Conversation Clock: Visualizing audio patterns in co-located groups

Tony Bergstrom and Karrie Karahalios
University of Illinois at Urbana-Champaign
{abergst2,kkarahal}@cs.uiuc.edu

Abstract

Aural conversation is ephemeral by nature. The interaction history of conversation fades as the present moment demands the attention of participants. In this paper, we explore the nature of group interaction by augmenting aural conversation with a persistent visualization of audio input. This visualization, Conversation Clock*, displays individual contribution via audio input and provides a corresponding social mirror over the course of interaction. This paper describes the implementation of* Conversation Clock*, provides observations on an initial pilot study, and outlines the future progression of this research.*

1. Introduction

In this work we create a series of social mirrors designed for conversation. The goal of this work is to augment live interaction by providing a social mirror and visual history to highlight social cues and signals for conversing individuals. The mirror provides the opportunity for an individual to observe and evaluate themselves. These visualizations display the rare view of one's interaction patterns from the third person perspective while participants are physically co-located with their conversational group.

As conversations vary in content and purpose, one can expect variation in structure and interaction patterns [11]. In addition to body language, voice tone, and gesturing, capturing this variation and making it salient provides an additional visual cue that can guide conversation. In everyday interaction a shared social mirror is obscured as individuals perceive themselves from a different perspective than others. The cues we provide with *Conversation Clock* in co-located spaces are meant to be communal. Every individual sees and affects the same mirror.

Talking and meeting with other people around a table is a common event. However, the aural nature of discussion that takes place is ephemeral. Asynchronous methods of review such as tape recorders and transcripts do not add value to the present moment but can be invaluable after the fact. As individuals converse, subtle patterns of interaction like interruptions, turn taking, mimicry, sentence repair, and conversational dominance fade away to make room for more immediate events [1][16]. When people remember events, a period as brief at 5 minutes is long enough to become susceptible to errors in their memory [15]. Even without memory errors, each participant interprets the interaction from their own perspective and may "remember" it quite differently from the other members. This memory may further be altered over time or as new content is associated with it. *Conversation Clock* presents abstracted forms of this history in order to provide socially translucent cues that depict the interaction rhythm and conversational patterns that are present but not obvious in typical aural group conversation.

Incorporating social cues into interactive systems has long been a goal of computer supported cooperative work. The focus has mostly examined remote collaborative spaces. Work such as Media Spaces and other video conferencing and support tools sought to convey many of these cues directly with real-time video and audio [2][10]. Others, such as Chat Circles and Babble, have attempted to abstract interaction to provide new social cues in the digital environment [4][5][6][7][8].

Research has not focused on the extension of new conversational cues into shared co-located physical interaction environments. Individuals in this environment already have a full set of familiar sensory cues that have been developed naturally via conversing and interacting over the course of his or her lifetime. People learn how take turns while conversing, know when an audience isn't understanding, and prompt speakers to continue by observing others around them. These cues establish the social translucence groundwork Erickson and Kellogg extend into remote environments [8]. Face-to-face interaction provides the visibility, accountability, and awareness to allow and support rich communication.

This paper examines interaction with augmented cues in a shared physical space. We describe the abstracted cues of the Conversation Clock visualization, how they developed, and the details of its implementation. Observations are presented from an initial pilot study highlighting positive and negative aspects of the visualization. We then conclude by presenting our path for future research using the Conversation Clock and its derivatives.

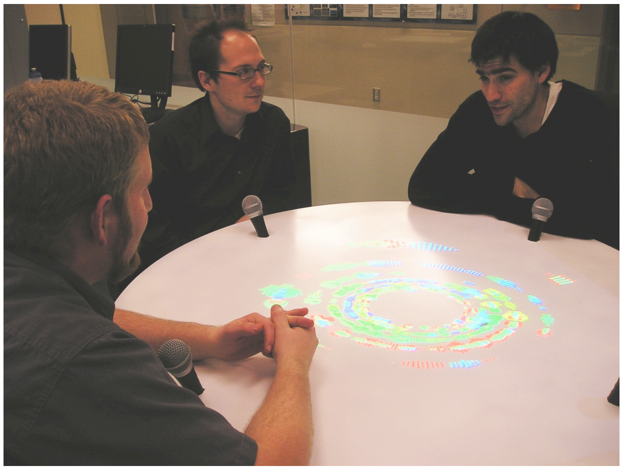

Figure 1. Interaction around the table is monitored via microphone while the *Conversation Clock* is projected on the center

2. Social Mirrors

Group interaction has a different set of social norms than dyadic interaction. The greater number of participants requires coordination of speaking turns. This negotiation takes different forms with regard to the relationship between participants. For example, a group of students might communicate informally and have several speakers vying for the lead, whereas a professor instructing a class might dominate the discussion and select other speakers explicitly. Individuals in a group naturally adopt different styles of interaction based on the composition of the group [19]. These styles might be affected by status, age, experience, skill set, and cultural norms among others. This group context affects patterns of interruptions, turn taking, mimicry, sentence repair and conversational dominance. In one group an individual might be reserved and wary to speak while loquacious and dominating in another.

