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*[4] 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

[4] 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.