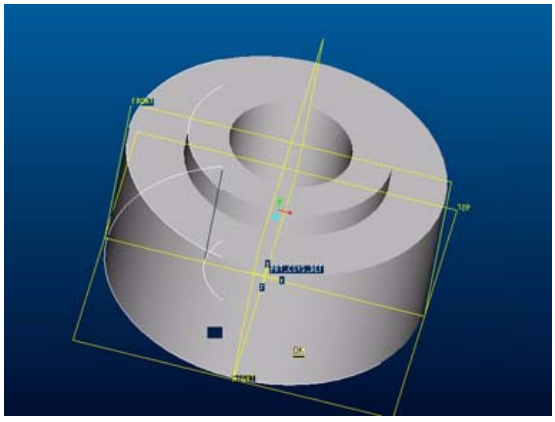

Reliance E-540
Servo Systems


Carson Gearhead 23
Servo Systems




AC4598 Slip Ring
PolySci





Bibliography

Agamanolis et al 1997 Agamanolis, S., Westner, A. and Bove, V.M. Reflection of Presence: Toward More Natural and Responsive Telecollaboration. *Proc. SPIE Multimedia Networks*, 3228A, 1997.

Arnheim 1974 Arnheim, Rudolf. *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley: University of California Press. 1974.

Arnheim 1969 Arnheim, Rudolf. *Visual Thinking*. Berkeley: University of California Press. 1969.

Barlow et al 1990 Barlow, H., Blakemore, C., and Weston-Smith, M., eds. *Images and Understanding*. Cambridge: Cambridge University Press. 1990.

Bly and Irwin 1993 Bly, S. and Irwin, S. Media Spaces: Bringing people together in a video, audio and computing environment. *Comm. ACM* 36,1, 28-47, 1993.

Boyle et al 2002 Boyle, M., Edwards, C. and Greenberg, S. The Effects of Filtered Video on Awareness and Privacy. *Proceedings of CSCW 2002*.

Buxton 1997 Buxton, W. Living in Augmented Reality: Ubiquitous Media and Reactive Environments. *Video-Mediated Communication*. 1997.

Buxton 1992 Buxton, W. Telepresence: integrating shared task and person spaces. *Proceedings of Graphics Interface 1992*.

Clifford and Marcus 1986 Clifford, J. and Marcus, G.E. *Writing Culture: The Poetics and Politics of Ethnography.* Berkeley: University of California Press, 1986.

Dourish et al 1996 Dourish, P., Adler, A., Bellotti, V. and Henderson, A. Your place or mine? Learning from long-term use of audio-video communication. *Computer Supported Cooperative Work*, v.5 n.1, p.33-62, 1996.

Dourish and Bly 1992 Dourish, P. and Bly, S. Portholes: Supporting awareness in a distributed work group. *CHI1992 Conference Proceedings*. New York: ACM Press.

Durlach 1999 Durlach, Nathaniel, I. *PRESENCE: Teleoperators and Virtual Environments*. Volume 8. Number 1. February 1999.

Eckman 2003 Eckman, Paul. *Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Times Books. 2003.

Finn et al 1997 Finn, K., Sellen, A. and Wilbur, S., eds. *Video-mediated communication*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

Fisher 1992 Fischer, Claude S. *America Calling: a social history of the telephone to 1940*. Los Angeles: University of California Press, 1992.

Galloway and Rabinowitz 1981 Galloway, K. and Rabinowitz, S. Hole in Space, 1981. http://www.ecafe.com/getty/HIS/

Goffman 1963 Goffman, Erving. *Behavior in Public Spaces: Notes on the social organization of gathering*. New York: The Free Press. 1963.

Goffman 1967 Goffman, Erving. *Interaction Ritual — Essays on Face-to-Face Behavior*. New York: Pantheon Books. 1967.

Goffman 1959 Goffman, Erving. *The Presentation of Self in Everyday Life*. New York: Doubleday. 1959.

Gooch and Gooch 2001 Gooch, Bruce and Gooch, Amy. *Non-Photorealistic Rendering*. Natick, MA: AK Peters. 2001.