Conversation Clock is a tool we use to mirror such interaction. The *Conversation Clock* aims to not only demonstrate these patterns, but to allude to the underlying roles of the participants.

Previous work with Visiphone looked at remotely mediated conversation between two locations [5][6]. It showed that the visualization allows you to "see things you know, but do not realize you know." Visiphone directly influenced the interaction between two parties without explicitly providing direction. Participants in the local spaces adjusted the volume of their own speech to visually match the volume in the remote space. How does this interaction change when there are no remote individuals? Does a visualization provide the same influence when participants are in the same space?

Social mirrors provide insight into the participants' culture and status. Rules of culture and language define the meaning of interaction as it progresses [3]. Each conversation, meeting, and interaction share similarities, but ultimately the social mirror must be viewed in the social context to make sense. For example, a hierarchical structure, such as boss/employee relationship should be interpreted based on this knowledge. One would expect the boss to demonstrate leadership characteristics in the social mirror. Even among peers, interaction has been shown to change drastically when the makeup varies by sex, degree program, or conversational style [19][20].

There are many instances where people find themselves in conversations around a table. A family eats dinner and discusses the day's happenings. Friends gather at a coffee shop or bar to relax, chat, and catch up on events. Coworkers and colleagues gather around tables to collaborate. Leaders gather advisors in order to make informed decisions. We believe each of these scenarios would benefit with the addition of a visual social mirror.

3. Related Work

DiMicco et al describe groups of four and how their participation might be affected by a shared display [4]. The display was projected onto a wall and mirrored on the opposite wall. Participants in this system must look away from the conversation in order to view and interpret the displayed information. The visualization consisted of a labeled histogram with each bar depicting the corresponding participant's contribution to discussion. Contribution was an aggregated value of historical data, but the interaction history was not made available. Based on the measured contribution, the shared display explicitly labeled participants as "over-participating", "participating", and "under-participating" on the shared screen. The goal of this system was to aid group decision making by balancing interaction during discussion. During a study of group problem solving situations, DiMicco measured the change of participation levels. They found that the over-participators were more likely to back off than the under-participators were to speak up.

Research conducted with Visiphone offered both visual and aural connections between remote locations [5]. The visualization conveyed aural contribution by adjusting the diameter of circles as they spiraled from the top of a dome and down the sides. The device was shown to display patterns of conversational dominance, turn taking, and interruption. During use, both remote and local individuals were found to similarly interpret the visual depictions of their conversations. Interaction was altered as participants reported finding themselves

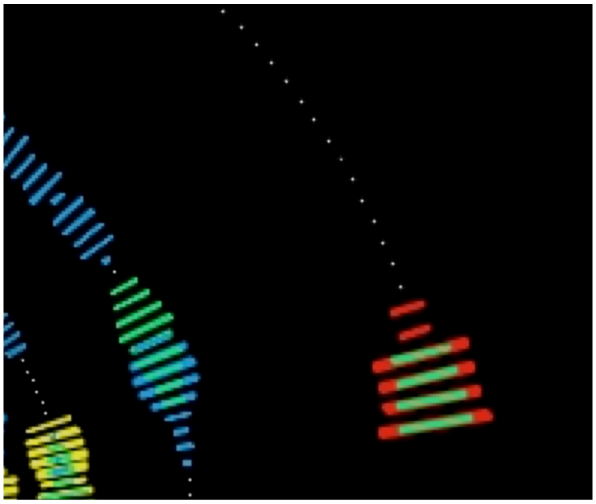

Figure 2. In this close up view, we see dots indicating a period of silence. These dots allow the participants to know the visualization is active and to find the radial position of the present moment.

matching each other's volume by sight rather than by ear. Additionally, conversational dominance became obvious as the dome's color becomes dominated by a single hue. Due to the salience of balance in communication, Visiphone was recommended for use in areas like marriage counseling.

