

# Talking in Circles:

Representing Place and Situation In an Online Social Environment

**Roy Alexis Rodenstein Kartoffel**

B.S. Computer Science, Georgia Institute of Technology, 1997

M.S. Computer Science, Georgia Institute of Technology, 1998

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

June 2000

© Massachusetts Institute of Technology, 2000  
All Rights Reserved

---

**Author**

Roy Alexis Rodenstein Kartoffel  
Program in Media Arts and Sciences  
May 15, 2000

---

**Certified by**

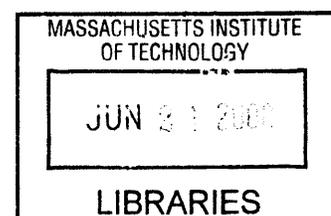
Judith Donath  
Assistant Professor of Media Arts and Sciences  
Thesis Supervisor

---

**Accepted by**

Stephen A. Benton  
Chair, Departmental Committee for Graduate Students  
Program in Media Arts and Sciences

**ROTCH**



# Talking in Circles:

## Representing Place and Situation In an Online Social Environment

Roy Alexis Rodenstein Kartoffel

Submitted to the Program in Media Arts and Sciences,

School of Architecture and Planning, on May 15, 2000

in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences at the Massachusetts Institute of Technology

### Abstract

This thesis presents work focused on the creation of a sociable space for communication online. Sociable communication requires the ability to converse with others using simple and meaningful mechanisms, supporting flexibility and expressiveness. Equally important is the ability for people to read the space they inhabit and make sense of it in socially significant ways, such as people-watching to observe others' interests and interaction styles. A third key to sociable communication is emphasis on identity and embodiment, giving participants a strong sense of themselves and others through their online representations.

These issues are approached through research in areas ranging from sociology to urban architecture, directed at finding bases for the design of capabilities that are useful and engaging in the context of computer support for distributed multiparty communication. The result of this research is *Talking in Circles*, a graphical audioconferencing environment that employs abstract graphics for representation and provides lightweight access to multiple expressive modes. This thesis discusses foundations for work towards sociable communication online as well as the design and implementation processes involved in the creation of the *Talking in Circles* system. User experiences with the system, lessons learned and directions for further research into sociable communication are then detailed.

Thesis Supervisor:

Judith Donath

Assistant Professor of Media Arts and Sciences

## Thesis Committee

---

**Thesis Supervisor**

Judith Donath

Assistant Professor of Media Arts and Sciences

Massachusetts Institute of Technology

---

**Thesis Reader**

Chris Schmandt

Principal Research Scientist

Massachusetts Institute of Technology

---

**Thesis Reader**

Elizabeth D. Mynatt

Assistant Professor

Georgia Institute of Technology

## Acknowledgements

I would like first to thank Judith Donath, my advisor over these two years at the Media Lab. Her broad range of knowledge and interests has been invaluable in the multidisciplinary work I have been involved in at the Sociable Media group, and her guidance as understanding as it has been ambitious.

My thesis readers, Beth Mynatt and Chris Schmandt, along with my advisor, provided very helpful comments on both my thesis project and this document. The complex concepts I have tried to discuss are, I trust, much clearer thanks to their feedback.

Special thanks go to fellow members of the Sociable Media Group. Karrie Karahalios, David Chiou, Dana Spiegel, Rebecca Xiong, Fernanda Viegas, Joey Rozier, Matt Lee and Brian Bower others, have been wonderful to work with, friendly and creative. Lisa Lieberson has been cheerfully helpful, even during the most hectic times.

Past colleagues –mentors and friends– have been instrumental in helping me on the road of academic research, among them Scott Hudson, Gregory Abowd, Keith Edwards, Steve Greenspan. Thanks to Ian Smith, especially, for showing me the value of a sense of history.

I'm grateful to the Media Lab student community at large. The variety of their expertise and their willingness to share it with their fellow students is simply heartening. Bakhtiar Mikhak and Rich Fletcher, in particular, were extremely helpful with the musical puzzle and the window display projects, respectively. To Luke Weisman, Kelly Dobson and Tamara Lackner, thanks for the best group-project experience I've ever had.

To the members of the Aesthetics and Computation Group for their zaniness, to the members of the Tangible Media Group, to Todd Machover for the best school field-trip ever, to Brian Smith for his frankness, to Erik Blankinship and Tamara Lackner for their kookiness and world-class good humor, to many others, thanks for keeping things interesting.

Finally, to my family –Ari, Mario, Graciela, Ñato, Dorita, Fany, Juan–, gracias.

# Contents

<b>Abstract</b>	<b>2</b>
<b>Thesis Committee</b>	<b>3</b>
<b>Acknowledgements</b>	<b>4</b>
<b>Contents</b>	<b>5</b>
<b>List of Figures</b>	<b>7</b>
<b>1 Introduction</b>	<b>9</b>
1.1 Motivation	10
1.2 System Overview	11
1.3 Organization of this Document	11
<b>2 Foundations for the Design of an Online Sociable Environment</b>	<b>13</b>
2.1 Communication Channels and Media	13
2.1.1 Text-based Communication	14
2.1.2 Graphical Text-chat Environments	15
2.1.3 Video-based Media Spaces	18
2.1.4 Audio as Primary Communication	20
2.2 Socio-spatial Grounding	22
2.2.1 Avatar Proxemics in Graphical Text Chat	22
2.2.2 Milgram's Studies of Crowd Formations	24
2.2.3 Whyte's Sociological Studies of Urban Architecture	25
<b>3 The Design Process of <i>Talking in Circles</i></b>	<b>28</b>
3.1 Speaker Identification	28
3.2 Spatialization	30
3.3 Circle Motion	36
3.4 Drawing and Gesture	36
3.5 Sound Booths	40
<b>4 The Implementation of <i>Talking in Circles</i></b>	<b>43</b>
4.1 Design Requirements and Iterative Prototypes	43
4.2 Shared and Subjective System Rendering	45
4.3 Basic <i>Talking in Circles</i> Graphical Elements	47
<b>5 User Experiences</b>	<b>50</b>

5.1 Evaluation of New Social Environments	50
5.2 Encouraging Experiences	51
5.3 Problems and Limitations	52
5.4 Specialized Users	53
<b>6 Summary and Conclusions</b>	<b>55</b>
<b>7 Future Directions</b>	<b>58</b>
7.1 Likely Fruitful Directions	58
7.2 Other Potential Directions	59
<b>Bibliography</b>	<b>61</b>

## List of Figures

<b>Figure 1.1</b> p. 12	Six participants conversing in Talking in Circles
<b>Figure 3.1</b> p. 29	Diagram of several moments' sample speech between James and Alice and corresponding graphical feedback
<b>Figure 3.2</b> p. 31	Output volume as a function of distance from a speaker
<b>Figure 3.3</b> p. 32	Diagram of sample speech between James and Alice and corresponding perceived speech volume
<b>Figure 3.4</b> p. 34	Three conversations in a Talking in Circles chat session. Al and Sandy are speaking, while Josh draws for Yef
<b>Figure 3.5</b> p. 34	Screenshot from the point of view of Al, the blue circle, as Sandy and Andy are speaking. Andy is within Al's hearing range, while Sandy is now outside it
<b>Figures 3.6a, 3.6b, 3.6c</b> p. 37	Participants' representations cannot overlap each other in Talking in Circles. Here, Albert approaches Norman, then moves swiftly around him, akin to crowd motion
<b>Figures 3.7a, 3.7b</b> p. 40	Two examples of drawings from Talking in Circles
<b>Figure 3.8</b> p. 41	Six participants conversing in Talking in Circles. Allen, Jackie and Sam are listening at the music booth, Kurt at the news booth
<b>Figure 4.1</b> p. 44	Typical desktop setup for a Talking in Circles session
<b>Figure 4.2</b> p. 46	Talking in Circles system diagram showing an overview of network transport of data and its rendering for participants
<b>Figure 4.3</b> p. 47	Talking in Circles login panel
<b>Figures 4.4a, 4.4b</b> p. 48	The music and sound booths

**Figures 4.5a, 4.5b**  
p. 48

Built-in icon bar and how each appears when displayed by a participant

**Figures 4.6a, 4.6b**  
p. 49

Albert's circle as it appears in Talking in Circles when not speaking (left) and when speaking at normal volume (right)

**Figure 4.7**  
p. 49

RAT control panel. Though not generally used during a Talking in Circles session, this control panel permits adjustment of the local microphone and speaker amplification

## 1 Introduction

*People's movements are one of the great spectacles of a plaza ...  
down at eye level the scene comes alive* – William H. Whyte

Computing represents a great and growing resource for human communication. Broad adoption of personal computing and, as importantly, adoption of the value of interconnectedness among computers has resulted in unprecedented possibilities for socialization online. Individuals and groups can increasingly reach others using novel digital media far more pliant than the traditional telephone or fax. The number of participants in a communicative act and their means for expression are no longer a fixed factor of a singular physical medium but a malleable function of the particular online environment they populate.

This thesis presents one point in the design space of online environments for sociable communication, inspired by the vitality Whyte describes. The design discussed herein is the result of exploratory research based in domains ranging from sociology to architecture and aimed at creating an amenable, engaging and flexible online space for people to communicate. The result of this work is *Talking in Circles*, a graphical audioconferencing environment that employs abstract graphical representation and allows lightweight access to multiple modes of expression.

## 1.1 Motivation

Recent years have seen an explosion in online communication. While asynchronous media such as electronic mail and USENET news continue to thrive, synchronous communication has emerged with vast popularity. Purely textual exchanges via Internet Relay Chat (IRC) and Multi-user Dungeons (MUDs) as well as chat rooms on the World Wide Web have become common avenues for group discussions and socialization, and environments with strong graphical components, such as *The Palace*<sup>1</sup>, are now commonplace. While much work has been done on task- and entertainment-oriented communicative systems, there remains a great need for research on the *sociable* aspects of online communication, the focus of this thesis work.

There are some notable systems that strive to support socialization. IBM's *Babble* [Erickson 1999-S], for example, is a graphical chat environment that gives participants a notion of others' interest in various conversations. Its focus generally remains on work, however; for example, discussions are segregated by named topic and participants are channeled into isolated rooms based on the discussion they explicitly choose to join. Work in rich media such as video spaces has also enhanced possibilities for social communication, due to video's ability to display gestures and facial expressions, but has similarly focused largely on office settings and a set of participants that share at least some pre-existing affiliation [Fish 1992, Gaver 1992].

In recent years, non-task-oriented approaches to online socialization have mushroomed. Systems emphasizing the use of graphics for entertaining discussions, such as *Virtual Places*<sup>2</sup> and Microsoft *V-Chat*<sup>3</sup>, afford the use of interesting graphical avatars for participants' representations. These avatars tend to be broadly caricatured, which can have a great (sometimes) unintended impact on interpretation of the caricature's messages [Donath 1999]. Further, participants often take advantage of the ability to switch avatars [Smith 2000] which does not necessarily help others maintain a consistent social grasp of them.

---

<sup>1</sup> <http://www.thepalace.com>

<sup>2</sup> <http://www.vplaces.net>

<sup>3</sup> <http://vchat.microsoft.com>

Though these systems collectively address certain issues relevant to sociable communication, they largely regard socialization as a means to an end, whether that end be entertainment or productive work. This thesis treats sociable communication as the primary goal, with sustained, lively and flexible discussion central objectives without regard to secondary ends. Important challenges exist for such a system, such as questions about the representation of the online environment and its participants to create a *legible* social space.

## **1.2 System Overview**

*Talking in Circles*, extending the design of the Sociable Media Group's *Chat Circles* graphical chat environment [Viegas 1999], addresses these challenges through the combination of graphics with audio for communication. Participants in a *Talking in Circles* session are represented as circles of various colors within a space and can move around this space and speak with each other via a microphone attached to their computer. They can also draw on their circle and use icons for reactive expression (see Figure 1.1). These capabilities, among others, are designed to allow for variety in communication and the ability to navigate the social space of participants.

## **1.3 Organization of this Document**

This thesis discusses the design and implementation processes of *Talking in Circles* as well as experiences with the system. Chapter 2 first sets the background through an elaboration of the foundations of this work in areas such as media spaces and urban architecture. Chapter 3 then describes in full the design of the system, detailing the various aspects and their rationale as based on relevant studies and related research. The implementation, including iterations and refinements, is covered in Chapter 4. Chapter 5 notes a range of user experiences in freeform use of the system, followed by general discussion and conclusions in Chapter 6. Finally, Chapter 7 illustrates potentially fruitful directions for extensions of this work..



Figure 1.1 Six participants conversing in *Talking in Circles*

## 2 Foundations for the Design of an Online Sociable Environment

*Any crowd can be seen as a group of points coming into aggregation ... evolving new shapes, and possessing certain self-distributing dynamics* – Stanley Milgram

The Introduction briefly discussed the work in this thesis as focusing exclusively on facilitating sociable communication, without specific regard to task- or entertainment-oriented aspects, but noted that work in these domains can still be of relevance. This chapter expounds on the foundations for the design of the *Talking in Circles* system –including work in Computer-Supported Cooperative Work (CSCW), chat systems, urban sociology and architecture, and studies of conversation over various media– highlighting both lessons and challenges gleaned. Some of the approaches *Talking in Circles* takes to these challenges are noted, and are then discussed in depth in Chapter 3.

### 2.1 Communication Channels and Media

In setting the goal of designing and implementing a distributed, multiuser sociable environment, considerations must be made for a variety of issues. The nexus of these issues is the determination and design of communicative modalities, such as text, speech or gesture. Visual and auditory channels for social communication have been embodied in many different media, ranging from text-only messaging, to environments focusing on graphical elements, to full-motion videoconferencing in numerous configurations, to systems focusing on speech for conversation. This section discusses systems and studies of the most relevant media for sociable communication, noting their strengths and limitations in the context of the decision to combine audio with abstract graphics as the foundation for *Talking in Circles*.

### 2.1.1 Text-based communication

A key decision in the design of an online sociable environment is the choice of communicative channels to provide. Many research and industry systems have chosen text as a primary communications medium. Text has obvious advantages, such as ease of implementation and low requirements in terms of users' computing resources. The popularity of text-only MUDs and of graphical text-chat systems attests to these advantages. What's more, much work has shown that textual communication in certain contexts can support rich and engaging interaction between people [Bruckman 1997, Erickson 1999-R, Turkle 1995].