Gould 1987 Gould, Stephen Jay. *Time's Arrow Time's Cycle: Myth and Metaphor in the Discovery of Geological Time*. Cambridge, MA: Harvard University Press. 1987.

Grudin 1988 Grudin, Jonathan. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. *Proceedings of CSCW 1988.*

Grundmann 1984 Grundmann, Heidi. *Art + Telecommunication*. Vancouver: Western Front, 1984.

Gumpert and Cathcart 1979 Gumpert, Gary and Cathcart, Robert, eds. *INTER/MEDIA: Interpersonal Communication in a Media World*. New York: Oxford University Press. 1979.

Hall 1981 Hall, Edward, T. *Beyond Culture*. New York: Anchor Books. 1981.

Hall 1982 Hall, Edward, T. *The Hidden Dimension*. New York: Anchor Books. 1982.

Hall 1981b Hall, Edward, T. *The Silent Language*. New York: Anchor Books. 1981.

Hauser 1997 Hauser, Marc D. *The Evolution of Communication*. Cambridge, MA: The MIT Press, 1997.

Harrison and Dourish 1996 Harrison, S. and Dourish, P. Re-place-ing Space. *Proceedings of CSCW 1996*.

Hindus et al 1996 Hindus, D., Ackerman, M., Mainwaring, S., and Starr, B. Thunderwire: A Field Study of an Audio-Only Media Space. *CSCW 1996*.

Hollan and Stornetta 1992 Hollan, J.and Stornetta, S. Beyond Being There. *Proceedings of CHI 1992*.

Horrigan 2000 Horrigan, Bill. *Shirin Neshat: Two Installations*. The Ohio State University: Wexner Center for the Arts, 2000.

Isaacs and Tang 1993 Isaacs E. and Tang J. What Video Can and Can't do for Collaboration: A Case Study. *Multimedia '93*.

Ishii et al 1992 Ishii, H., Kobayashi, M. and Grudin, J. Integration of inter-personal space and shared workspace: ClearBoard design and experiments. *Proceedings of CSCW 1992.*

Jacobs 1993 Jacobs, Jane. *The Death and Life of Great American Cities*. New York: The Modern Library, 1993.

Jancke et al 2001 Jancke, G., Venolia, G., Grudin, J., Cadia, J. and Gupta, A. Linking Public Spaces: Technical and Social Issues. *Proceedings of CHI 2001*.

Karahalios 2001 Karahalios, Karrie. Communication Systems: A comparison along a set of major axes. http://web.media.mit.edu/~kkarahal/generals/communication/index.html. 2001.

Karahalios and Donath 2003 Karahalios, Karrie and Donath, Judith. Scale, Form, and Time: creating connected sociable spaces. Ubicomp 2003.

Karahalios and Donath 2004 Karahalios, Karrie and Donath, Judith. Telemurals: Linking Remote Spaces with Social Catalysts. *CHI 2004*.

Kiesler et al 1984 Kiesler, S., Siegel, J., and McGuire, T. Social Psychological Aspects of Computer-Mediated Communication. *American Psychologist,* vol. 39, no. 10, 1123-1134. October 1984.

Koolhaas 1994 Koolhaas, Rem. *Delirious New York: A Retroactive Manifesto for Manhattan*. New York: The Monacelli Press. 1994.

Krueger 1991 Krueger, M. *Artificial Reality II*. Addison-Wesley, 1991.

Kubovy 1986 Kubovy, Michael. *The Psychology of Perspective and Renaissance Art*. Cambridge: Cambridge University Press. 1986.

Lynch 1972 Lynch, Kevin. *What Time is this Place?* Cambridge, MA: The MIT Press. 1972

Mackay 1999 Mackay, W. Media Spaces: Environments for Informal Multimedia Interaction. In *Computer Supported Co-operative Work*. John Wiley and Sons. 1999.

Matheson 1991 Matheson, K. Social Cues in Computer-Mediated Negotiations: Gender Makes a Difference. *Computers in Human Behavior*, vol. 7, 137-145. 1991.