Numerous artistic installations have visualized live conversation and sound [12][14]. These installations have been successful at encouraging individual and group interaction by allowing users to manipulate and construct a visual depiction of their aural input. Rendering the conversation in one installation, Telemurals, increased conversation to at least five times it's normal level [12].

In the following sections, we describe the interaction visualization, *Conversation Clock,* and the iterative design process that resulted in the current implementation.

4. Conversation Clock

The *Conversation Clock* visualizes the conversation of up to four individuals sitting around a table that doubles as a projection surface. During group interaction, table microphones capture individual audio streams from all participants. The visualization of all streams is then projected onto the tabletop surface in real time.

4.1. Visualization

4.1.1. Abstraction. Participation is measured by the volume of aural input. Aside from capturing the binary indicator of talking and not talking, volume can indicate the degree of emphasis a person gives his or her speech. Louder volumes can indicate confidence of the speaker, desire to lead the conversation, or attempts to be heard over other speakers.

Individuals sit at the *Conversation Clock* table (*Figure 1*). Each seat is associated with a microphone to capture input. The speech captured from individuals is represented by distinct colors (red, yellow, blue, and green). As the individual chooses to participate, aural contribution is rendered as a rectangular bar drawn in the associated color along a circular timeline. The length of this rectangle indicates the average amplitude one's speech and utterances. A series of these provide an abstracted visual history of conversation over time (*Figure 2*).

4.1.2. Simultaneous contribution. Multiple speakers often contribute to a conversation simultaneously. Turn-taking, arguments, and back channel conversation provide simple and common examples of simultaneous contribution [17]. Simultaneous contributions can indicate negotiation between speakers, back-channels, or a common reaction such as laughing. To capture simultaneous contribution, each bar along the clock is actually a layering of rectangles from largest to smallest (*Figure 2*).

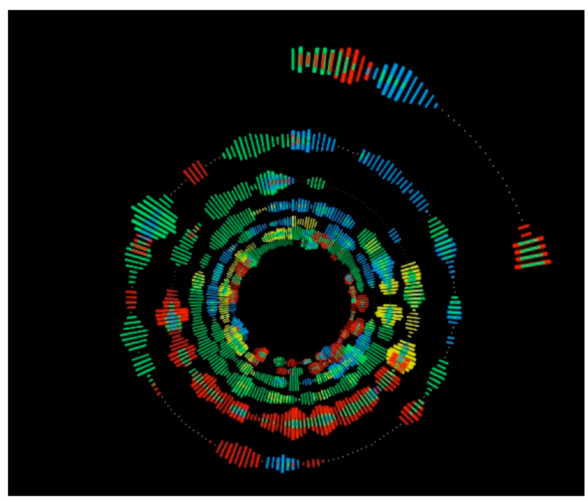

Figure 3. The *Conversation Clock* provides a visual history of interaction and communication. Each participant's contribution is highlighted by different colored bars. The lengths of these bars indicate the degree of participation measured in volume. As a conversation progresses, a history is built with concentric rings reminiscent of the rings on a tree.