There are several important problems with textual environments, however. An obvious but no less relevant one is the impact of typing skills on communication ability in such environments. Jensen et al, for example, found that participants in a cooperative game who were allowed to communicate by typing rated each other's intelligence as roughly the same as participants who had no direct communication at all or conversed via text-to-speech messaging; participants who could converse by using a speakerphone rated each other's intelligence significantly higher. The authors note there is evidence in the literature suggesting that typing speed can affect such judgments of conversational partners [Jensen 2000]. As far as the outcome of the game, text-only communication similarly was found to foster a significantly lower degree of cooperation than the speakerphone case and no significant difference in degree of cooperation from the no-communication and text-to-speech cases, supporting the use of audio over text for sustainable, mutually enriching conversation.

Other problems with text-only communication are mentioned by Viegas and Donath. They note that messages in text chats serve the dual purposes of conveying content and also conveying *presence* of the participant to others, an overloaded functionality which is not always conducive to free and focused communication [Cherny 1995]. Similarly, the linear accumulation of new messages and the use of discrete phrases as the communicative unit (as delimited by users through application of the Enter key or 'send' button) do not support fine-grained temporal management of conversation, such as "negotiation of conversational synchrony" [Viegas 1999]. Distribution of participants into subgroups or threads is problematic due to the aggregation of all users' messages into a singular common buffer, a problem shared by audio-only media as will be discussed later and an important reason *Talking in Circles* employs graphical representation of participants to enable naturalistic subgroup formation.

Another issue raised by Viegas is the low differentiation in appearance of users in a pure-text chat, generally limited to differences in name. This low representational figurativeness and embodiment also permits participants to easily modify the gender or other personal qualities they convey as well as to engage in impersonation or switch between multiple personas. While freeing in some ways, researchers have questioned whether the impact of these capabilities on those who take advantage of them as well as those who are exposed to others who employ them is on the whole favorable [Turkle 1995]. From the point of view of sustained sociable communication, it is not clear that such a sparse degree of participant embodiment is optimal, an important problem *Talking in Circles* addresses through its emphasis on voice communication for more stable and direct identity cues.

### **2.1.2 Graphical Text-chat Environments**

Another popular approach to online environments is that of textual communication accompanied by graphics. A variety of graphical chat systems exist that employ different approaches for conversations among participants. An important advantage of these systems, such as *The Palace*, is the use of a permanently visible proxy, or 'avatar,' used to represent a participant. One might hypothesize that breaking apart the fused purposes of sending a message – conveying the particular content, and conveying the sender's presence in the system – would reduce the tendency of participants toward brief or content-free utterances employed at least partly to remind others of the sender's presence in the chat room.

Recent empirical findings indicate this is not necessarily the case. Smith, Farnham and Drucker have conducted one of the most extensive studies to date of graphical chat environments. Their study focuses on *V-Chat*, an avatar-based chat system that provides a scrolling text window for messages and a variety of chat rooms with background graphics for the avatars to populate. Though the study's results are specific to *V-Chat*, that system is representative of current graphical chat environments.

On the question of whether avatars reduce the incidence of brief presence-oriented messaging, the authors report that a subset of the study's 119 days of conversation logs, comprising 31,529 messages, shows that fully 23% of messages included greetings of some form. Average message length was 5 words, and 61.3% of the sessions recorded over those 119 days resulted in participants sending no messages at all [Smith 2000]. These findings suggest that a graphical representation for participants, even one as potentially lush as provided by the avatars and backgrounds of *V-Chat*, does not necessarily foster sustained or complex social interaction. These

challenges are approached in *Talking in Circles* through the rich medium of speech communication in addition to representing participants graphically.

Another aspect observed in this study was usage of higher-level features of *V-Chat*, including motion of avatars around the chat room, use of avatars' 'gestures,' and display of a custom avatar. The authors found that motion around the chat room ranged from 5.9 positions per minute for participants' first use of the system to 2.0 positions per minute when participants had been around for more than 40 sessions. Such frequent and sustained use of the ability for motion supports strong consideration for including it in an online social space. The following chapter will discuss further what kind of motion capabilities may be most advantageous for such purposes of *social navigation*, for example the ability for members to observe others' activities as they evolve over time and decide their desired involvement in these activities.

Smith and his colleagues found that use of their particular implementation of gestures peaked at 0.57 per minute for first-time users and dropped approximately linearly to 0.13 per minute for those with experience in more than 40 sessions. This indicates that gesture was a moderately useful mechanism for participants. However, in light of the finding that 61.3% of sessions resulted in no messages sent and that average message length was 23 characters (less than one-third the width of a typical Unix console, for comparison), the possibility remains that gestures may be more useful in situations where communication is heavier or more complex. The literature suggests that this may be the case, as gestures can serve to clarify ambiguity, convey agreement or express approval [Rosenfeld 1967]. While this thesis does not address gesture in detail, Chapter 5 notes that users of *Talking in Circles* have sometimes used the multiplicity of modalities in similar ways, combining speech, motion of their circle and sketching for complex expression.

The third aspect of high-level features in Smith et al's study of *V-Chat* was use of custom avatars. Such use increased from 21% of sessions during first login to 76% of sessions for participants beyond their fortieth login. While these numbers indicate enthusiasm for the possibility of customizing one's graphical representation, survey results show a variety of motivations for the use of custom avatars. 43% of the 150 survey respondents said they use them to express their individuality, 24% said they use them to stand out, 23% for dislike of the system's built-in avatars, and 11% for the challenge. Two thirds of subjects said their avatars represented their true gender, a rather low number, especially for self-report. Overall these results make it unclear whether unbounded graphical customizability of participants' look may be more of a positive or

negative feature in the design of an online sociable space. While *Talking in Circles* does not allow unbounded customizability of participants' representation, it affords capabilities such as sketching to allow a high degree of individuating self-expression.

A few other issues related to avatars bear mentioning. One is the tendency toward screen clutter in graphical chats. These systems often draw speech balloons near the corresponding avatar, provide a separate scrolling window that displays participants' messages, or do both. The *V-Chat* survey results say 76% of respondents looked equally at the main graphical window containing the avatars and at the scrolling message window above it. While it's encouraging that participants attended to both graphics and text, the necessarily disjoint display of participants' avatars and messages results in real-estate for messages being limited and in some cases messages covering up part of the graphics. Of greater concern is the breakup of participants' holistic embodiment, as looking at messages requires looking at a separate space on the screen from that occupied by the avatar of the participant who sent the message. One goal of *Talking in Circles* has therefore been a more robust and holistic notion of embodiment.

A related aspect of graphical clutter is the distracting need for view management it fosters. The authors of *Comic Chat* note that chatters in both 2D and 3D rooms need to take care not to obstruct or be obstructed by others' avatars. Speech balloons can also cover up avatars or other speech balloons, one problem addressed by that system, albeit only for dyadic or triptic conversations [Kurlander 1996]. In addition, observation of *Active Worlds*<sup>4</sup> over a three-month period suggested that *proxemics*, briefly, the management and conceptualization of space by people [Hall 1966], is applicable to graphical chat environments. Specifically, Jeffrey found that *collisions*, that is, overlap of one avatar on another, were a very contentious issue for participants. Positioning one's avatar in very close proximity to or in direct contact with another was often considered forward and made the other person uncomfortable [Jeffrey 1998].

Flimsy though digital graphics may be from a physical point of view, they are real enough to elicit protectiveness of personal space, whether out of a general perception of being crowded or out of concern for what onlookers may think of seeing avatars very near each other. Indeed, the recipients of this virtual crowding responded with such comments as "This is a nice distance to keep :)" (after moving away from the encroaching avatar) and "[people] will see you..cut that

---

<sup>4</sup> <http://www.activeworlds.com>

out” and often referred to the notion of being “in someone’s face.” This kind of crowding also occurs by accident, which does not necessarily make the person caught underneath any less bothered by it. These results highlight another important aspect of embodiment in online environments, namely physical integrity and the ability to maintain some notion of personal space. As the following chapter details, *Talking in Circles* prevents participant overlap to preserve their physical integrity while highlighting the dense feel of a crowd.

A third issue of potential concern is users’ tendency toward caricatured avatars, as has been touched on. Viegas comments that “the avatars can distort expression and intent by providing a small range of (often broadly drawn) expressions that overlay all of a user’s communications. Even if an avatar has several expressions, and many do, it is still a far cry from the subtlety of verbal expression” [Viegas 1999] (also see [Eisner 1990, Thomas 1995] for examples of the biasing power of caricature). While avatars are not necessarily realistic, they are figurative enough to raise these difficult issues. *Talking in Circles*, like Viegas’s *Chat Circles*, employs abstract representation, among other reasons, to minimize unintended communicative biases.

### **2.1.3 Video-based Media Spaces**

While text and graphics provide useful possibilities for communication, audiovisual channels allow for much more immediate conversation in ways people are used to. Faces and voices, for example, are direct components in the conveyance of identity. Though both the audio and video components allow for rich communication, this section and the following one discuss why audio currently provides the most fruitful portion of these immediacy and identity benefits.

Because of its potential for a clear depiction of the face and physical gestures, video has long been expected to be of great value as a collaborative medium [Gaver 1992]. So far this hope has met with limited success. While subjects usually report greater overall satisfaction with video-mediated communication than with audio-only scenarios, sustained or significant qualitative or quantitative differences in the actual conversation have been difficult to find. Particularly where they are meant to be used by more than a handful of people, as is the focus of this work, video spaces present a number of complex problems.

One video space that found moderate success was Bellcore’s *CRUISER*. Employed by summer interns and their mentors, this system supported some interesting behaviors such as maintaining a video connection open for an extended period in order to have quick and informal access to a colleague. *CRUISER* was primarily designed for one-on-one conversation, however, and in the

end was judged to be useful generally only for tasks for which the telephone is normally used. The great degree of figurativeness also resulted in users feeling video calls were as intrusive as face-to-face interruptions and in some cases more invasive of privacy due to insecurity about potential eavesdropping by third parties [Fish 1992]. Although these concerns are partly the result of the workplace environment the system was used in, taken together with the findings on proxemics in graphical text chats it's not clear that people would be comfortable being viewed up close by groups of strangers for extended periods.

While video images could be degraded as a palliative to their intrusive feel, studies have generally found that even pristine full-motion video may not result in measurably superior communication. Subtle gestures and even explicit attempts to achieve eye contact can be easily lost on video [Heath 1991]. Indeed, studies have had trouble formalizing the differences in communicative ability afforded by video over audio-only communication. Given the steep hardware requirements of video spaces and their uncertain value, speech remains a reliable foundation for distributed human communication, as this thesis supports.

Some of the most extensive comparisons in this area have been done by Rutter, who has studied communication between pairs and among groups of four subjects in conditions ranging from face-to-face, to telephone conversation, to co-presence with a screen between participants. While face-to-face cases were generally found to allow for greater informality and discussions on a personal level, telephone conversations (often labeled *cueless* in the literature) evidenced adjustment over time that resulted in increased spontaneity. In fact, Rutter found that in the case of a college-campus nightline service telephone conversation functioned as well as or better than face-to-face communication for discussions on a personal level [Rutter 1987].

Over time Rutter's stance has shifted toward a broader model where subjects gauge each other's *psychological distance* not so much through the number of cues available but more importantly through what *usable* cues exist. He hypothesizes that "if cues do vary in salience, perhaps people attend only to the most important and simply ignore the remainder, making much of the available information redundant" [ibid, p. 137] (see also [Cook 1972]). Hopper notes that complex features as high-level as syntactic and lexical language characteristics and as low-level as pitch contour of terminal phrases, for example, are involved in speakers' aptitude for guessing which conversational pauses are open for taking the floor [Hopper 1992]. The robustness of speech communication, detailed in the following section, further supported the decision to employ audio as the primary communicative channel in designing a sociable online environment.

More recently, Sellen has performed comparisons of three kinds of multiparty video-mediated communication with face-to-face and audio-only conditions. Her experiments showed no significant differences whatsoever between the audio-only and video conditions. In fact, even the elaborate video setups used brought to light difficult problems in videoconferencing such as conveying who is attending to whom in a group or how individuals themselves are being seen by each of the others on video [Sellen 1995]. In general, while multiple video viewpoints are important to unlocking the benefits of videoconferencing, cognitive grasp and mapping thereof to physical view-management control is quite challenging [Gaver 1993]. As will be detailed in the following sections, *Talking in Circles* addresses these problems through a cohesive shared view of the space and its participants and a strong spatial grounding to exploit intuitive behaviors such as approaching an audio source to be more closely engaged with it.

#### **2.1.4 Audio as Primary Communication**

As the preceding sections have detailed, audio emerged as the most promising starting point for the design of a sociable online space. The high hardware and networking requirements for videoconferencing, its unclear benefits and its highly personal representation make it problematic for such a space. By the same token, audio retains great flexibility for expression and conveyance of identity and tone, while being less invasively figurative than video. Audio is the primary channel used in *Talking in Circles*.

As Schmandt summarizes, “the expressiveness of speech and robustness of conversation strongly support the use of speech in computer systems ... as a medium of interaction” [Schmandt 1994]. Speech is easy and natural, being the normal way most people have historically communicated and continue to communicate, and bears a smaller cognitive load than text generation [Kroll 1978] (see also [Pinker 1994]). The human voice is rich in dynamic identity and intonational cues, revealing many more than text communication generally supports.

Speech is, as Schmandt notes, resilient to many kinds of filtering, distortion and noise. Cherry, for example, recorded two passages read by a single person and overlaid them on a tape, then asked a listening subject to write down the two stories, playing the tape back and forth as necessary. Though difficult, subjects can complete this task for many different kinds of texts, even though the passages are read by the same voice and no directionality in the audio is available to help the subject separate the streams [Cherry 1957] These findings relate to Rutter’s and Sellen’s research on the robustness of so-called “cueless” communication. In Cherry’s words, “constraints which

exist in language are said to introduce *redundancy* – a rather unfortunate term in view of the important role it plays [ibid, p. 115].

