

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].

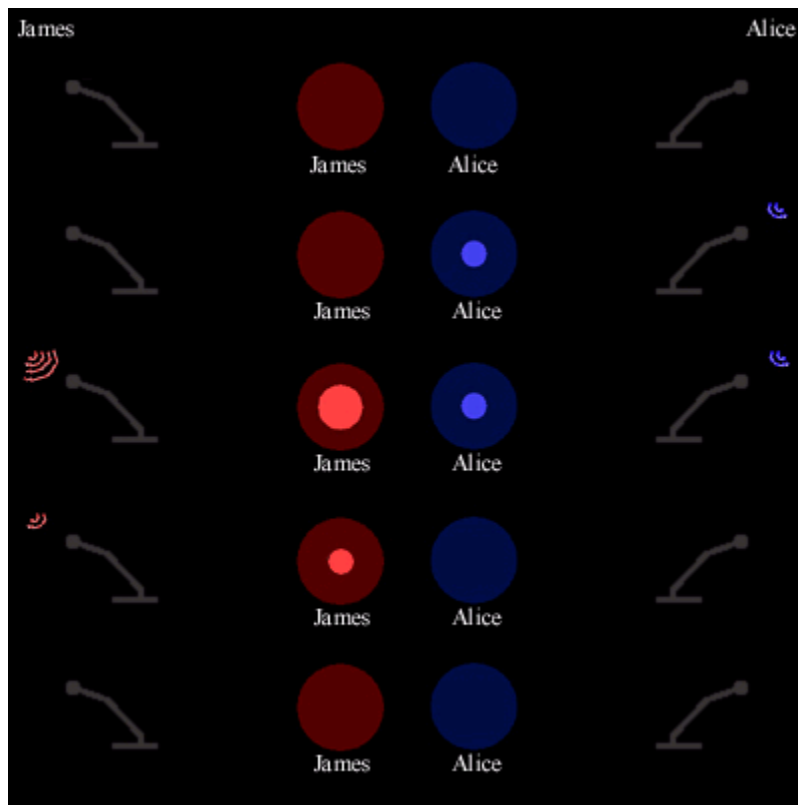


Figure 3.1: Diagram of several moments’ sample speech between James and Alice. As each speaks into their microphone (left and right sides), their onscreen circles (center) display graphical feedback showing their instantaneous audio energy over several moments.

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.

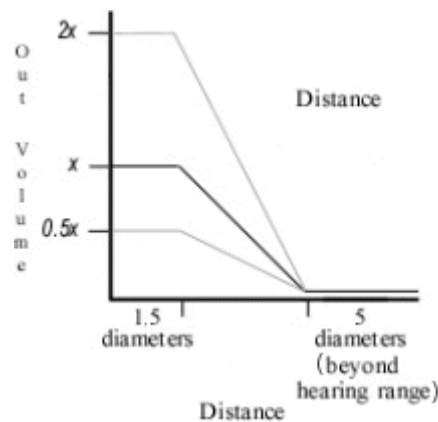


Figure 3.2: Output volume as a function of distance from a speaker, for input volume x (darkest line), $0.5x$ and $2x$.

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.