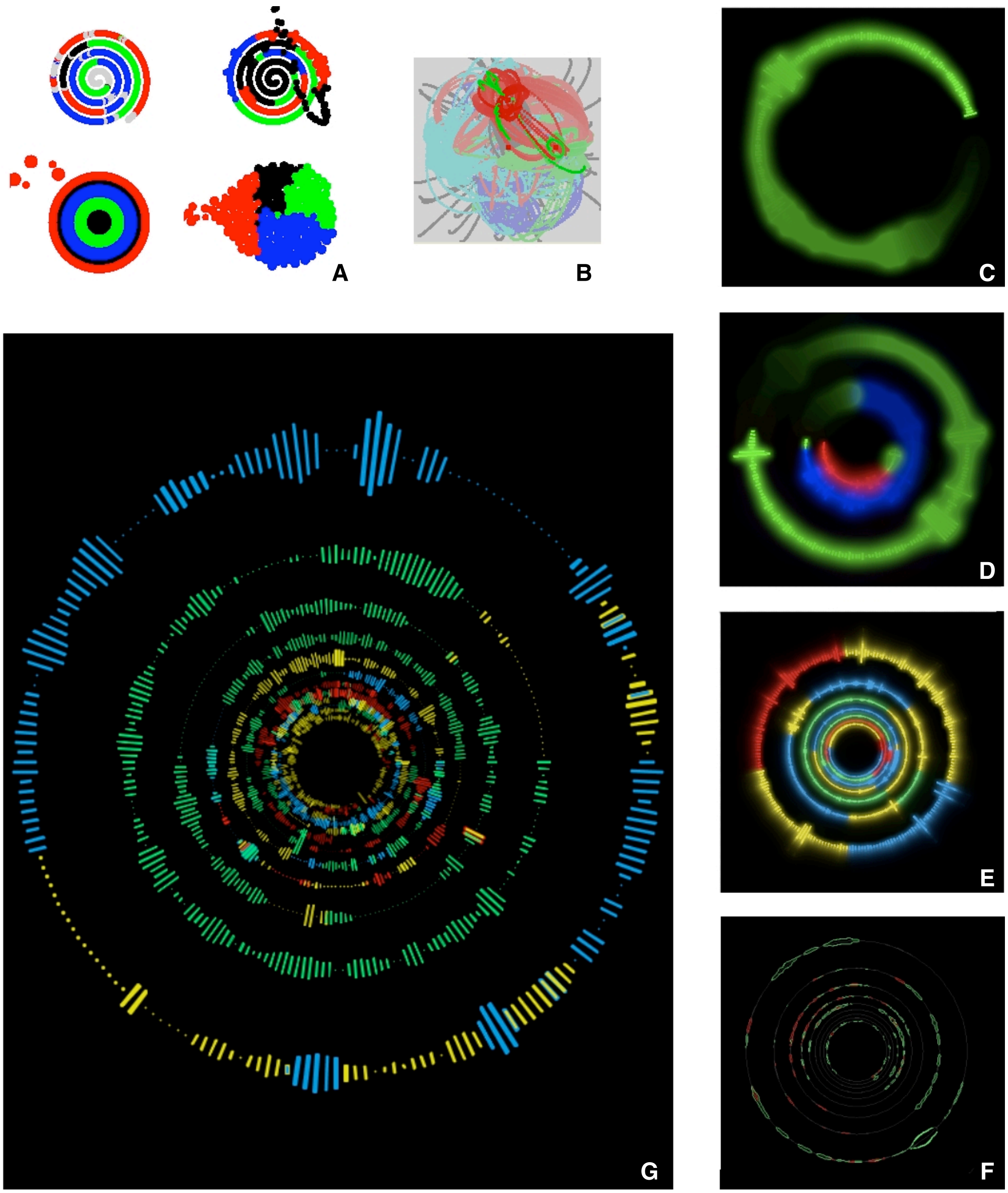

Figure 4: The conversation visualizations progressed through many iterative designs during its development. Each prototyped design provided merits and flaws to be considered in the next iteration. Early stages, (A) and (B), focused on showing individual contribution and highlighted the need for displaying history. History became more explicit, as seen in *(C)*, *(D)* and (E) along the right column, but the the visualization failed to present simultaneous contribution. *(F)* Among other methods, we tried overlapping bezier curves; however as speakers approached similar volumes, one was always obscured. *(G) Conversation Clock* began to overlap bars to show this contribution, ensuring each contributing color is visible.

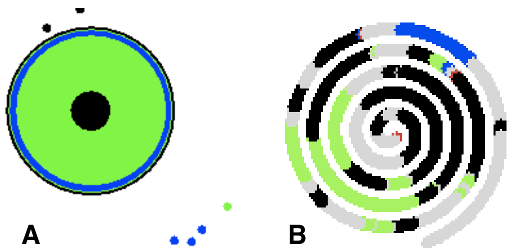

Figure 5. The bubble visualization (A) grows as more people contribute. Each bubble represents a volume sample, they combine when they collide. This visual was not effective, new contributions (the outside) cannot easily be distinguished. The spiral (B) attempted to show a more explicit history. However, as the colors spiraled inward, many commented on the resemblance of a flushing toilet, which was not a desirable metaphor.

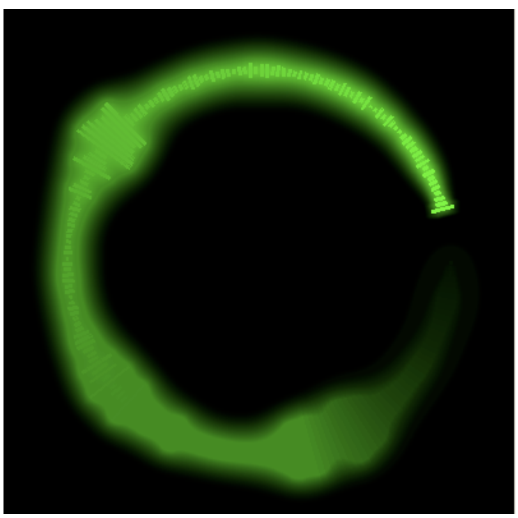

Figure 6. This *Conversation Clock* prototype incorporates a glow that solidifies over time. The radar-like fade provided a less distracting transition element than previous prototypes (Figures 4a, 4b).