A further advantage of speech is the ability to leverage users' experience with the telephone, an audio-only channel that people are already well-adapted to [Rutter 1987], having spent 3.75 trillion minutes on it in 1987 in the U.S. alone [Hopper 1992]. In practical terms, an audio channel serves many of the conversational functions outlined by Goffman, including two-way acoustic capability; back-channels for on-the-fly feedback on reception; means of initiating, confirming and breaking a conversational connection; turn-taking signalling; pre-emption signalling for interruptions; framing capabilities for distinguishing special readings such as jokes and asides; and support for pragmatics and other communicative social norms [Goffman 1981]. Beyond these basic capabilities of the audio channel both speech and musical audio support ready discernment of emotional content by people [Beldoch 1964].

These qualities of audio communication have been tested in telephone systems for decades, of course, but have also been demonstrated in *Thunderwire*, a CSCW audio-only media space. One user described this system as “a lightweight sort of social space,” while the authors note it fostered sociable conversations including interchange, play and personal warmth [Hindus 1996]. In fact, *Thunderwire* was only mildly successful in its task-oriented goals but greatly successful in its social goals. Nevertheless, the system suffered from important shortcomings which kept participants from feeling fully comfortable or satisfied with it.

As Rutter notes, one common problem with both speakerphones and conference calls is recognition of who the current speaker is, facilitated only if they identify themselves each time [Rutter 1987]. *Thunderwire*, having no graphical display, was also prone to this problem, although over time users got to know each others' voices. In fact, even graphical conferencing systems such as *MASSIVE* [Greenhalgh 1995] and *FreeWalk* [Nakanishi 1996] provide scant information to aid in speaker identification. More importantly, however, *Thunderwire* users were never sure of the currently active membership of the space, and sometimes found it distressing to be unaware of who may be listening to them. In addition, as with standard conference calls, the system was not well-suited to multiple conversational threads, since all participants shared a singular audio channel. In short, audioconferencing is poor at conveying participants' continuous presence and supporting their identification, particularly when used by more than two speakers, when speakers are distributed, or when they do not know each other a priori.

*Talking in Circles* addresses these issues through the complementary use of graphics, displaying which participants are logged in at all times, and additionally employs distance-based fading of the audio as well as a maximum sound dissipation threshold to enable subgroup conversations. Participants control their location in the space in relation to others, and thus their exposure and membership in these subgroups, through the position of their representative circle. A related point to note is that *Thunderwire*, consisting of a single modality, does not integrate other communication channels which might be advantageous. The use of graphics in *Talking in Circles* for such purposes as displaying conversational membership also opens up the possibility of complementary channels such as pictographic or iconographic ones, as will be detailed in the following chapter. Finally, graphical feedback, visible as the bright circles in Figure 1.1, accompanies participants' speech colocated on their circle, resolving the problem of speaker identification.

## **2.2 Socio-spatial Grounding**

The ability of participants in *Talking in Circles* to move their circle around the space is generally not constrained. In order to create an environment that can be viewed and understood by the other participants, and thus enables them to engage in conversation and manage their social interactions, it was necessary to strive for *legibility* of the space and its members. Kevin Lynch defines legibility in the classic *The Image of the City* as “the ease with which [a cityscape's] parts can be recognized and can be organized into a coherent pattern” [Lynch 1960, p. 2-3]. The cityscape is but the stage, however— “Moving elements in a city, and in particular the people and their activities, are as important as the stationary physical parts” [ibid, p. 2]. Lynch continues, “A vivid and integrated physical setting, capable of producing a sharp image, plays a social role as well. It can furnish the raw material for the symbols and collective memories of group communication” [ibid, p. 4]. Achieving both social and spatial legibility, indeed, was one of the primary goals for the design of *Talking in Circles*.

### **2.2.1 Avatar Proxemics in Graphical Text Chat**

As has been noted, participants in *Talking in Circles* can move about freely. What expectation can be had about the structure and legibility of their motion? One data point is provided by Smith's aforementioned study of the avatar-based chat environment *V-Chat*. Smith and his colleagues studied proxemics in their system by breaking up the chat room's space into a 40x40 grid, then measuring the average distance a subject's avatar stood from a randomly selected other participant as well as to the target, the intended recipient of the subject's message. They found a

statistically significant difference, with subjects standing closer to their intended recipient than to random others.

While this is an encouraging result for graphical social environments, it must be noted that the difference in distance was quite small. Subjects were, on average, approximately 10.8 grid squares away from their target, and 12.4 squares away from randomly selected others. This is a difference in distance of only 8% if we consider avatars as tending toward the center of the room, or 4% if we assume they might be positioned off to the side as well. It's not clear that these small differences are robust enough to be reliably read by other participants observing the space; perhaps a follow-up study will reveal whether *V-Chat* participants can in fact predict others' intended recipients based largely on the formers' location within the room.

In *Talking in Circles*, as in its predecessor *Chat Circles*, distance from others is directly tied to perceived magnitude of their input, motivating participants to modulate their visual representation's distance from others in ways that naturally reflect their attention and interests, facilitating social navigation [Viegas 1999]. The low significance of participant location in graphical chat environments such as *V-Chat*, by contrast, is a result of the fact that motion in these systems creates no functional differentiation. All areas of their chat rooms are generally the same, regardless of the specific part of the background users move over or their location with respect to others in the room.

*Babble* is another text-chat system that employs graphics to spur what the authors term *social translucence*, or the provision of "perceptually-based social cues which afford awareness and accountability" [Erickson 1999-S]. In *Babble* participants are represented as colored dots within a small window. The window contains a circle, and members of the same conversational topic (chat room) as the user are displayed inside this circle. Participants who speak or "listen" by interacting with their *Babble* window move toward the center of the circle, showing high engagement in this conversation. Those who are idle for a prolonged period drift toward the periphery of the circle. Finally, those in other chat rooms appear outside the circle. This use of graphics for display of conversational membership is compelling, although the inability to know which conversations those not in this chat room are engaged in, are moving toward or leaving from, is a major obstacle to the social legibility of the space. *Chat Circles* and *Talking in Circles*, on the other hand, employ a cohesive view of all participants, allowing everyone constant observation of others' changing conversational membership.

## 2.2.2 Milgram's Studies of Crowd Formations

In order to foster legibility of the *Talking in Circles* space, studies of real-life individual and crowd mobility patterns are essential. Such social domains as cocktail parties, where those present tend to break up into small conversation groups and wander around the room mingling with other groups, indicate that distance-based audio attenuation, akin to the physical laws in real conversation spaces, might foster similar recognizable behavior among participants.

Sociological studies support the legibility that stimulus attenuation can impart. Stanley Milgram performed studies of crowds to determine various factors about their formation and composition. The primary formation noted by Milgram is the ring, the tendency toward which is due to its being “the most efficient arrangement of individuals around a point of common interest” [Milgram 1977, p. 207]. He further notes that such circular gatherings are robust, and can accommodate subsequent aggregation of members, even in concentric layers.

Often the point of common interest may be an individual, while in *Talking in Circles* the point of common interest is generally the intangible entity that is the conversation itself. People's natural tendency toward a ring formation is enhanced, as in cocktail parties, by the proxemics of participants' graphical embodiment (shown to have at least some applicability in [Smith 2000] and [Jeffrey 1998], as discussed) as well as by audio attenuation, which as in real spaces creates a tangible perceptual correlation between distance and clarity of the input source. To support naturalistic mobility patterns and thus legibility of the online space, the design of *Talking in Circles* thus calls for distance-based attenuation of participants' speech.

Milgram notes, interestingly, that the shape of crowd formations long eluded study. He places part of this blame on the custom observers generally have had of viewing crowds from the plane of the crowd itself, that is, at ground level. The most useful vantage point for observation of people and their motion in a crowd, however, is from a position directly overhead [Milgram 1977]. Indeed, while *V-Chat* is a 3D graphical system, the view onto its chat rooms is primarily 2D from slightly above ground level, which makes the z-axis difficult to use and prone to causing occlusions (as discussed, these are quite undesirable in graphical social spaces). *Talking in Circles* employs an overhead 2D view that encompasses the entire area the participants occupy, the viewpoint suggested by Milgram's studies. As will be discussed in Chapter 3, this viewpoint permits surveying of the space as well as detailed observation of movement by individual parties as they wander from conversation group to conversation group.

### 2.2.3 Whyte's Sociological Studies of Urban Architecture

In addition to factors such as participants' motion and distance-based audio fading, the design of the "cityscape" of the online space itself is key to legibility. Before getting carried away with complex architectural designs, it is useful to heed William H. Whyte's summary of a decade of study on what makes a successful social space— "What attracts people most, in sum, is other people. If I labor the point, it is because many urban spaces are being designed as though the opposite were true" [Whyte 1988, p. 10]. What is it *about* people, and by extension spaces, that attracts people to them?

Whyte provides several answers to this question. Unfortunately, many are not easy or impossible to apply well to online spaces. Sun easements, trees, and availability of food and ample seating room are among these. There are, to be sure, other aspects Whyte cites which can be of great use in creating a sociable networked environment. Variety and liveliness, for example, are provided to an extent in *Talking in Circles* by the ability to observe fellow citizens in the space, note the formation of groups, the joining and departing by subsequent participants, the collective activity or individuals' speech that indicates conversational interest, boredom, excitement.

Beyond the features of the crowd of participants itself, there are structural factors that can contribute to a legible and amenable space. Whyte notes a predilection by people for areas with nearby objects, whether flagpoles or statues. "They like well-defined places, such as steps, or the border of a pool. What they rarely choose is the middle of a large space" [Whyte 1980, p. 22]. Thus, objects within the space, perhaps including "something roughly in the middle" as Christopher Alexander puts it in *A Pattern Language* [Alexander 1977, p. 606], may be beneficial in allowing people to choose how to arrange themselves over the open areas. A related observation by Whyte is the importance of some choice over one's situation in a space. "The possibility of choice is as important as the exercise of it," he notes wisely [Whyte 1980, p. 34]. Differentiation among central and edge areas, internal boundaries and other such reference points can provide some choices within the space.

A third suggestion is for some form of *sound cover*. In urban plazas this can often take the form of waterfalls, which though extremely loud in and of themselves, provide a feeling of seclusion for conversants because it increases the difficulty of making out their speech above the din. Ideally, of course, the sound may be relaxing, pleasant or interesting, besides masking conversation. An indoor example of this kind of pleasant and utilitarian sound is the relatively loud playing of a band at a jazz club.

In addition to functioning as a pleasant form of sound cover, there is another beneficial aspect to an enjoyable audio source. This is what Whyte terms *triangulation* or “the process by which some external stimulus provides a linkage between people and prompts strangers to talk to each other as though they were not” [ibid, p. 94]. Triangulation in an urban space can be provided by a sculpture, a street performer, or even a striking building. In *Talking in Circles*, it is provided by *sound booths* within the space. As Donath notes, however, simple aggregation of inputs does not guarantee an interesting online system. The idea is “to create environments that combine a rich information landscape with the ability to communicate with others – information spaces that provide a context for community” [Donath 1995-S].

The sound booths, the two central circular areas that can be seen in Figure 1.1, are designed to perform many of the functions Whyte suggests within the context of an engaging social environment. First, the booths break up the space, creating central areas near them, boundary areas at their edges, as well as areas distant and secluded from them. In addition to breaking up the space in the visual channel, the sound booths break up the space with audio, both music and news. In short, the booths are designed to provide variety in the space, differentiation of areas, choice, pleasant sound sources to mask conversation, and likely centers of triangulation. These capabilities are detailed in the following chapter.

As mentioned earlier, many of the lessons from Whyte’s work and other sociological studies are not directly applicable to the online world. Displaying a chat room with a picture of a sun, of a crowd, or of chairs, as some graphical chat environments do, unfortunately does not provide the benefits these factors have in real life. The sound booths exemplify the approach taken in the research that led to *Talking in Circles*, that is, carefully discriminating whether notions carry over from the physical to the virtual world and designing ways to fruitfully apply these notions to online interaction.

While merely placing an image of a sculpture into an online environment would probably carry little of the triangulation benefits Whyte noticed in three-dimensional pieces larger than human scale which one can walk around or under, a dynamic spectacle such as music is easy to carry over. The music and sound booths provide not only a focus for conversation but an important source of common ground that participants share [Cherny 1995]. Lastly, the booths exist as entities independent of conversation within the space. While most online environments rely solely on other users to make a visit worthwhile, the streaming news booth and interactive music booth

are continuously available for enjoyment or enrichment even when no conversational partners are online.

### 3 The Design Process of *Talking in Circles*

*The simple social intercourse created when people rub shoulders in public is one of the most essential kinds of social “glue” in society* – Christopher Alexander

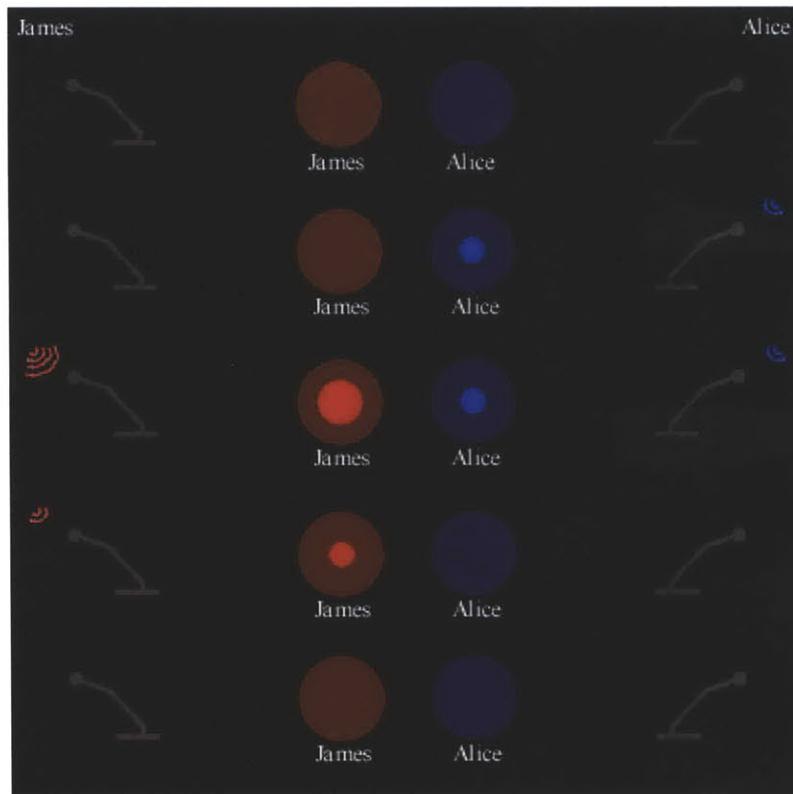
The preceding chapter detailed studies and systems which provide the scaffolding for the design of *Talking in Circles*. This chapter discusses the design process in full, including specific design decisions and omissions as well as an analysis of each major system feature.