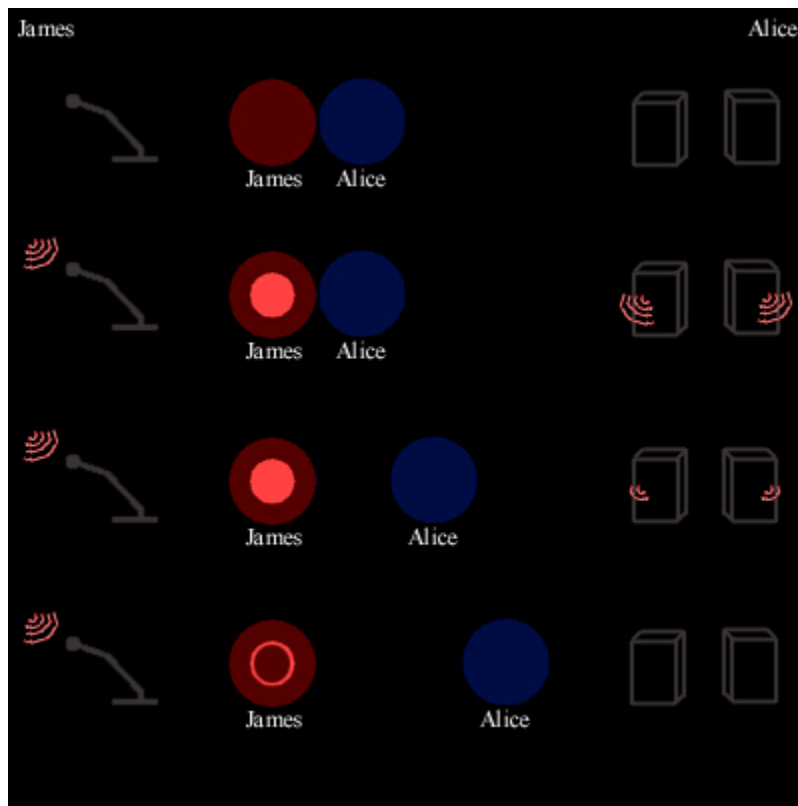


Figure 3.3: Diagram of sample speech between James and Alice. As James speaks into his microphone (left side), Alice moves her onscreen circle farther from his (center). As she moves away, her perceived volume for James's speech (through her speakers, right right) decreases.

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.

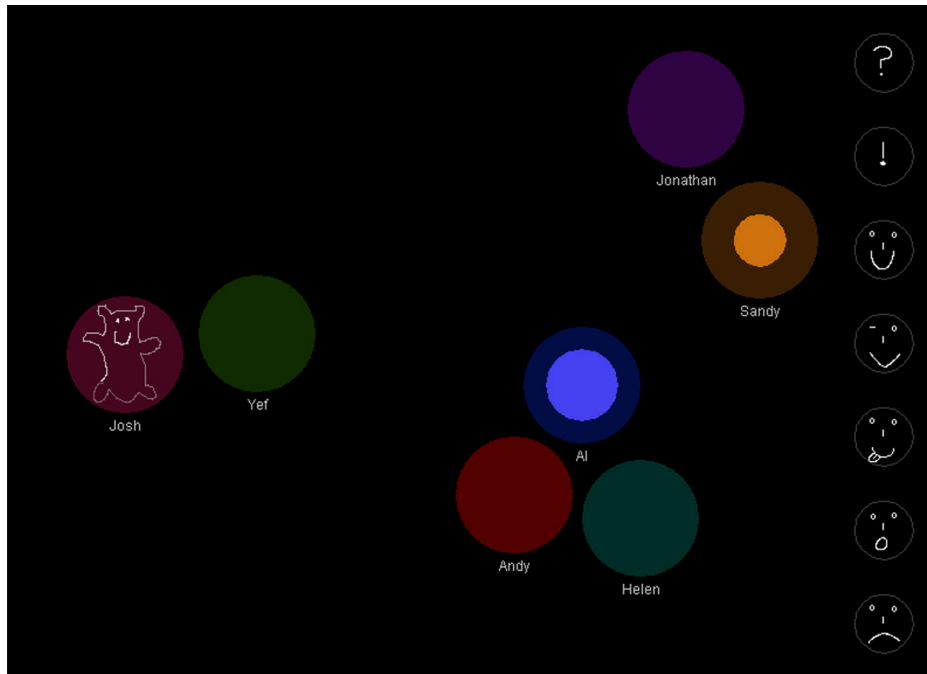


Figure 3.4: Three conversations in a *Talking in Circles* chat session. Al and Sandy are speaking, while Josh draws for Yef.