4.1.3. Time metaphors. The *Conversation Clock* utilizes the metaphor of the clock as the primary representation of time. As time passes, the progression of bars is drawn clockwise around a circle. A full circle represents a minute of conversation. A complete minute contracts into the center, making room for the next minute rendering. The resulting concentric circles

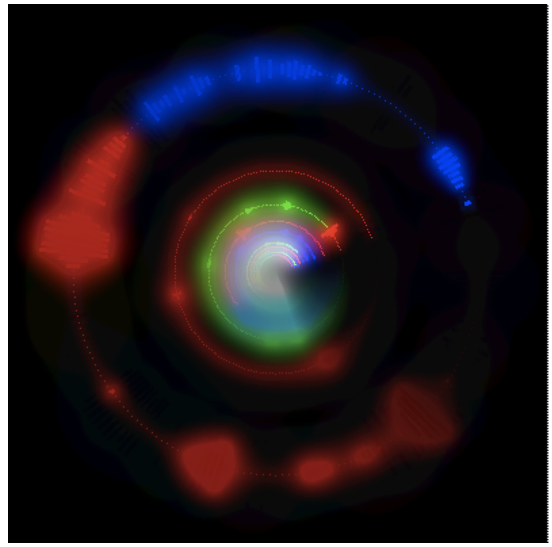

Figure 7. This prototype began the incorporation of concentric rings to display the history. The addition of this history contributed to a visual distraction as the large wedge of fading and reappearing tic marks drew the eye to the table. Additionally, the glow of the rings began to obscure the central portion of the visualization, negating benefit of the larger history.

show the passage of minutes analogous to the passage of years in a tree's rings (*Figure 2, 3*).

4.1.4. Silence. Work done in Visiphone indicated that displaying small dots was useful in conveying that a connection was present between remote spaces [5]. Though we did not deal in remote connections, we also chose to adopt this indicator for two reasons. The silent dots offer a visual cue that the system is active. Additionally, it allows individuals to visually judge the length of the current silence.

4.2. Initial Prototype

The interface of the *Conversation Clock* is a round table, accommodating four individuals. Using the table as a projection surface allows individuals to attend to their conversation and examine the visualization without drastically averting their gaze. With the central location and the circular nature of the visualization, each position around the table offers an entire view of the *Conversation Clock*. For our pilot study, the table held four directional microphones to capture the aural data from each individual at the table. The microphones were mounted directly into the table, allowing one's hands to remain free to gesture and

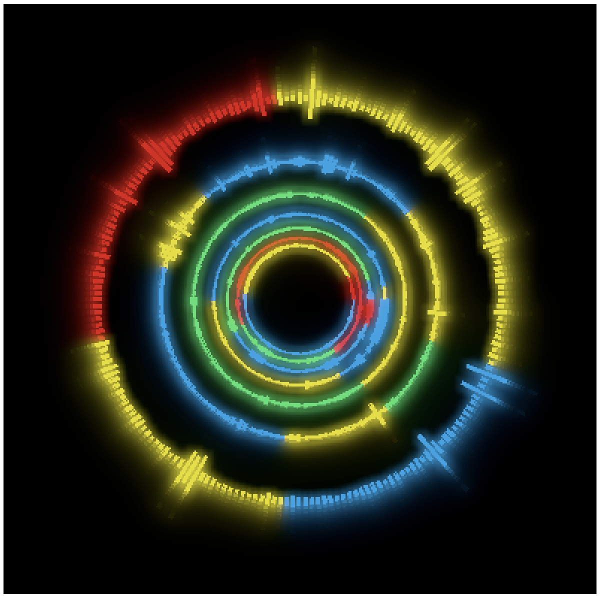

Figure 8. This iteration of *Conversation Clock* separated the animated inward progression of circles from the progression around the circumference in favor of a more readable center. However, the inability to show multiple people speaking simultaneously forced further iteration.

interact with the other participants at the table. The *Conversation Clock* visualization was implemented as a Java applet utilizing Processing libraries [9]. Java Sound API was also utilized in order to access each of the four incoming channels.