#### 3.1 Speaker Identification

Having settled upon audio as the primary communicative channel, a first basic faculty required for a successful multi-party audioconferencing environment was the ability to identify the participants, particularly the current speaker. As discussed in the previous chapter, such identification is difficult in traditional audioconferencing environments. This problem, among many others, benefits from the graphical component of the *Talking in Circles* interface. The aim was helping users map the voices they hear to the circles representing the respective participants, that is, to help transform the user experience from disembodied voices to a more cohesive perception of fellow participants. Collocated graphics, displayed on participants’ circles in accompaniment of their speech, are used for this purpose.

The system uses a bright inner circle displayed inside participants’ darker circle to represent the instantaneous energy of their speech. Figure 3.1 shows James and Alice speaking, with their inner circles’ size showing how loud their speech was at several instants. The pervasive natural pauses in speech and conversation thus leave only particular circles showing speech activity, making this cognitive matching problem much simpler. Distance-based spatialization, discussed below, also

helps, as speech from circles farther from the user's circle sounds fainter. Finally, identity cues such as learning a participant's voice or their circle's labeled name also resolve matchings. Since graphical features such as each participant's circle, circle color and name are constantly available, others can identify not only the current speaker but the set of current participants, and over time employ these various cues –voice, name, location in the space– to gain a fuller and more permanent sense of the participants than current-speaker-identification provides. This is particularly important in the case of a sociable, distributed space such as *Talking in Circles* because participants will often initially be complete strangers to each other. Even mere awareness of presence and minor identifying details suffice to forge a powerful sense of shared experience and can eventually conspire to lead *familiar strangers* to interact [Milgram 1977].



**Figure 3.1: Diagram of several moments' sample speech between James and Alice. As each speaks into their microphone (left and right sides), their onscreen circles (center) display graphical feedback showing their instantaneous audio energy over several moments.**

I conducted an informal test of the graphical feedback provided by the dynamically-changing bright inner circle. Six subjects were shown two circles with non-identifying names, equidistant from their own circle and at equal audio volume. The two test circles each played a different

RealAudio news stream, and I asked participants about their experience in trying to match the two speakers they heard to the correct circles.

Although this scenario is challenging, with constantly overlapping speech and no individuating cues for the circles, all subjects successfully matched each stream to its corresponding circle within a few seconds. Though simultaneous speech was at first confusing, the subjects mentioned that the occurrence of short natural pauses in the speech soon made the matching apparent, as only one voice was heard and only one circle was bright. A similar situation happened when one speaker said something loud, causing one bright inner circle to grow visibly larger than the other. In general, the subjects said the speech-synchronized updating of the bright inner circle's helped them differentiate and identify the speakers by highlighting the matching rhythms of the speech.

In addition to displaying instantaneous speech energy, an early design attempted to show recency of participant activity, allowing the bright inner circle to fade out slowly over time when a speaker stopped speaking. This experiment provided a slightly enhanced short-term history but it interfered with the real-time feel of the inner circle's accompaniment of speech. While recent activity can also be useful in indicating speaker availability, the two are not necessarily equivalent. As Ackerman found, local disruptions can cause frequent changes in listening or speaking availability of users without their remembering to turn off their microphone, even when it resulted in unwanted eavesdropping [Ackerman 1997]. Similarly, lack of speech by a user for several minutes does not guarantee that they are not still listening. Thus, manual availability indication is unreliable, and accurate automated detection of participant availability is currently highly problematic as well. The system therefore does not currently attempt to display availability status.

## **3.2 Spatialization**

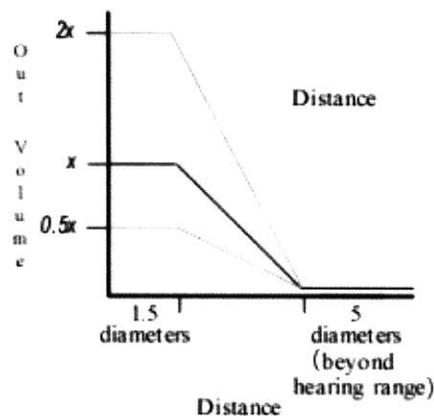
While the system supports various user capabilities, some are supported directly while others arise out of a combination of modalities. The most salient behavior is that of circle motion as an indication of interest and membership in a conversation. As Milgram noted, rings are a naturally-emerging configuration of people engaged around a common activity [Milgram 1977]. In *Talking in Circles*, as in face-to-face situations, standing close to someone lets one hear them clearly and reduces distraction from other sources farther away, due to distance-based audio attenuation.

This natural tendency toward physical alignment, besides being a functional conversational feature and serving to a limited extent the role of gaze, has additional benefits. It allows other

participants to view the formation of groups or crowds around a particular discussion, letting them gauge trends in participants' interest and advising them of conversations that are potentially interesting. Crowd motion does not necessarily require explicit attention; as in real life large gatherings stand out, and can continue to draw people as users notice the traffic and wonder what the fuss is about.

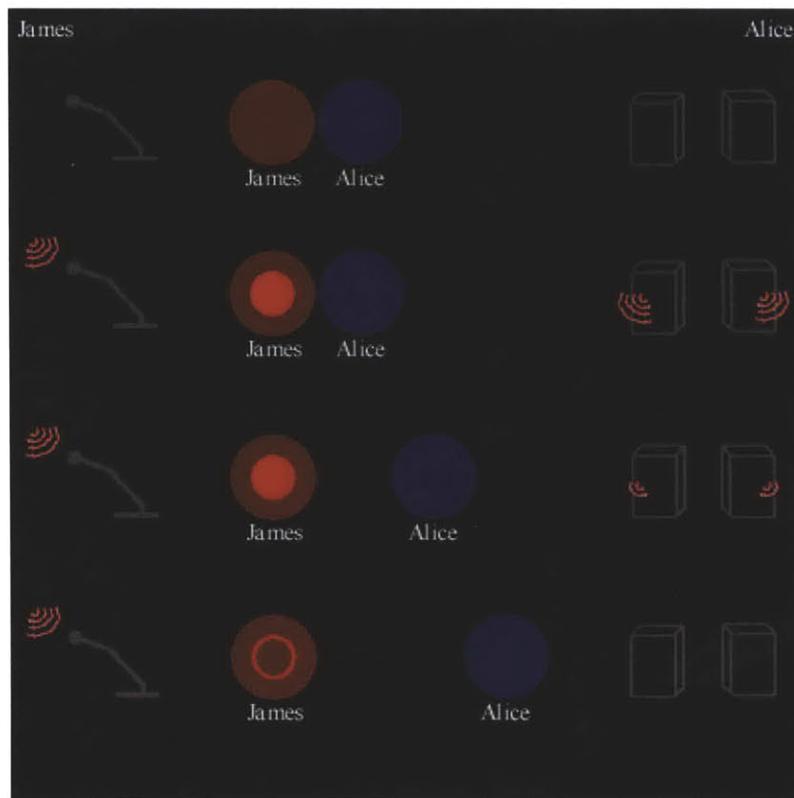
An additional important benefit of this crowd motion is simply the vitality with which it imbues the space. Whyte remarks on the fact that the biggest draw for people is other people, and notes the popularity of people-watching as a form of triangulation— simply stated, a stimulus source which can be observed by multiple members of the population, potentially giving rise to conversation between strangers. The grounding in a 2D space may also bring in features such as traveling conversations, where conversants move across a space to find a comfortable spot along the perimeter [Whyte 1988]. As in real life, it is possible, with some effort, to be near a particular speaker but attend to another, or to stand between two groups and attend to both conversations.

Selective attention, enabled by the physical grounding and audio attenuation, also provides some of the benefits related to the 'cocktail-party effect' [Cherry 1953]. Though audio from those one is closest to is heard most clearly, nearby conversation can be heard more softly within a certain distance threshold. This helps a user concentrate primarily on the conversation group they have joined while preserving peripheral awareness with the possibility of 'overhearing,' such that the mention of a name or keyword of interest can still be noticed. Thus, social mingling is fostered, as participants can move between subgroup conversations as their interest changes or move to an unoccupied physical space and start their own conversation.



**Figure 3.2: Output volume as a function of distance from a speaker, for input volume  $x$  (darkest line),  $0.5x$  and  $2x$ .**

In order to allow clear audio for participants in a conversation, no audio fading is done within a distance of 1.5 diameters from the center of each speaker's circle. This allows participants located next to or very close to each other to hear the full volume of speakers' speech, while fading is performed for circles in conversations farther away. Figure 3.2 plots the shape of the audio-fading function to show how output volume varies by distance from a speaker. The function remains the same but is parameterized by the instantaneous input volume, as shown by the upper and lower lines in the figure. This modification to the spatial rules of our environment preserves the positive qualities of audio fading but helps members of a conversation hear each other clearly; the system's focus on spatial grounding is always rooted in fostering a sociable space. Though detailed user control over fading parameters could be beneficial, such as in the case of a very widespread conversation group, customizing the physical rules of the space can lead to inconsistent user experiences [Smith 1996] as well as unnecessary GUI clutter [Singer 1999], contrary to the system's design emphasis on minimalist, highly-visible, abstract representation.

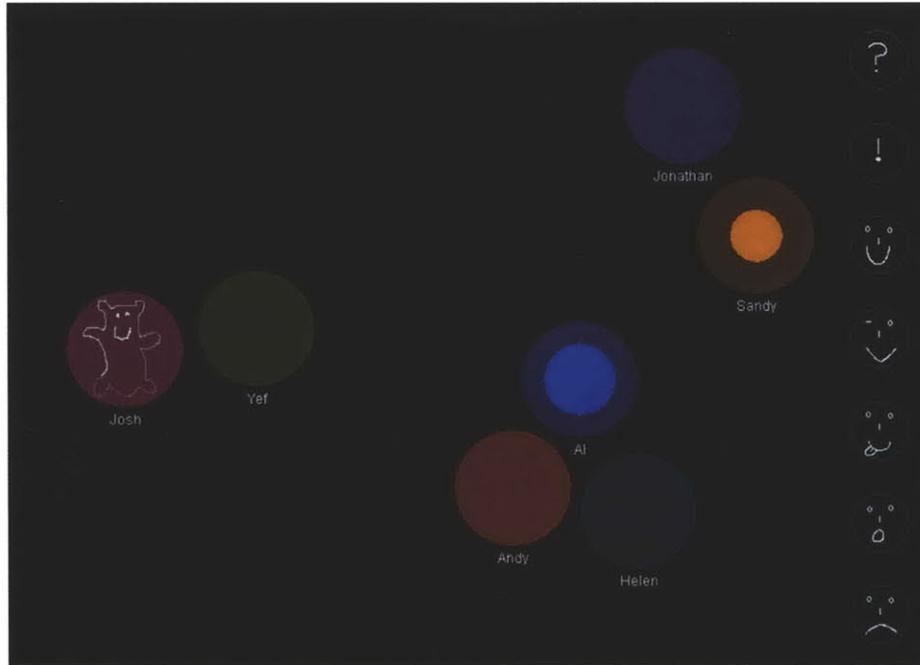


**Figure 3.3: Diagram of sample speech between James and Alice. As James speaks into his microphone (left side), Alice moves her onscreen circle farther from his (center). As she moves away, her perceived volume for James's speech (through her speakers, right right) decreases.**

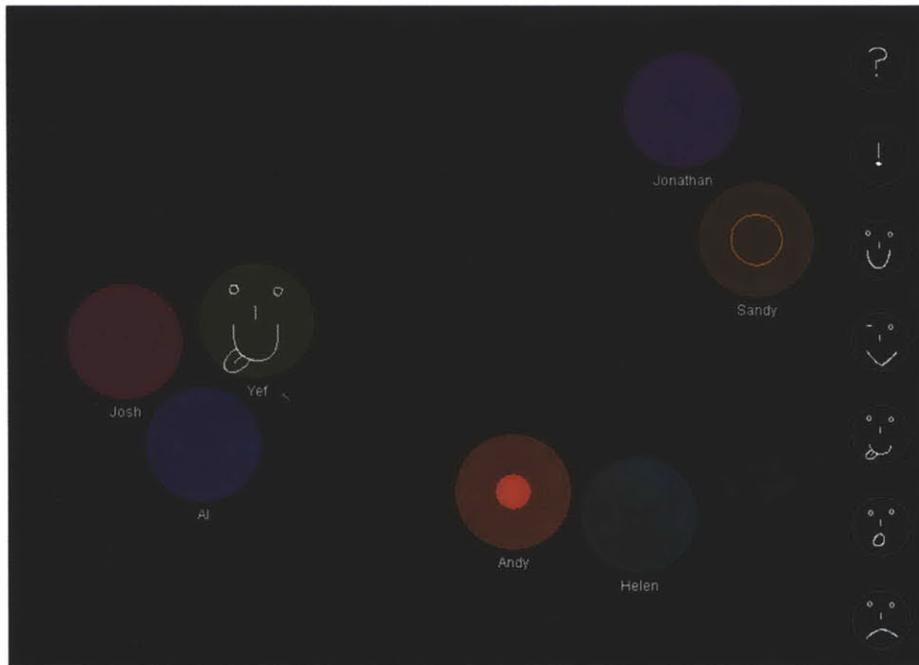
The distance threshold for audio to be heard, currently 5 diameters, serves multiple functions. Naturally, it aids performance optimization by obviating the need for audio playback for clients beyond the threshold. The major benefit, however, is letting the user know that they cannot hear someone, as activity by those beyond the hearing range is rendered as a hollow circle. For example, screen shots of a *Talking in Circles* chat from the screen of participant Al, the blue circle, shows he has moved from a conversation with Andy and Helen in Figure 3.4 to one with Josh and Yef in Figure 3.5. The hollow orange inner circle shows that Al is now beyond the hearing range of Sandy. Since the hollow circle still indicates speech, however, a participant can note a spurt of activity even if they cannot hear it, and can move closer to see what the discussion is about if they so desire. Figure 3.3 summarizes mechanics of the audio fading and graphics.