McCloud 1993 McCloud, S. *Understanding Comics*. New York: Kitchen Sink Press, 1993.

Meyrowitz 1985 Meyrowitz, Joshua. *No Sense of Place: The Impact of Electronic Media on Social Behavior*. Oxford: Oxford University Press. 1985.

Milgram 1992 Milgram, Stanley. The Individual in a Social World: Essays and Experiments. New York: McGraw-Hill, Inc. 1992.

Mitchell 1995 Mitchell, William. *City of Bits*. Cambridge, MA: The MIT Press. 1995.

Olson et al 1995 Olson, J., Olson, G,. and Meader, D. What mix of video and audio is useful for small groups doing remote design work? *CHI1995*.

Ong 2002 Ong, W. J. Orality and Literacy. New York: Routledge. 2002.

Pederson and Sokoler 1997 Pederson, E.R. and Sokoler, T. AROMA: abstract representation of presence supporting mutual awareness. *Proceedings of CHI 1997*.

Pool 1977 Pool, Ithiel de Sola, ed. *The Social Impact of the Telephone*. Cambridge, MA: The MIT Press, 1977.

Rosen 2000 Rosen, Jeffrey. The Unwanted Gaze. The Destruction of Privacy in America. New York: Random House. 2000.

Saville-Troike 1989 Saville-Troike, Muriel. The Ethnography of Communication: An Introduction. New York: Basil Blackwell, Inc. 1989.

Scherer 1986 Scherer, K.R. Vocal Affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165. 1986.

Singer et al 1999 Singer, A., Hindus, D., Stifelman, L., and White, S. Tangible Progress: Less is more in Somewire audio space. *Proceedings of CHI'99*.

Smith and Hudson 1995 Smith, I. and Hudson, S. Low disturbance audio for awareness and privacy in media space applications. *Proceedings of Multimedia '95*.

Smith and Scott 1997 Smith, Craig A. and Scott, Heather S. A Componential Approach to the meaning of facial expressions. *The Psychology of Facial Expression*, eds. Russel, James A. and Fernando-Does, Jose Mogul. New York: Cambridge University Press, 1997.

Spears and Lee 1994 Spears, R. and Lea, M. Panacea or Panopticon? The Hidden Power in Computer-Mediated Communication. *Communication Research*, vol. 21, no. 4, 427-459. August 1994.

Sproull and Kiesler 1991 Sproull, Lee and Kiesler, Sara. *Connections: New ways of Working in the Networked Organization*. Cambridge, MA: The MIT Press. 1991.

Standage 1998 Standage, Tom. *The Victorian Internet*. New York: Berkley Books. 1998.

Tang and Minneman 1991 Tang, J. and Minneman, S. VideoWhiteboard: video shadows to support remote collaboration. *Proceedings of CHI 1991*.

Venkatesh 1996 Venkatesh, A. Computers and other interactive technologies for the home. *Communications of the ACM*, 39:10.

Viégas and Donath 1999 Viégas, Fernanda and Donath, Judith. Chat Circles. *CHI1999*.

Walther 1996 Walther, Joseph. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, vol. 23, no. 1, 3-43, February 1996.

Wilson 2002 Wilson, Stephen. *Information Arts*. Cambridge, MA: The MIT Press, 2002.

Whittaker 1995 Whittaker, Steve. Rethinking Video as a technology for interpersonal communication: Theory and design implications. *International Journal of Human-Computer Studies, 42*, 501-529. 1995.

Whyte 1984 Whyte, William Foote. *Learning From the Field*. London: Sage Publications, Inc., 1984.

Whyte 1988 Whyte, William H. *City: Rediscovering the Center*. New York: Doubleday, 1988.

Whyte 1980 Whyte, William H. *The Social Life of Small Urban Spaces*. New York: Project for Public Spaces, 1980.

Youngblood 1970 Youngblood, Gene. *Expanded Cinema*. New York: E.P. Dutton & Co. 1970.

Personal Conversation 2002 Personal communication between author and public art critic. 2002.