To reduce noise, the captured audio had to surpass an amplitude threshold. This value was calibrated for the ambient noise of the room and the response of the microphone. When a microphone's input exceeded the assigned threshold, its associated color was rendered on the *Conversation Clock*. In order for a second, third, or fourth user to be displayed simultaneously, their input must have surpassed a percentage of the dominant speaker's amplitude. This threshold existed to compensate for cross input that occurred when participants were not talking directly into the microphone and their audio was captured by a microphone other than their own.

This initial design utilized table mounted microphones so that individuals would not be bound to the table. A table modeled using this design would be appropriate in public places where anyone could walk up and engage others at the table without needing specialized hardware.

5. Conversation Clock Design Process

Various iterations of the *Conversation Clock* can be seen in *Figure 4*. We chose the table as the object of our social space because it is a commonplace object that is often present during conversations. Additionally, a table also provides a flat surface on which to project the visualization.

With these constraints in mind, initial prototypes of visualizations focused on measuring aural contribution of each individual (*Figures 4a,4b,5*). Most of these early versions highlighted the need for a more explicit and easily readable history. They further revealed that animations showing contribution and history must not distract participants from the conversation. The early visualizations contained too much motion and focused the users attention to the visualization instead of the conversation.

The next iteration incorporated a slow moving radar-like animation (*Figure 6*). The present moment's audio cues would first be drawn clearly and brightly. As the leading edge of conversation progressed around the circle, older contributions began to blend with the colors of their neighbors and then fade to black similar to traditional radar screens that would fade until the antennae made another sweep in that direction. This fade mitigated distraction by gradually altering the image minimally in each moment. This version also incorporated illumination, an effect achieved by blurring the previous visual before redrawing the present state. Colors blended when multiple people spoke as a result of this effect. However, this radar visualization did not provide a history that was long enough to act as social mirror in the shared space. A single person could easily monopolize a full ring around the circle. Though the aesthetics were appealing, the timing of this visualization was not adequate to provide a persistent archive.

The next iteration began representing history as concentric rings (*Figure 7*). Each ring followed the radar-like metaphor. Initially, a bar depicting current input would be drawn, begin to blend, and then fade. As the leading edge of the radar overlapped with older visuals, the faded bars were redrawn closer to the center. This design created a longer history; critiques of this visualization noted that the radar-like metaphor was no longer meaningful with the extended history. Viewers questioned why previous minutes would fade , disappear, and reappear. Additionally, the illumination of the rings blended into a grey color and obscured more than it revealed. It was noted that the blending of color was enhanced by the projection lighting; thus the degree of blending used in illumination could be reduced.

The next revision removed the radar-like effect and retained the concentric clock design presented earlier (*Figure 8*). To reduce the distraction from the animation, rings still form and approach the center but no longer fade. Illumination was toned down drastically but was still incorporated to show a mix of contributors to the audio input.

This latest iteration brought the *Conversation Clock* to it's present state (*Figure 4*). Overlapping the graphics of simultaneous speakers provided the added observation of simultaneous speech.

6. Pilot Study

With our functional prototype, numerous informal and one formal pilot study were conducted to evaluate various features of our system. We sought feedback on aspects of the *Conversation Clock* that were annoying, useful, informative and influential. In this section we describe the feedback from these inquiries.

6.1. Methodology

We performed a pilot study of four graduate students from an HCI research group of the University of Illinois at Urbana-Champaign. These four individuals were familiar with the *Conversation Clock*, but had not interacted with it previously. The group consisted of four graduate students: 3 males and 1 female.

The participants were not given a preassigned topic for the pilot discussion: they met with the goal of discussing the design of a website that they were developing as a research group. The website represented the projects and status of their research group. Links to individuals in the research group, research, news, and publications comprised the main requirements of their design. A site already existed, but in a temporary form that had become obsolete. The members needed to discuss this topic with or without this augmented table. They agreed to use the *Conversation Clock* table and let us videotape their meeting.

Following the meeting, participants were interviewed as a group while still seated at the table.

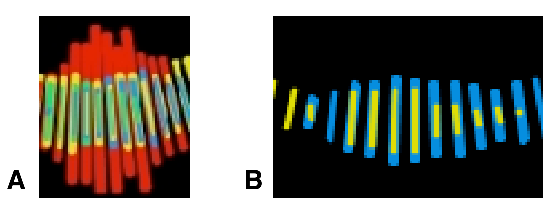

Figure 9. (A) We see how the noise detected in other microphones (inside) can obscure speaker information (outside). (B) A transition between two speakers

6.2. Setup

Microphones were calibrated for each participant prior to the conversation. Each participant was asked to speak in their normal tone of voice while the other individuals remained silent. The microphone's volume was adjusted based on a running average of the last 20 audio samples. When the average stayed within a preset range, the microphone calibration was complete.