The system's space is currently 10 circle diameters diagonally, enough room for several groups to distribute themselves around. The 5-diameter maximum hearing threshold and linear fading of the audio are generous, designed to allow a good amount of mild perceptual awareness of the speech in much of the space, as one has at a cocktail-party. Traditional inverse-square-of-distance audio attenuation would result in exact positioning of one's circle being important and would likely result in the suboptimal requirement for increased view management, a requirement Kurlander warns against [Kurlander 1996]. No extensive studies of users' mental maps of an online graphical audioconferencing environment exist. Such studies, for example performed by varying the audio-fading function, could generate interesting comparisons with physical spaces. What are users' expectations for audio attenuation in a system such as *Talking in Circles*? Do they expect to be able to hear everyone throughout the room, or for speech to carry only a small distance outside the speaker's circle? Detailed study of such questions could prove beneficial to the design of online graphical audioconferencing.

The audio-fading threshold is symmetric, so that if user X is too far to hear user Y, Y is also too far to hear X. This feature lets a user easily find a spot where they cannot be heard by a certain group, by noting when their inner circles appear hollow. Thus, as in a real cocktail party, one can move to the side to have a semi-private conversation, although this privacy relies only on social rules and is not enforced or guaranteed by the system. These interaction possibilities address some shortcomings of video-mediated conversation, such as the lack of a "negotiated mutual distance" and of a sense of others' perception of one's voice [Sellen 1992]. The system does not provide mechanisms for breaking this perceptual audio symmetry, such as the ability to mute a particular participant or raise the volume of another at will.



**Figure 3.4: Three conversations in a *Talking in Circles* chat session. AI and Sandy are speaking, while Josh draws for Yef.**



**Figure 3.5: Screenshot from the point of view of AI, the blue circle, as Sandy and Andy are speaking. Andy is within AI's hearing range, while Sandy is now outside it.**

It's important to emphasize the role such features as symmetric audio fading play in social environments. Erickson et al define *socially translucent* systems as those providing "perceptually-based social cues which afford awareness and accountability," and suggest translucence is key for users "to carry on coherent discussions; to observe and imitate others' actions; to engage in peer pressure; to create, notice and conform to social conventions" [Erickson 1999-S]. Ethnographer Joshua Meyrowitz echoes the applicability of these notions to online social interaction.

When we find ourselves in a given setting we often unconsciously ask, "Who can see me, who can hear me?" "Who can I see, who can I hear?" The answers to these questions help us decide how to behave. And although these questions were once fully answered by an assessment of the physical environment, they now require an evaluation of the media environment as well [Meyrowitz 1985, p. 39].

One group of visitors to the Media Lab who tried the system suggested that, in their corporate setting, they would find it useful to have private breakout rooms for a couple of participants each, as well as a larger full-group meeting room. Although *Talking in Circles* can easily be adapted to support such a mode, our focus on a purely social space makes relying on existing social behaviors more interesting to us than technologically-enforced boundaries. A related concern is that of rudeness or other undesirable behavior by participants. Once again, the system's varied interaction design can support existing or emergent social mores that help sort things out; just as people can move closer to conversations or people they are interested in, they can move away from conversations which become uninteresting or people who show hostility.

Beyond these pragmatic features of distance-based audio attenuation, other potential sociable applications exist. For example, with greater audio attenuation the popular children's game of "telephone" could be played, in which a large circle is formed by all attendees and a short phrase or story is whispered from person to person around the circle, becoming increasingly distorted, until it gets back to the originator and the starting and ending phrases are revealed to everyone. The current attenuation function and thresholds already allow for various high-level speech behaviors. Though not always desirable, yelling to get someone's attention or to be heard more loudly even by those a certain distance away are possible, as is moving to a central location between disjoint parties to say something for all conversants in them to hear.

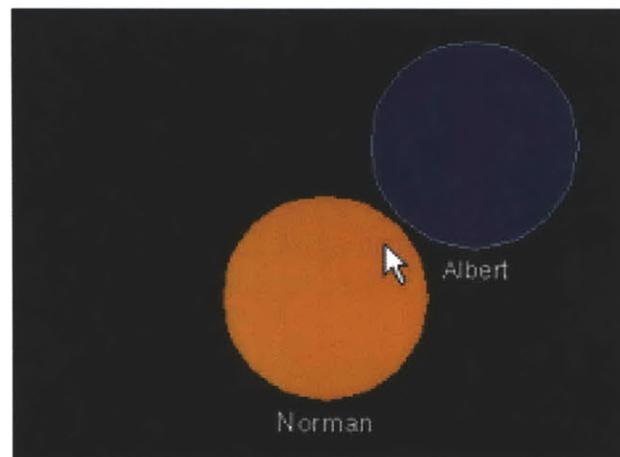
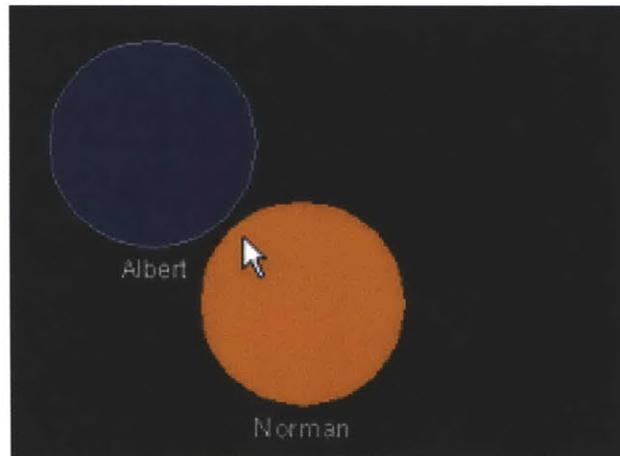
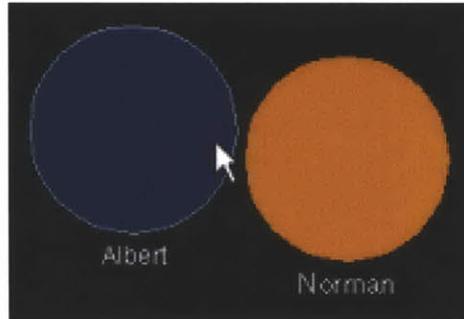
### 3.3 Circle Motion

As Christopher Alexander notes in the chapter-opening quotation, the “rubbing shoulders in public” that urban crowds exhibit is an important social phenomenon. Traditionally, graphical online environments do not support any such “tangible” interaction, as participants jump directly to any point in the space they wish to travel to. This teleportation ability has interesting parallels to modern urban architecture. Alexander goes on to discuss how cities now permit motion of large numbers of people to occur mostly through nondescript indoor passageways, thus “robbing” the street of people [Alexander 1977, p. 489]. An additional downside of such teleportation in online environments is its potential result in overlap of participants’ representations. As noted in Chapter 2, avatars partially obscuring each other, whether intentionally or accidentally, can arouse strong negative responses from those caught underneath. Visibility, as well as bodily integrity of one’s representation, are important social factors in graphical chats. *Talking in Circles* deals with these issues by not permitting overlap, simultaneously preserving visibility of participants’ presence and of their drawing space.

A circle’s motion is stopped by the system before it enters another circle’s space. In this situation, a participant can drag their pointer inside the circle blocking their path and their own circle follows around at the outer edge of the circle that is in the way, which provides the feel of highly responsive orbiting (see Figures 3.6a, 3.6b, 3.6c). Thus, at close quarters, participants still preserve their personal space and can move around in a manner which provides a certain physical interaction with other participants, an attempt at enhancing the feeling of being in a crowd as well as encouraging observable, lively motion in the space. Swift motion across large areas is still immediate, as obstructing the participant with all circles along the way to the new location would make for a cumbersome interface. As always, our aim is to leverage spatial grounding with a primary focus on social interaction design, hence our differing policy for motion at close quarters versus over greater distances.

### 3.4 Drawing and Gesture

One major benefit of audioconferencing, of course, is that it frees the hands from being tied to a keyboard. This freedom can be employed to run *Talking in Circles* on a keyboardless tablet, as mentioned earlier, or on a wearable computer. Unlike traditional chat systems we need not display large amounts of text, which takes up a lot of screen real estate, resulting in great freedom in maximizing the potential usage of the space and the graphical area marked off by each circle.



**Figures 3.6a, 3.6b, 3.6c: Participants' representations cannot overlap each other in *Talking in Circles*. Here, Albert approaches Norman, then moves swiftly around him, akin to crowd motion.**

Since the circles' interior space is used only momentarily during speech, this space can be used for drawing. Though the space on one's circle is limited, it is large enough for diagrams, bits of handwriting, and so on. Drawing strokes appear in bright white, visible even over the graphical feedback during speech, and fade away in 30 seconds. Although this makes long-lasting sketching more difficult, a tradeoff worth noting, our design intent is akin to letting people at a cocktail party use a napkin to share sketches on, and obviates potential distraction from cumbersome drawing controls. The relatively fast refresh rate keeps the drawing space available, which is important for drawing to be useful for gesturing.

Drawing faces is a natural tendency, and it is particularly inviting given the circular shape of the user's representation. The circle's space is enough to permit much more than simple emoticons, and even drawing-unskilled users immediately took to writing short phrases and drawing faces. Combined with moving one's circle, drawing a face can be targeted at a particular user both by drawing the face as facing in that user's direction and by moving toward that user with the face showing on one's circle. Coordinating motion with drawing has been popular with users, such as drawing a face with the tongue sticking out and moving quickly up and down next to the intended viewer, enhancing the facial expression with bodily motion.

Shared drawing is also useful for showing explanatory diagrams [Tang 1991], which Isaacs and Tang note as a user-requested capability in their study [Isaacs 1993], for certain kinds of pointing, and potentially for other meta-conversational behaviors such as back channels. These uses are important in creating a social space since studies of telephone conversations have found reduced spontaneity and increased social distance compared to face-to-face discussions [Rutter 1989]. Employing drawing, confusion can be indicated not just by explicit voicing but by a question mark or other self-styled expression on one's circle. Drawing has also been found by Rutter to help in clarifying complex discussions and to increase listeners' involvement with speakers when speakers draws [Rutter 1987].

As the system is used in various environments we are very interested in studying the development of novel drawing conventions and gestures for conveying various data. We have already observed novice users effect floor control, for example, by displaying an exclamation mark in their circle upon hearing something surprising, or simply by shaking their circle a bit to indicate they have something to say. Again, although voice by itself is useful in these tasks to some extent, both audio-only and videoconferencing studies have found complex tasks such as floor control to be less effective than can be done in face-to-face communication. Thus, the complementary

combination of voice, circle motion and drawing is aimed at overcoming some of these limitations.

In order to make the pictorial modality more accessible, we also provide a set of clickable icons that display drawings in the user's circle, similar to the availability of preset graphics in *The Palace*. As shown in Figure 4.5a, the system currently includes a question mark and exclamation point, as well as expressions indicating happiness, humor, surprise and sadness. These built-in icons are designed, as the rest of the system, with relative minimalism; the expressions are designed to look informal enough to be similar to users' own drawings and to not appear overly figurative and trite with repeated use. The drawings available on the icon bar are standard graphics files editable in any graphics editor, and the drawings the system includes can be removed or modified or new ones added simply by putting them in the *Talking in Circles* directory. The ready access to showing these iconic drawings on one's circle and the ability to customize this set of drawings makes the pictorial channel more available than requiring the user to draw everything from scratch each time.

The icons appear in one's circle for five seconds when clicked, as shown in Figure 4.5b. This brief display duration is designed to allow for relatively *reactive* use of the icons' expressiveness. For example, the exclamation point icon can be used during a conversant's speech to display interest or surprise, without having to interrupt the speaker or other participants in the conversation. In this way the icons are a rudimentary attempt at back-channel capability; people often display reactions on their face, giving the person they are meeting with instantaneous feedback without obstructing the audio channel. In addition, the icon-bar drawings can be clicked on in sequence and are updated immediately, which allows for higher-level expressive sequences such as pictorially sticking one's tongue out while making a humorous remark, followed by displaying the winking face and then the smiling face. Since the system is ideally used with a pen tablet, reader Beth Mynatt has made the excellent suggestion that a pie menu, centered on the participant's circle, would allow much more rapid, fluid and gestural interaction with the available display icons.

Lastly, drawing can of course be used strictly for doodling, whether out of boredom or to accompany music one is listening to, and for other purely aesthetic ends. Individuals' use of their drawing space—whether they draw constantly or rarely, make abstract doodles, draw faces or words— may provide others a sense of the person's identity. Besides being an expressive

channel, these behaviors serve as another form of triangulation, giving participants a spectacle to watch and gather around. Figures 3.7a and 3.7b show two drawings done in *Talking in Circles*.



Figures 3.7a, 3.7b: Two examples of drawings from *Talking in Circles*

### 3.5 Sound Booths

As discussed in Chapter 2, work by Lynch, Milgram and Whyte emphasizes the importance physical features have on social interaction in spaces. Their research concurs with the pragmatic guidelines Mynatt et al suggest based on their experiences with the *Jupiter* network community, such as designing online spaces to fit social activity and providing degrees of interaction and awareness through spatial techniques, including the use of audio and graphical elements [Mynatt 1997]. Along with the features discussed in the preceding sections, such as circle motion and audio fading, *Talking in Circles* employs *sound booths* to these ends.

The system contains two booths, a music booth and a news booth (near the bottom-left and top-right, respectively, in Figure 3.8). The music booth has nine musical notes on it, each representing a song, that can be played by a participant by clicking on the note; the song is then broadcast to all participants. This is another case where individual participants could be allowed to choose songs that play only for them. However, this capability is already filled by external, physical music devices participants can have available. The music booth, like the rest of the system, is designed to emphasize interpersonal experience, hence the music booth plays the same song, at the same time, for all listeners, enabling shared perception of rhythms and moods in the songs. The news booth plays a single audio stream which can be pre-recorded or live, generally streaming RealAudio news from CNN. While a cohesive interpersonal experience can be extremely powerful in the case of music listening, streaming news makes its importance even

clearer, as participants listening simultaneously to live news have those particular developments as topics for discussion with each other.



