For our project, we propose to implement a highly scalable, real-time visualization of face-to-face conversation among multiple users within a shared social space. Through this we hope to represent a variety of salient social patterns in sound-space. Among them is the overall nature of conversation within the space as well as that of discrete conversations. Within those discrete conversations, we can visualize speakers' individual phonemes or highlight key phrases using speech-to-text. Ultimately we hope to augment the audible aspect of social interaction in shared spaces with a highly information-rich visual one. Just as one's hearing is immersed with the sound of conversation when audibly interacting with others, one's vision will also be enveloped with light and visual patterns that arise from that conversation.

The ideal environment and setup for this visualization is a dimly-lit four-sided room upon whose walls we can project the visualization while multiple groups people are standing within the room and in proximity to those walls. An array of microphones would be installed in each wall and would be the source for audio data to be processed. An array of light sensors would detect the cast of shadows against the walls and place visual elements outside of the silhouette or onto the person blocking the light sensor. This ideal case can be expanded to larger rooms, making provisions for more projectors per wall, or more wall surfaces per room, or scaled down to just a single wall with a single microphone only. The choice at which level to scale depends on the type of social interaction one would wish to illuminate. A large cocktail party would demand the widest-scale implementation, while an intimate conversation between close friends would require just the one surface to be illuminated.

For the purposes of this project, the implementation would be on the small side of this scale, beginning first with a single wall and microphone and moving towards two channels of audio input and beyond, adding multiple projections and light sensors. The single microphone would pick up on all conversationalists' speech and recognize the patterns without distinction to geography or source. Two channels would be able to pick up on multiple sources and provide a semblance of linear distance from the sensors. Adding a third would enable the system to provide complete information as to the location of a source, and, given a lack of mobility of speakers, distinction between participants.

The following illustrations highlight the concept of our proposal:
Figure 1. the first user speaking - simple color analogy to represent tone of voice / timbre / etc.

Figure 2. the second user speaking, potentially in a different hue to represent the user, but with different saturation values to represent the same parameters.

Figure 3. both users speaking, but with a gradient to highlight this fact. The implementation would use colors and not grayscale values, but I'm not sure how to do this in Illustrator.
Figure 4. Both users speaking, but with additional visual features to reflect the users' phonemic or key phrasing. Taking a page from Köhler’s book, the first user is clearly speaking with a lot of rounded sounds, while the other many guttural ones.