Following calibration, participants interacted with the visualization to ensure familiarity with the system and their representative color. They then began their meeting.

6.3. Interaction

Conversation focused on the task at hand. Initially, individuals explicitly watched the table for their contribution as they talked. However, conversation focused on the group's predefined task within the first few minutes of interaction. Conversation generally swapped back and forth between the website and social conversation. Participants indicated the conversation resembled a typical conversation amongst these individuals. Twice during interaction, the group transitioned from website related conversation to explicit discussion about their representation on the *Conversation Clock*. All four participants reported that the *Conversation Clock* did not hinder their ability to converse naturally.

Speakers avoided looking at the table. It was noted in observation and by participants' reports that during the conversation, the speaker rarely looked at the table. The speaker focused predominately on the other participants. As reported in the post-conversation interview, the speaker was looking for the reactions and cues from the other participants. One participant made clear that he often focused on the person he was speaking to and used peripheral vision to monitor the other listeners. As the speaker's gaze commonly connected with that of a listener and not the table, the visualization seems to have less effect on the current speaker.

Listeners observed the table during conversation. Non-speakers were much more likely to watch the *Conversation Clock*. Though there was little self reporting on observing the visualization in this state, video review showed non-speakers shifting attention between the table and the speaker. As these participants have more flexibility in their gaze, they stand to be most affected by the *Conversation Clock*.

Touch-based interaction should be considered as an extension. Though there was no interaction based on computer vision or touch-based interaction, participants intuitively felt it should be present. The non-speaking participants played with the edge of the visualization on their fingers, but were wary to enter further toward the center. One participant specifically wanted the ability to access the audio by touch. Dragging an inner circle to the edge might replay that conversation. Another felt that participants should be able to annotate parts of the conversation by first indicating the segment then tagging it with a label. This could be used later as a highlight for important points or breakpoints during later review or reference. Without any form of interaction with the table, the participants reported a feeling of intrusion when they entered the projection space. Hands, arms and papers remained at the fringes for a majority of the conversation, the exception being explicit references to the *Conversation Clock*.

6.4. Discussion : Problems and Solutions

The pilot test served to highlight areas for improvement in upcoming studies and development of the *Conversation Clock*.

Microphone noise can obfuscate social cues. The most pressing concern of the participants was the microphone noise. The visualization would often show multiple individuals speaking simultaneously in spite of the fact that there was only one active speaker (*Figure 9a*). This is likely a combination of multiple factors. As the microphones are mounted in the table, jarring, tapping, and physical interaction with the table can create noise in the audio signal across all microphones. Activity such as writing and striking the table can be considered notable interaction that should be represented, but the noisy signals (*Figure 9a)* indicated this activity should be ignored for our implementation. The distance between the individual and the microphone also contributed to the noise. A speaker is often speaking over the top of their microphone when addressing the group. Microphones must be sensitive enough to detect their speech in these cases. False positives have been noted to be particularly prevalent in speaking to a participant directly across the table as both microphones, speaker and listener are aligned.

In light of this problem, we have opted to use lapel microphones in our upcoming study. The lapel microphones are much better at detecting their associated individual, however the wire to the microphone creates a tether to the table that restricts motion and requires the user to wear a device. We had previously attempted to keep individuals unrestricted from devices, allowing them to walk up and use the table freely. While this is still a goal, the benefits of using a more intimate microphone are more productive at present for capturing individual speaker participation.

Dynamic calibration should be considered. Calibration also created a problem. During the calibration phase one individual spoke much more quietly than during the majority of the conversation. In contrast, another individual chose to speak loudly and clearly for calibration and softly during the remainder of the conversation. As a result, some of the interaction proved to be unbalanced for these individuals. A more adaptable calibration technique should be utilized in future work. Automatic calibration was anticipated previously, as the goal is to have a freestanding table that is approachable in a social environment or used as a tool for meetings with very little setup.

Aim for smooth transitions. One individual noted that some of the movement was distracting. Specifically, the bars on the opposite end of the table appeared too abruptly. The participant indicated a fade-in mechanism of some kind might be useful to soften the impact of this effect.

Color contrast improves interpretation. Lastly, participants desired more contrast and resolution on the table image. After having seen the display on a monitor, participants noted that the clarity of the display was vastly improved due to the color contrast available. We have acquired a new LCD projector with better contrast and will be used in our future studies.