**Figure 3.8: Six participants conversing in *Talking in Circles*. Allen, Jackie and Sam are listening at the music booth, Kurt at the news booth.**

Indeed, the sound booths are not merely background graphics, as is common in graphical chat environments, but are interactive, dynamic features of the space, serving as good potential sources for *triangulation*. In order to facilitate listening to or discussing the music and news, the audio fading differs for the booths. The maximum distance at which their sound can still be heard is two circle diameters, rather than the five for participants' own conversation. The graphic representing the booths shows their fading function, being saturated at the center and becoming desaturated and fuzzy toward the outer edges. Figure 3.8 shows Jackie, Allen, Sam and Kurt within the sound booths, at varying distance from the center. The tighter audio fading for the booths creates permeable, compartmentalized spaces without hard graphical boundaries. Thus several participants can distribute themselves along the edge of the news booth and hear that stream at moderate volume while still being able to hear each other and converse about the news.

The booths also serve as background for participants who prefer some sound cover for their conversation, while others who prefer the quiet can use the spaces outside the booths' area.

The booths are generally graphically static, but they display the name of the song or news stream that is playing when a participant moves their pointer near them (see Figures 4.5a and 4.5b). In the case of the music booth, moving the pointer over each note shows the corresponding song name. The selection of songs for a shared social environment is another area that could benefit from more research. Although originally the system design included a third sound area playing ambient or nature sounds, such sounds do not always work well in an environment whose look and feel are not nature-like. Whyte, for example, found that while waterfalls are enjoyable and relaxing to conversants in plazas, they sound noisy and unfamiliar when played back from tape outside the context of a plaza.

Finally, it is worth noting some of the variety of possibilities the booths provide. Some participants, for example, might use the music booth as background muzak while for others listening to the music might be the primary activity or even something they do while waiting for interesting people to show up, but even then they do so within the context of a social space with others and ongoing activity. The booths' differentiation of the space can also let people observe that certain others can usually be found in the quiet areas away from the booths, while a particular person often listens at the news booth and thus might be good to talk to about current events. The booths, that is, work best in combination with the system's other features to create a lively, legible social environment.

## 4 The Implementation of *Talking in Circles*

*Sociofugal space is not necessarily bad, nor is sociopetal space universally good. What is desirable is flexibility and congruence between design and function so that there is a variety of spaces, and people can be involved or not, as the occasion and mood demand* – Edward T. Hall

### 4.1 Design Requirements and Iterative Prototypes

The requirements for *Talking in Circles* focused on full-duplex audioconferencing between a substantial number of simultaneous users, where we defined substantial as approximately 15. Figure 4.1 shows a typical desktop setup for use with the system, including a microphone and headphones to prevent feedback from speakers. Even experimentally, we were interested in low-latency audio. Unlike with pre-recorded streaming, in live audioconferencing it's not possible to pre-buffer or pause when there is network congestion. Lag is known to be detrimental to the use of speech for social interaction, for example leading to greater formality [O'Conaill 1993]. However, we were also interested in creating a system that could have as wide a user base as possible, important given our focus on social applications as well as to facilitate wider, extended study of the system's use. This meant that we could not use proprietary broadband networks or high-speed LAN's, as previous systems have typically done.



**Figure 4.1: Typical desktop setup for a *Talking in Circles* session**

The initial aim was to design for the internet as a whole, but this proved intractable, as even highly compressed protocols such as RealAudio occasionally suffer from unpredictable network delays and must pause to rebuffer [Progressive]. A prototype was then attempted using the Java Sound API [Java Sound], but resulted in measured end-to-end lag of two to three seconds for machines on the same high-speed LAN subnet. Finally the *RAT* package was taken as a basis and modified for use with *Talking in Circles*. *RAT* is the *Robust Audio Tool*, an open-source audioconferencing tool from University College, London [UCL]. *RAT* uses the MBONE, the internet's multicast backbone, for network transport [Savetz 1996]. Thus we avoided inefficient strictly client-server and peer-to-peer architectures. A client-server architecture, adding an extra hop and centralized processing between each transmission, does not scale when there is a large number of clients or when these clients are geographically distributed. Peer-to-peer connections also require  $O(n)$  work for sending out the audio to each participant. Multicast, by contrast, allows each client to transmit its audio only once and then replicates it to the other clients, resulting in a fully-connected graph architecture where the outgoing edges need be traversed only once for constant  $O(1)$  transmission cost.

*RAT* was adapted to support communication of participant state ( $x/y$  location, circle color, instantaneous audio energy, and so on) and end-to-end lag was then measured at approximately 0.5 seconds, considerable but not detrimental unless participants can also hear each other directly

[Krauss 1967]. The bottlenecks in our current implementation are local ones, primarily high-frequency reading of the microphone, due to the Windows Task Manager's coarse threading, as well as screen redraw, and accordingly we have noticed no substantial performance degradation when varying the number of users from one to eleven. An additional advantage of using multicast throughout for communication is that data besides audio, particularly drawing, are also transmitted to all participants with very low latency, which enables participants to observe and discuss drawing as each pixel or stroke is added.

Although the audio code, including compression/ decompression and MBONE transport, is written in C, the user interaction portion is maintained in Java using the Java Native Interface in Sun's Java 2 platform. For example, computations including instantaneous audio energy, background noise suppression and logarithmic normalization to map the energy value onto the circle's area are performed in C, communicated to the *Talking in Circles* interface via the Java Invocation API, and the bright inner circle (see Figure 3.4) is then updated several times per second in the Java component. Lag was a problem with Java's mouse-motion reporting during freehand drawing, which was adequately resolved using Bresenham's line-interpolation algorithm. A port of the system was made to Linux, but the Java Native Interface was not well-developed on Linux and thus the system did not run with full capabilities; performance of the basic system, however, was far superior to the original Windows version. Audio bandwidth use is moderate at 5KB/s per client, too high for scalable use with modems but quite appropriate for use on various forms of broadband network connections. Ongoing and future developments—including the spread of always-on, high-bandwidth internet access; the deployment of networking protocols such as Ipv6 and RSVP, which can provide for bandwidth guarantees for data requiring low-latency transmission; and the development of new data-compression algorithms—promise to make real-time, multimodal social environments such as *Talking in Circles* useable at high quality by anyone on the open Internet perhaps within two years.

## **4.2 Shared and Subjective System Rendering**

By way of summarizing, this section notes the various data elements of the system as produced by the conversants and how these are rendered for each. Figure 4.2 shows the path of these data through the system.

Each participant directly multicasts the following:

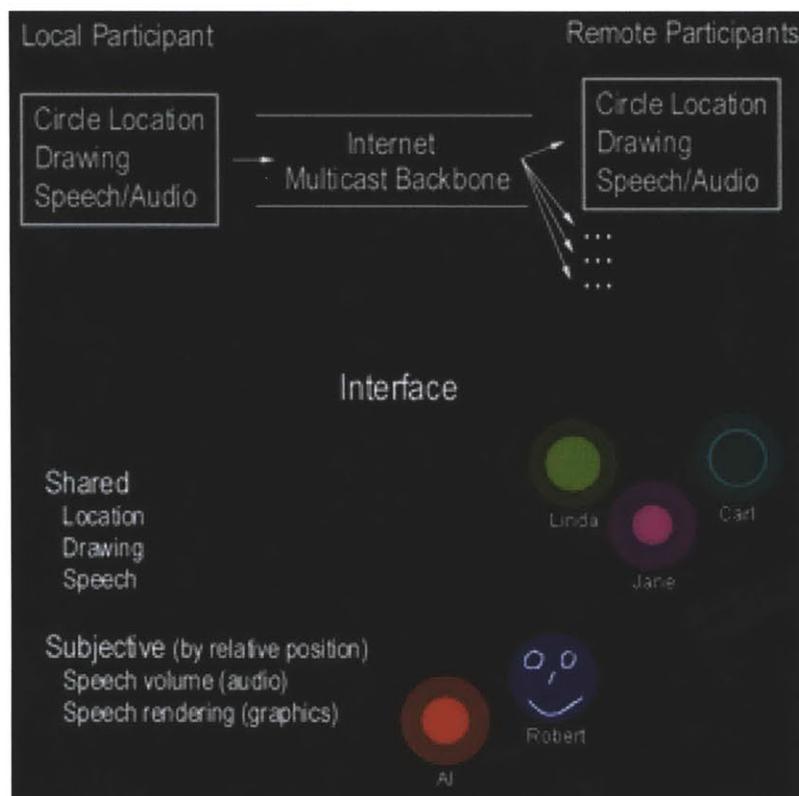
- circle x/y coordinates
- freehand drawing
- icon selection
- instantaneous audio energy
- speech/audio

The interface is rendered from these features, and all participants' displays share identical views of:

- circles' location
- circles' drawing/icon display
- participants' audio

Finally, the local user's relative location produces a subjective rendering of:

- speech volume (audio)
- speech rendering (graphics)



**Figure 4.2: Talking in Circles system diagram showing an overview of network transport of data and its rendering for participants**

### 4.3 Basic *Talking in Circles* Graphical Elements

Figure 4.3 shows the *Talking in Circles* login panel. Users enter their name, then choose a color for their circle and join the session. Circle colors were hand-designed for equal perceived brightness, to minimize potential biasing effects of, for example, certain colors seeming more “happy” or “sad” than others. While color design is a complex perceptual problem, Jacobson and Bender note that color perception, while subjective, is in many ways free of cultural and individual influences, and that “predictable visual sensations can be elicited by adjusting the relationships among colors” [Jacobson 1996]. The rest of this section’s figures show the other basic graphical elements used in *Talking in Circles*.

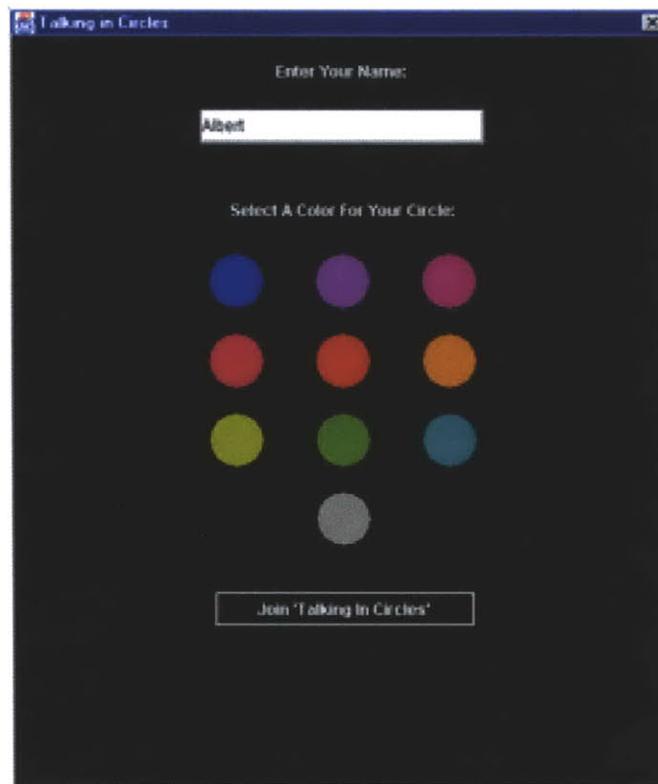
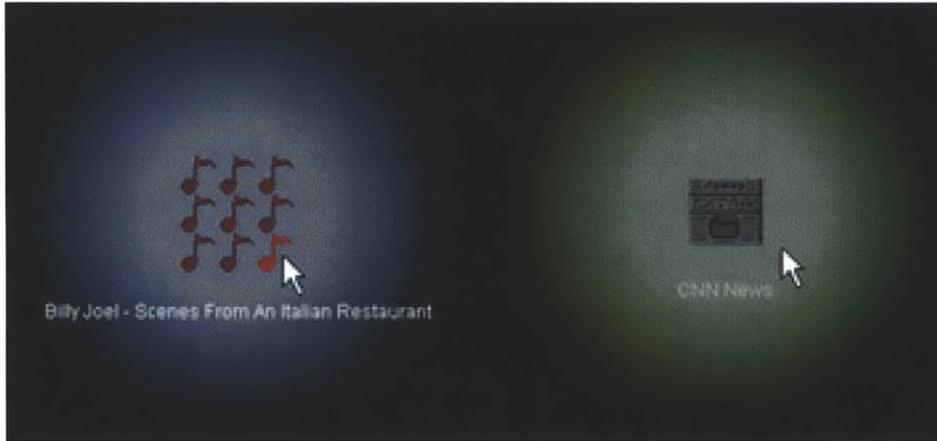
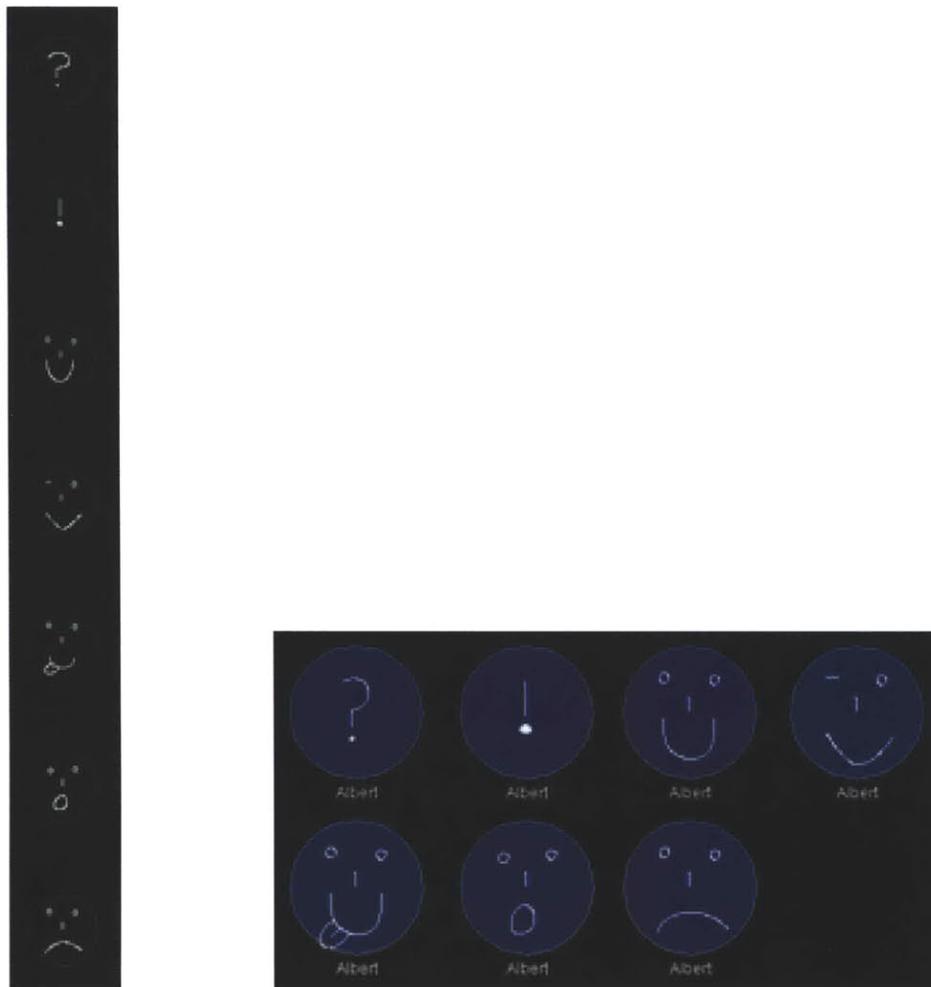


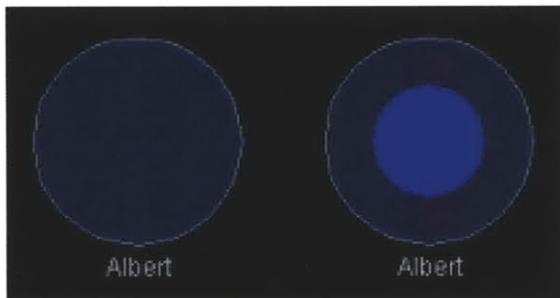
Figure 4.3: *Talking in Circles* login panel.