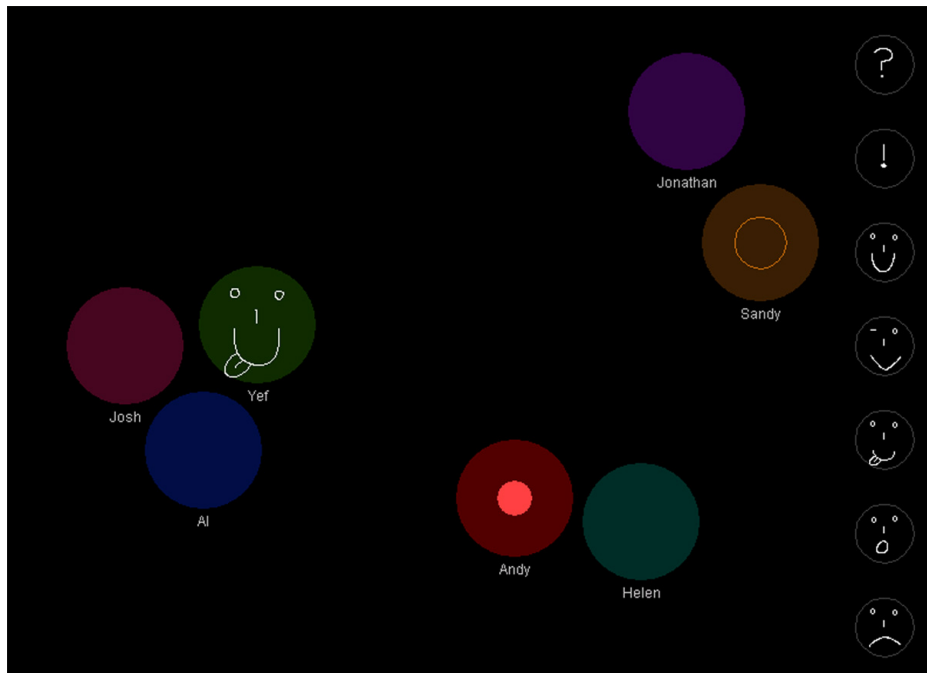


Figure 3.5: Screenshot from the point of view of Al, the blue circle, as Sandy and Andy are speaking. Andy is within Al's hearing range, while Sandy is now outside it.

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.

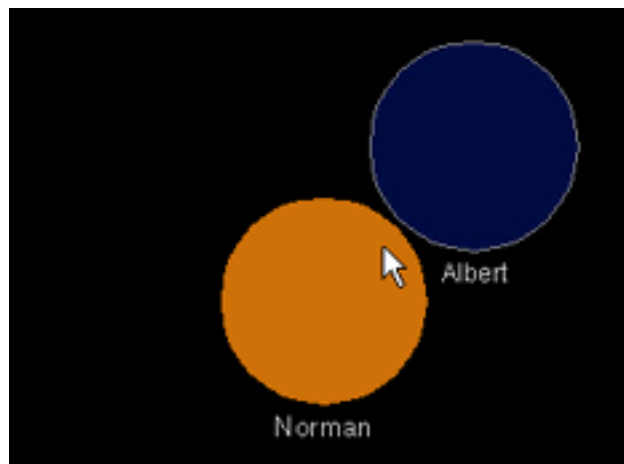
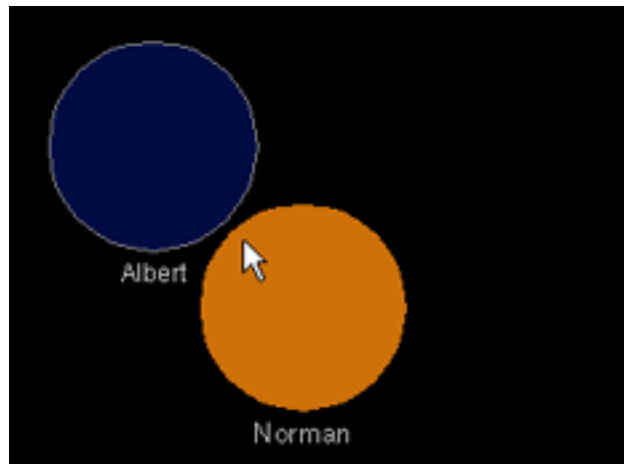
3.3 Circle Motion