7. Future Work

The next step is to begin a larger scale user-study in order to thoroughly examine how the social mirror affects group dynamics. In the next series of studies, groups will have their own agendas rather than an experimentally provided one. Groups will meet multiple times with and without the tabletop visualization. This will provide a means for comparing the group dynamics with and without the graphics.

We will be studying groups that already meet periodically. Specifically, we intend to use groups of students that are working collaboratively on group projects, small research group meetings where an agenda is already present, faculty discussing course curriculum, and informal social chats. We hope observing these groups interact will provide insight into meaningful quantitative measures for future studies.

From our observation, we propose observing the following in our upcoming study:

- *How often does a person's gaze fall to the Conversation Clock?*
- *In what states does a person look at the Conversation Clock?*
- *How often does a participant reference the Conversation Clock?*
- *How often and to what degree do participants utilize the central area of the table with and without the visualization?*
- *After becoming familiar with it, how often is the visualization referenced in conversation?*

- *How does the visualization affect objects placed on the table?*
- *How much history is most satisfying?*

In addition to continuing work on the *Conversation Clock*, we are also pursuing a similar visualization, *Conversation Votes*. This visualization allows individuals a silent and anonymous voice to provide direct feedback into the social mirror. In this way, members gain access to an anonymous channel in which they can dissent or applaud without becoming the focus. It allows users to actively send a signal to the table and alter the social mirror with their input. Both visualizations, the *Conversation Clock* and *Conversation Votes*, create new channels to augment live co-located conversation.

8. Acknowledgments

Thanks to members of the Social Spaces Group and members of the ORCHID research group of University of Illinois at Urbana-Champaign.

9. References

[1] Bernard, H. R., P. Killworth, et al. (1984). The Problem of Informant Accuracy: The Validity of Retrospective Data. *Annual Reviews Anthropology* 13: 495-517.

[2] Bly, S., Harrison, S., and Irwin, S. Media Spaces: Bringing People Together in a Video, Audio, and Computing environment. *Communications of the ACM*, 1993.

[3] Chaika, Elaine. *Language: The Social Mirror*. Rowley, MA: Newbury House Publishers, 1982.

[4] DiMicco, J. M., A. Pandolfo, et al. Influencing group participation with a shared display. *CSCW 2004*. : 614-623.

[5] Donath, J., K. Karahalios, et al. VisiPhone. *Proc. of ICAD 2000*. Atlanta, Georgia, United States, 2000.

[6] Donath, J., Karahalios, K., and Viégas, F. Visualizing Conversation. Journal of Computer Mediated Communication 2 (4), 1997.

[7] Donath, J. and F. B. Viégas. The chat circles series: explorations in designing abstract graphical communication interfaces *Proc. of Designing interactive systems*. : 359-369, 2002.

[8] Erickson, T. and W. A. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact*. 7(1): 59-83, 2000.

[9] Fry, B. and C. Reas Processing, <http://processing.org>.

[10] Isaacs, E. A. and J. C. Tang. What Video Can and Cannot do for Collaboration: A Case Study. *Multimedia Systems* 2(2): 63-73, 1994.

[11] Johnson, C., Funk, S. J., Clay-Warner, J. Organizational Contexts and Conversational Patterns, *Social Psychology Quarterly,* 61(4): 361-371, Dec., 1998.

[12] Karahalios, K and J. Donath. Telemurals: linking remote spaces with social catalysts, *Proceedings of the SIGCHI*, p. 615-622, 2004.

[13] Karahalios, K. and T. Bergstrom. Visualizing Audio in Group Table Conversation. *Proc. on Horizontal Interactive Human-Computer Systems IEEE*: 131-134, 2006.

[14] Levin, G. and Z. Lieberman. http://tmema.org/messa/, 2003.

[15] Loftus, E. F. and H. G. Hoffman. Misinformation and Memory, The Creation af New Memories, *Journal of Experimental Psychology:* General 118 (1):100-104 March 1989.

[16] Norman, D. A. The Design of Everyday Things. New York, Doubleday, 1988.

[17] Sellen, A.J. Speech patterns in video-mediated conversations. In *Proceedings of CHI'92*. Monterey, CA: ACM, 1992.

[19] Tannen, Deborah. “Teachers’ Classroom Strategies Should Recognize That Men and Women Use Language Differently.” The Chronicle of Higher Education 37:40 (June 10, 1991).

[20] Tannen, Deborah. *Gender and Discourse*. Oxford University Press, New York. 1994.