Figures 4.4a, 4.4b: The music and sound booths.



Figures 4.5a, 4.5b: Built-in icon bar and how each appears when displayed by a participant.



Figures 4.6a, 4.6b: Albert's circle as it appears in *Talking in Circles* when not speaking (left) and when speaking at normal volume (right).

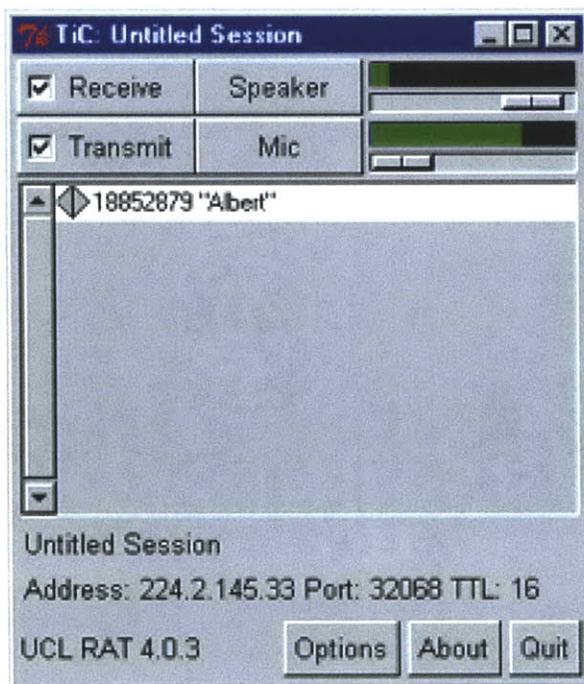


Figure 4.7: *RAT* control panel. Though not generally used during a *Talking in Circles* session, this control panel permits adjustment of the local microphone and speaker amplification.

## 5 User Experiences

*It is my heart-warm and world-embracing Christmas hope and aspiration that all of us-the high, the low, the rich, the poor, the admired, the despised, the loved, the hated, the civilized, the savage-may eventually be gathered together in a heaven of everlasting rest and peace and bliss-except the inventor of the telephone – Mark Twain, 1890<sup>5</sup>*

### 5.1 Evaluation of New Social Environments

The creation of an online environment for social interaction requires consideration of a large variety of factors. Issues ranging from visual design, to computational requirements, to complex social mores all can play key roles in the ultimate success of the design and implementation of such an environment. As Meyrowitz puts it, “The introduction and widespread use of a new medium of communication may restructure a broad range of situations and require new sets of social performances” [Meyrowitz 1985, p. 39].

*Talking in Circles*, as a graphical audioconferencing environment, includes a large number of factors that play a role in these social performances. Several of these have been mentioned, such as the choice of color palette for participants’ representations and the physical model used for audio attenuation through the space. Quality and latency of the audio transmissions and interface responsiveness as a whole are particularly important aspects for interactive communication, and

---

<sup>5</sup> [Hopper 1992, p. 197]

disambiguating their impact on the user experience from that of individual design elements is not easy.

In addition to these implementation-dependent performance factors, the development of advantageous uses of an environment's communicative affordances and of new social performances requires several weeks' use of the system by a stable population [Erickson 1999-R, Erickson 1999-S]. Though many users, for example, called *Talking in Circles* "very intuitive" and were able to use it as they expected, only a long-term field study of use by a moderately-sized distributed group in which at least some participants are strangers can verify that these reactions apply to broad use of the system in its intended context. As this scale of study is beyond the scope of this thesis, the following sections concentrate on observations of common experiences by a large number of non-expert users.

*Talking in Circles* was in development for approximately a year. Over this time around 80 people have used the system to various degrees, usually after a brief introduction to the basic functioning of motion and speech, up to five simultaneous users at a time for up to half an hour per session. Users have ranged from Media Lab students, to Lab sponsors from research labs and industry, to student and faculty visitors from various universities.

## **5.2 Encouraging Experiences**

Reactions have generally been highly positive across the range of design features of the system. Speech communication, foremost, has been very welcome, and people (generally non-expert users of text-only computer-mediated communication) have stressed their experience of greater freedom of expression and corroborated the importance of tone of voice for their social conversations. A few users who heard the half-second latency asked about it, but in all cases where users were not colocated or could not hear each other directly they felt the audio was immediate and responsive.

The graphical feedback for speech, in the form of the bright inner circle, was initially unfamiliar to users but they quickly grasped and used it. As discussed in Chapter 3, an informal experiment had suggested that the feedback was indeed useful and allowed participants to identify speakers they hear within a few seconds. The icon bar, particularly the ability to click on icons in sequence to produce basic animations with the expressions, has been easily understood and used often. The combination of motion with the icons, such as showing a winking face and moving toward a

particular conversant or showing the gasping icon and backing away from someone, has also been popular, although not used as often.

The system's visual design itself received occasional comments, with users saying they found it fit the underlying interaction design. The informal look of the icon bar images was mentioned favorably and, interestingly, almost no users requested the ability to use custom avatars. As users become expert through working with the system for longer periods, visual design is an area they will likely have greater feedback on.

Social factors in the design were often mentioned. Users liked the sound booths' flexibility as foreground listening or background sound during conversation, and the booths' more contained audio dissipation which allowed them to converse with others outside the booths. They also were interested in the booths' contribution to the variety of the space and the possibility for shared context from seeing others' listening patterns over time. A few users joked about the lack of explicit barriers to others following them around the room or moving away from them during a conversation but said they liked the ability to mediate their distance from others through motion and conversational grouping.

Motion around the space and the corresponding changes in audio were regarded very favorably. Even when users had been told about the system's distance-based audio fading they were often surprised when they began to try it extensively and often commented on how smoothly and intuitively it worked. The generous thresholds for the maximum distance at which audio could be heard seemed to work well. Users liked having ample room for forming conversational groups while being able to faintly hear others far away. They also had no trouble positioning themselves near a sound booth so that they could hear its music or news as well as converse with others outside the booth.

While users liked the notion and experience of the audio fading, the drawing capability, and so on, they particularly enjoyed the holistic design of the space and the liveliness the different activities provided. The music and news booths, the last features implemented, were cited as central to the variety and appeal of the system for extended use.

### **5.3 Problems and Limitations**

The major problem was overall system performance. While generally-available hardware resulted in good performance for up to eleven users, the most the system was tested with, simultaneous

activity could cause problems. People moving quickly around the space while talking and while others spoke and drew, for example, caused jerky updates in the graphics. While the audio was generally not affected, the lack of responsiveness in the interface was enough of a detraction to some users.

A particular limitation is responsiveness while drawing. As mentioned in Chapter 4, the Windows implementation had inconsistent performance due to poor thread management by the operating system. Occasionally UI updates would become very slow, rendering the system difficult to use for all modes but speech, and just as quickly the system would become fast and responsive again. Even when the system was highly responsive to motion and other graphical updates, drawing was difficult due to Java not reporting much of the mouse dragging, due to the processor being temporarily in exclusive use for other tasks.

Though drawing was still engaging and useful, it could be much improved. In addition, most users tried the system using a mouse rather than a pen tablet, which contributed to the drawing capability being useable but not ideal. A few used the system with a Wacom LCD pen tablet and found the form factor of the tablet and pen, which they could hold on their lap away from a desk, superior to the traditional desktop setup with a vertical monitor and mouse. The system was tested with a high-quality directional desktop microphone, but while it could be left on the desk or a microphone stand for hands-free speech, transmission was optimal only when it was directly facing users. An LCD pen tablet with built-in microphone or a lapel mike would probably be less constraining.

## 5.4 Specialized Users

Among users, particular groups with specific needs or suggestions emerged. Corporate users reported similar satisfaction and problems as others, but often requested private break-out rooms for small groups of about three people in addition to the larger common meeting space. Though users in general were happy with the distance-based audio attenuation, a few more familiar with audio technology suggested 2D or full 3D audio spatialization.

Another group of users who have found the system highly compelling have been those involved in counseling and other areas involving patient interaction, what Strauss terms *sentimental work* [Strauss 1991]. Like Rutter, Lester has found the telephone quite well-suited to certain uses in counseling and crisis intervention for a wide range of age groups, due to its mixture of audio intimacy with physical distance. In fact, the biggest problem Lester notes in telephone counseling

is the facility with which the medium leads to conversation (not necessarily therapy-oriented) [Lester 1977]. While this is another strong vote of confidence for audio communication in a sociable environment, it's important to keep these findings in mind when considering specialized uses of online environments such as in counseling applications.

## 6 Summary and Conclusions

*A highly imageable (apparent, legible, or visible) city in this peculiar sense would seem well formed, distinct, remarkable; it would invite the eye and the ear to greater attention and participation – Kevin Lynch*

This thesis presented the design of *Talking in Circles*, research toward the creation of a legible and engaging medium for sociable communication. Many factors are key to such a medium, which this design focuses on through the combination of abstract graphical representation with underlying audio communication. *Talking in Circles* employs audio as its primary communication channel, taking advantage of the ease people have in speaking with each other and leveraging their experience in using the telephone for social conversation. Sociological studies of crowd formations underlie the system's distance-based audio attenuation, supporting behaviors such as approaching people one is interested in speaking with and informal distribution of participants into conversational subgroups. The attenuation mechanism allows for clear conversation among participants in close proximity and for maintaining a sense of other conversations faintly in the background

The system's use of graphics allows for participants' speech to be embodied in a holistic onscreen representation, anchoring their voice for others to recognize and combine with their name and circle color into a multifaceted perception of identity. Graphical feedback on participants' circles, matching the rhythm and volume of their speech, eases speaker identification, while the circles also serves as platforms for individual graphical display. Users can sketch on their circle freehand, complementing their speech with explanatory or expressive drawings, and can display icons on them to fluidly convey reactions during conversation.

Graphics are also used to strengthen the system's spatial grounding. As in urban plazas, different areas of the space have different functional properties supported by graphical differentiation. While most of the space has empty backgrounds and is available as quiet conversational areas, two central sound booths add flexibility and variety to the space. Users who prefer some background sound for cover, who would like to enjoy shared music while interacting with others, or who prefer another concrete information source for discussion with others, can move toward areas reached by sound from the music and news booths. Others can meanwhile observe which participants enjoy the various areas of the space, another way in which users differentiate themselves and get to know others, and can take advantage of the booths' content when engaging in conversation with strangers.

Together, participants' cohesive screen representations, ability to navigate the audio space, and lightweight access to displaying drawings on their circle combine to provide *Talking in Circles* participants with broad support for expressiveness, or "multiplicity of cues, language variety, and the ability to infuse personal feelings and emotions into the communication" [Chalfonte 1991]. As an online environment for distributed multiparty communication, the design presented in this thesis resolves issues such as speech-source identification, comprehensive viewing of others in the space and the ability for mutual negotiation of distance with them. The system's spatial grounding supports simple formation of conversational subgroups, and its novel emphasis on embodiment integrity preserves users' personal space while adding a subtle tangibility to motion within a crowd.

User reactions to *Talking in Circles* have on the whole been emphatically positive. The system's cocktail-party metaphor and functional support for it are discussed as refreshing and highly useable. The music and sound booths are considered by users an important aspect of the system due to their augmenting the variety of available settings within the system as well as being interactive, dynamic sources of information and entertainment. Above all, users favor the cohesive system design and the easy access to multiple communicative modalities, which make for a lively space with interesting, observable activity. The most-cited problem is performance, such as reduced responsiveness when a number of participants are highly active at once. Nevertheless, the implementation has given rise to multimodal behaviors combining motion, speech and drawing.

The research presented in this thesis, embodied in the design of the *Talking in Circles* graphical audioconferencing environment, its implementation and user observations, support the literature's

broad findings on the flexibility and richness of human speech for communication. Further, this work exemplifies the great potential in the combination of speech communication with interactive graphics, and illustrates challenges and successes in carrying a multifaceted design based on social factors to fruitful implementation. Chapter 7 now looks at research directions likely to further inform the design of sociable communication online.

## 7 Future Directions

### 7.1 Likely Fruitful Directions

Creating an online sociable environment is a complex challenge, and in working toward it several directions for further research have presented themselves, ranging from graphical techniques to ambitious problems such as browsing conversational history. Among the most accessible is allowing more flexible display of participant identity. Unobtrusive popup displays showing user-selected personal data on hobbies or affiliations, for example, can help participants learn about each other and perhaps give rise to casual conversation about these bits [Rodenstein 1999].