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

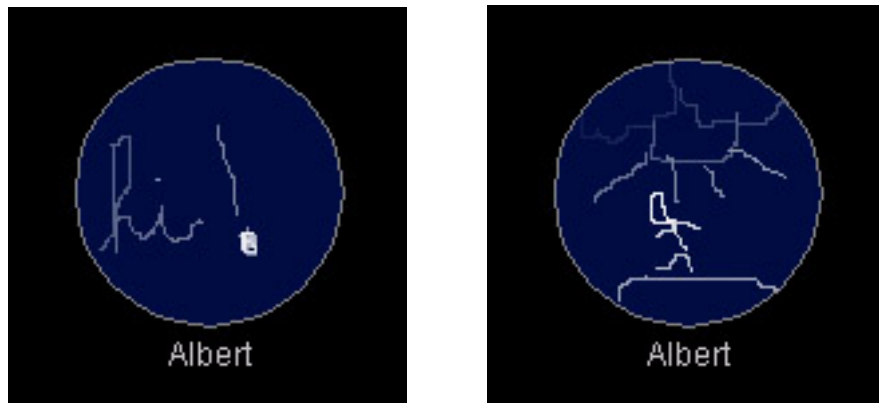
combination of voice, circle motion and drawing is aimed at overcoming some of these limitations.

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.

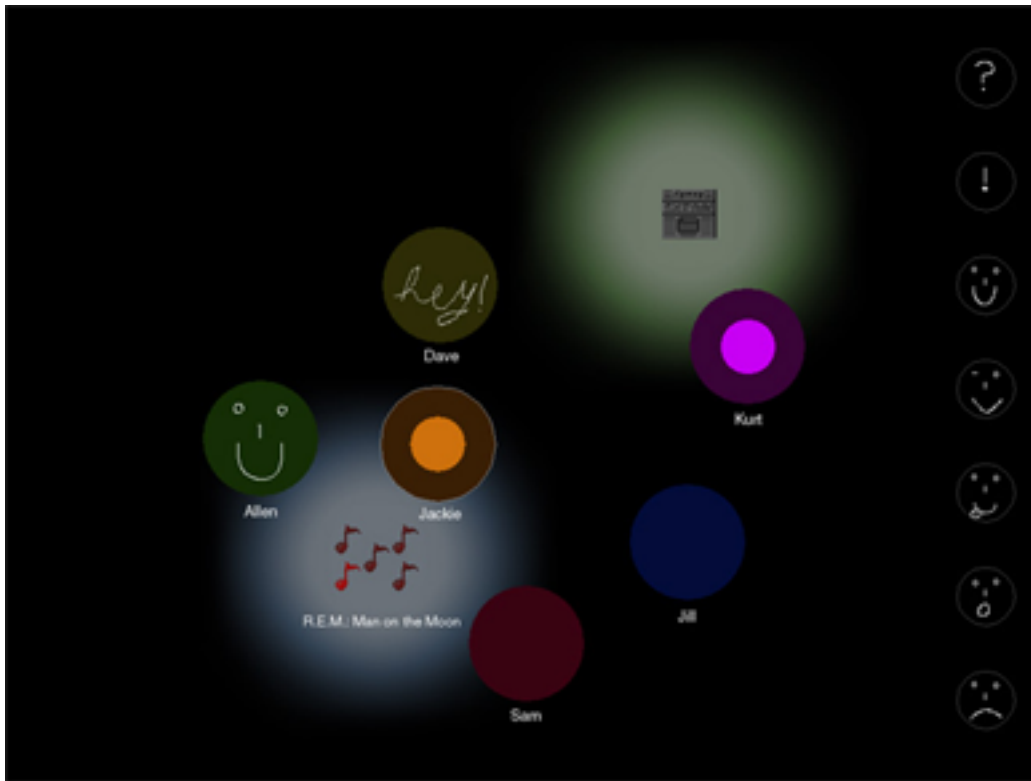


Figure 3.8: Six participants conversing in *Talking in Circles*. Allen, Jackie and Sam are listening at the music booth, Kurt at the news booth.

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.