The level of identity information is always an important consideration in social interaction. Voice is a primary identity mechanism in *Talking in Circles*, but some participants may feel their real voice is too personally identifying in certain conversations. Optional pitch shifting of a user's speech could provide them with a comfortable degree of anonymity while retaining the multiplicity of communicative cues it provides. An area within the environment or separate room, for example, could perform varying degrees of anonymizing such as modification of voice and display of all circles in the same color.

Also useful would be further work on the environment's background (or backgrounds, for multiple rooms or a more extended space). While *Talking in Circles* provides sound booths as graphical backgrounds with numerous useful qualities, the space as a whole is still relatively bare or, more to the point, immutable. Some persistence in the backgrounds, whether explicitly allowing participants to paint on backgrounds collaboratively or perhaps to implicitly display

some history of activity or motion as part of the background, some informative wear on the space [Hill 1993], could be quite interesting.

History itself, of course, can be very useful for conversation. *Chat Circles*, which uses a similar spatial metaphor, provides a linear history of messages for each participant plotted over time, which allows one to piece together conversational threads [Donath 1999]. A more spatially-contextualized history function that can allow visual playback of the activity in the space to aid in selecting points of interest in the history and then let users browse the audio from certain groups or areas within the space would be quite powerful. Audio browsing is generally difficult due to its single-threaded, sequential nature, but techniques such as *braided audio* can help explore several segments more fluidly [Schmandt 1998]. Since global speech recording can be invasive, differentiated areas within the space are also useful here, so that participants who would like their conversational history maintained could move within a marked recording area.

Other input/output modalities are also of interest. On the input side, some form of non-explicit input, to complement the existing system's focus on direct manipulation, could work well. Some display of skin temperature or heartbeat, for example, could function as an implicit differentiator of participants as well as add a human element [Picard 1997]. An output device such as a force-feedback mouse, meanwhile, could provide ground for experimentation with greater degrees of tangibility than the current collision-prevention and orbiting provide, as well as permit different parts of the space's background to have a different feel as participants move over them.

## 7.2 Other Potential Directions

In addition to the promising directions outlined in the previous section, two areas for extension of *Talking in Circles* of unclear potential are the addition of a notion of gaze direction and higher-dimensional audio spatialization. Gaze direction has long been known to play an important role in face-to-face conversation. A recent study, for example, suggests that gaze direction is highly correlated with the speaker someone is attending to or the listener they are speaking to [Vertegaal 2000]. One interesting approach is the use of directionalized graphics as participants' representations to indicate their gaze direction [Donath 1995-I]. Explicit indication of gaze direction on the computer screen, however, requires constant manual user input, while automated gaze-direction detection has hardware and performance requirements the average computer does not yet fulfill. It would be interesting, however, to study how gaze direction is affected and used when constrained to a screen-sized space containing representations of numerous people.

As mentioned in the User Experiences chapter, an oft-mentioned feature is 2D or full 3D audio spatialization. While the benefits of spatialization for an increased spatial feel as well as for audio stream separation are important, the mapping from users' 3D environment to the 2D overhead-viewpoint used in *Talking in Circles* may be more confusing than beneficial. Without doing head-tracking, the audio spatialization would depend wholly on the participant's position in the space and not on their gaze direction, unlike spatialization in real spaces [Cherry 1957]. Nevertheless, in limited use left-right spatialization may still be helpful.

## Bibliography

- [Ackerman 1997] Ackerman, M., Hindus, D., Mainwaring, S., Starr, B. Hanging on the 'wire: A field study of an audio-only media space. In *ACM Transactions on Computer-Human Interaction*, March 1997, 4(1), pp. 39-66.
- [Alexander 1977] Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., Angel, S. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, 1977.
- [Beldoch 1964] Beldoch, M. Sensitivity to Expression of Emotional Meaning in Three Modes of Communication. In *Social Encounters: Readings in Social Interaction*, Michael Argyle, ed., Aldine Publishing Company, 1973, pp. 121-131.
- [Bruckman 1997] Bruckman, A. *MOOSE Crossing: Construction, Community, and Learning in a Networked Virtual World for Kids*, MIT Media Laboratory, Massachusetts Institute of Technology, 1997.
- [Chalfonte 1991] Chalfonte, B., Fish, R. S., Kraut, R. E. Expressive Richness: A Comparison of Speech and Text As Media for Revision. In *Proceedings of CHI '91*, ACM, 1991, pp. 21-26.
- [Cherny 1995] Cherny, L. *The MUD Register: Conversational Modes of Action in a Text-Based Virtual Reality*. Department of Linguistics, Stanford University, 1995.
- [Cherry 1953] Cherry, E. C. Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, 25, 1953, pp. 975-979.
- [Cherry 1957] Cherry, C. *On Human Communication: A Review, a Survey and a Criticism*. The Technology Press of Massachusetts Institute of Technology, John Wiley & Sons, Inc., Chapman & Hall Limited, 1957.
- [Cook 1972] Cook, M., Mansure, L. G. Verbal Substitutes for Visual Signals in Interaction. *Semiotica*, 6, 1972, pp. 212-221.

- [Donath 1995-I] Donath, J. S. The Illustrated Conversation. In *Multimedia Tools and Applications*, Vol. 1, March 1995.
- [Donath 1995-S] Donath, J. S. *Sociable Information Spaces*. Second IEEE International Workshop on Community Networking, Princeton, NJ, June 20-22, 1995.
- [Donath 1999] Donath, J. S., Karahalios, K., Viegas, F. Visualizing Conversations. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences*, HICSS, 1999.
- [Eisner 1990] Eisner, W. *Comics and Sequential Art*, Poorhouse Pr., 1985.
- [Erickson-R 1999] Erickson, T. Rhyme and Punishment: The Creation and Enforcement of Conventions in an On-Line Participatory Limerick Genre. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences*, HICSS, 1999.
- [Erickson-S 1999] Erickson, T., Smith, D. N., Kellogg, W. A., Laff, M., Richards, J. T., Bradner, E. Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of "Babble." In *Proceedings of CHI'99*, ACM, 1999, pp. 72-79.
- [Fish 1992] Fish, R. S., Kraut, R. E., Root, R. W. Rice, R. E. Evaluating Video as a Technology for Informal Communication. In *Proceedings of CHI '92*, ACM, 1992, pp. 37-48.
- [Gaver 1992] Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K., Buxton, W. Realizing a Video Environment: EuroPARC's RAVE System. In *Proceedings of CHI '92*, ACM, 1992, pp. 27-35.
- [Gaver 1993] Gaver, W., Sellen, A., Heath, C., Luff, P. One is Not Enough: Multiple Views In a Media Space. In *Proceedings of InterCHI '93*, ACM, 1993, pp. 335-341.
- [Goffman 1981] Goffman, E. *Forms of Talk*, University of Pennsylvania Press, 1981.

- [Greenhalgh 1995] Greenhalgh, C., Benford, S. MASSIVE: A Collaborative Virtual Environment for Teleconferencing. In *ACM Transactions on Computer-Human Interaction*. 2(3), 1995.
- [Hall 1966] Hall, E. T. *The Hidden Dimension*, Doubleday & Company, 1966.
- [Heath 1991] Heath, C., Luff, P. Disembodied Conduct: Communication Through Video in a Multi-media Office Environment. In *Proceedings of CHI '91*, ACM, 1991, pp. 99-103.
- [Hill 1993] Hill, W., Hollan, J. History-Enriched Digital Objects. In *Proceedings of CFP '93*, Computer Professionals for Social Responsibility, 1993.
- [Hindus 1996] Hindus, D., Ackerman, M. S., Mainwaring, S., Starr, B. Thunderwire: A Field Study of an Audio-Only Media Space. In *Proceedings of CSCW '96*, ACM, 1996, pp. 238-247.
- [Hopper 1992] Hopper, R. *Telephone Conversation*. Indiana University Press, 1992.
- [Isaacs 1993] Isaacs, E. A., Tang, J. C. What Video Can and Can't Do for Collaboration: a Case Study. In *Proceedings of Multimedia '93*, ACM, 1993, pp. 199-206.
- [Jacobson 1996] Jacobson, N., Bender, W. Color as a Determined Communication. *IBM Systems Journal*, Vol. 35, 1996, pp. 526-538.
- [Java Sound] Java Sound API. Sun Microsystems, Inc. <http://java.sun.com/products/java-media/sound>
- [Jeffrey 1998] Jeffrey, P. Personal Space in a Virtual Community. In *CHI '98 Extended Abstracts*, ACM, 1998, pp. 347-348.
- [Jensen 2000] Jensen, C., Farnham, S. D., Drucker, S. M., Kollock, P. The Effect of Communication Modality on Cooperation in Online Environments. In *Proceedings of CHI 2000*, ACM, 2000, pp. 470-477.

- [Krauss 1967] Krauss, R. M., Bricker, P.D. Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustic Society of America*. 41, 1967, pp. 286-292.
- [Kroll 1978] Kroll, B.M. Cognitive egocentrism and the problem of audience awareness in written discourse. *Research in the Teaching of English*. 12, 1978, pp. 269-281.
- [Kurlander 1996] Kurlander, D., Skelly, T., Salesin, D.. Comic Chat. In *Proceedings of SIGGRAPH*, 1996, pp. 225-236.
- [Lester 1977] Lester, D. The Use of the Telephone in Counseling and Crisis Intervention. In *The Social Impact of the Telephone*, Ithiel de Sola Pool, ed., MIT Press, 1977, pp. 454-472.
- [Lynch 1960] Lynch, K. *The Image of the City*. The MIT Press, 1960.
- [Meyrowitz 1985] Meyrowitz, J. *No Sense of Place: The Impact of Electronic Media on Social Behavior*. Oxford University Press, 1985.
- [Milgram 1977] Milgram, S. *The Individual in a Social World: Essays and Experiments*. McGraw-Hill, Inc., 1977.
- [Mynatt 1997] Mynatt, D., Adler, A., Ito, M., O'Day, V. Design for Network Communities. In *Proceedings of CHI '97*, ACM, 1997, pp. 210-217.
- [Nakanishi 1996] Nakanishi, H., Yoshida, C., Nishimura, T., Ishida, T. FreeWalk: Supporting Casual Meetings in a Network. In *Proceedings of CSCW '96*, ACM, 1996, pp. 308-314.
- [O'Conaill 1993] O'Conaill, B., Whittaker, S., Wilbur, S. Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human-Computer Interaction*. 8, 1993, pp. 389-428.
- [Picard 1997] Picard, R. *Affective Computing*. Cambridge, MA: MIT Press, 1997.
- [Pinker 1994] Pinker, S. *The Language Instinct*. Penguin, 1994.

- [Progressive] Progressive Networks, Inc., <http://www.real.com>
- [Rodenstein 1999] Rodenstein, R. *Lightweight Rendering of Individual Identity Online*. Sociable Media Group Tech Report 101, MIT Media Lab, 1999.
- [Rodenstein 2000] Rodenstein, R., Donath, J. S. Talking in Circles: Designing a Spatially-Grounded Audioconferencing Environment. In *Proceedings of CHI 2000*, ACM, 2000, pp. 81-88.
- [Rosenfeld 1967] Rosenfeld, H. M. Non-Verbal Reciprocation of Approval: An Experimental Analysis. In *Social Encounters: Readings in Social Interaction*, Michael Argyle, ed., Aldine Publishing Company, 1973, pp. 163-172.
- [Rutter 1987] Rutter, D. R. *Communicating by Telephone*, Pergamon Press, 1987.
- [Rutter 1989] Rutter, D.R. The role of cuelessness in social interaction: An examination of teaching by telephone. In Roger, D. and P. Bull (ed.), *Conversation: An Interdisciplinary Perspective*. Multilingual Matters, 1989.
- [Savetz 1996] Savetz, K., Randall, N., Lepage, Y. *MBONE: Multicasting Tomorrow's Internet*. IDG, 1996.
- [Schmandt 1994] Schmandt, C. *Voice Communication with Computers: Conversational Systems*, Van Nostrand Reinhold Computer, 1993.
- [Schmandt 1998] Schmandt, C. Audio Hallway: a Virtual Acoustic Environment for Browsing. In *Proceedings of UIST '98*, San Francisco, CA: ACM, 1998, pp. 163-170.
- [Sellen 1992] Sellen, A. J. Speech Patterns In Video-Mediated Conversations. In *Proceedings of CHI '92*, ACM, 1992, pp. 49-59.
- [Sellen 1995] Sellen, A. Remote Conversations: The Effects of Mediating Talk With Technology. *Human Computer Interaction*, Vol. 10, No. 4, 1995, pp. 401-444.

- [Singer 1999] Singer, A., Hindus, D., Stifelman, L., White, S. Tangible Progress: Less Is More In Somewire Audio Spaces. In *Proceedings of CHI '99*, ACM, 1999, pp. 104-111.
- [Smith 1996] Smith, G. Cooperative Virtual Environments: lessons from 2D multi user interfaces. In *Proceedings of CSCW '96*, ACM, 1996, pp. 390-398.
- [Smith 2000] Smith, M. A., Farnham, S. D., Drucker, S. M. The Social Life of Small Graphical Chat Spaces. In *Proceedings of CHI 2000*, ACM, 2000, pp. 462-469.
- [Strauss 1991] Strauss, A. *Creating Sociological Awareness: Collective Images and Symbolic Representations*. Transaction Publishers, 1991.
- [Tang 1991] Tang, J.C. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*. 34(2), 1991, pp. 143-160.
- [Thomas 1995] Thomas, F., Johnston, O. *The Illusion of Life: Disney Animation*, Hyperion, 1995.
- [Turkle 1995] Turkle, S. *Life on the Screen: Identity in the Age of the Internet*. Touchstone Books, 1997.
- [UCL] University College London. Robust Audio Tool. <http://www-mice.cs.ucl.ac.uk/multimedia/software/rat>
- [Vertegaal 2000] Vertegaal, R., Slagter, R., van der Veer, G., Nijholt, A. Why Conversational Agents Should Catch the Eye. In *CHI 2000 Extended Abstracts*, ACM, 2000, pp. 257-258.
- [Viegas 1999] Viegas, F. B., Donath, J. S. Chat Circles. In *Proceedings of CHI '99*, ACM, 1999, pp. 9-16.
- [Whyte 1980] Whyte, W. H. *The Social Life of Small Urban Spaces*. The Conservation Foundation, Washington, D.C., 1980.
- [Whyte 1988] Whyte, W. H. *City: Rediscovering the Center*. NY: Doubleday, 1988