Re: Thunderwire

Party lines provide continuous aural access to a community. They provide functionality similar to the take-a-penny/leave-a-penny dishes at checkout, but with people instead of pennies. Many people can be in the dish - connected - at one time, and have access to all of the other members. Penny dishes tend to be small, so only one hand fits at a time and it’s difficult for more than one person take a penny at a time. Similarly, only one person can be talking at a time on a party line if maximum utility for the group is desired. Thunderwire is essentially a high quality party line.

Historically, the medium used for party lines has been the consumer grade phone system. Interactions over this system are normally characterized as more formal than those conducted face-to-face. That is, face-to-face interactions typically feature more interruptions and expressiveness. This is mainly due to the low quality of the phone system. When the audio quality is low, it is much more difficult to distinguish nuances in conversation and greater use of correction mechanisms is required. Thunderwire addresses this problem, as it is a custom solution implemented with high quality audio as a top priority.

Low quality audio does indeed significantly impair attempts at communication. One way to realize this importance is to contrast the reaction to low quality audio with that to low quality video. Many video conversation clients have difficulty maintaining a constant framerate, especially when the user’s connection is not superb. Yet the fact that people continue to use these clients indicates that audio quality is much more important. While a lag in the video may be overlooked, inconsistent and low quality audio is a conversation killer. Some will even go so far as to mute the video chat audio and carry on their conversation over the phone system if audio quality is a problem. What can be taken from this is that while video is a useful supplement for maintaining presence, only audio is crucial for communication.

Thunderwire is much like an aural analog to IRC today. Although the nature of the medium does not enforce discrete utterances, most productive conversations will adhere to this format. Each member has access to the conversation between any of the members, yet conversations tend to be amongst a relatively small subset of the membership at any one time. This analogy breaks down when the awareness of presence is considered. Many IRC clients today provide a visual list of participants; even text-based clients have a command to display membership. Lacking the ability to readily provide this information at a glance, audio-only spaces like Thunderwire inspire the development of new norms to cope.

A more accurate comparison would be drawn to tactical communications systems used by the military. These systems are comprised of a high-quality audio channel as well, and many lack a video supplement. A norm embraced by military units for the use such systems is that of “reporting in”. The officer in charge of the unit polls his or her charges; since all members are privy to this group communication, all are able to maintain a sense of the presence of the rest of the unit. While the implication is that an authority figure is central to this model, a poll could be initiated by any member of a relatively cohesive group.

Arguably, text-based systems must address the issue of presence and awareness as well. Use of vocalizations such as “Um-hm” to indicate presence must be explicitly typed out in such mediums. A reasonable conclusion to draw is that while only audio may be necessary for communication, it is not sufficient to provide the richness of communication humans desire. Before learning to speak, a newborn knows its
mother from her scent, her touch, and the sound of her voice. While these senses are not required to communicate information, they do communicate presence better than a thousand words.

Re: In Situ Speech Visualization
Consider a standard keyboard, and the manner in which a typist interacts with it. Now, imagine someone sitting hunched over holding his or her wrists at approximately the same height and distance apart. Then imagine that person’s fingers start to move rapidly and in no discernible order. Communication via the typed word is not natural. Perhaps the closest analog of typing in nature is the removal of lice. While human hands are very dextrous, they naturally supplement rather than embody a primary source of communication.

Movements needed to interact successfully with a keyboard are not natural; they are tediously learned behaviors. Natural methods of interaction such as speech or body language would lead to richer, more satisfying exchanges. This gets at one of the underlying observations of the paper; additional layers of indirection frustrate interaction. A keyboard is certainly not a direct means of communication with a computer or of communication with others mediated by a computer. In essence, the keyboard is a layer of indirection between the user and the user’s goal of communication. Removing these layers advances the ease of use and consequently the richness and satisfaction derived from the experience.

Speech visualizations presented in the paper follow this core principle. People naturally associate visualizations of sound spatially with their source. The environment that we as humans are most adept at manipulating and gathering meaning from is the three dimensional world around us. It is, therefore, natural that the most direct representation of audio should originate from the same source and be displayed in three dimensions. It is also natural to represent sounds with a sharp change in pitch with jagged lines, and those having a gradual change in pitch with smooth lines. As natural as locating the hollow man by searching out his footsteps in the rain.

Re: Seeing More
Bertrand Russell once said;

The whole problem with the world is that fools and fanatics are always so certain of themselves, but wiser people so full of doubts.

With this in mind, it is apparent why so many of our officials are so incapable of governing. The paper suggests that only about 7% of face-to-face communication is composed of the words themselves. Knowing this, politicians don’t need to choose useful words - they just need to be good at delivering the ones they do choose.

The Conversation Clock is a product of this mindset; words and text are discarded in favor of volume and subtler cues such as patterns and structural elements of conversation. Back channels as well as changes over time are readily observable. Indicating feedback with the display of overlapping speech is much more useful than showing only the most prominent. Similarly, using dots to visually emphasize silence draws attention to the role a cessation in conversation plays in communication.

Given both these gauges of their performance in a very accessible way, the Conversation Clock enables awareness and provokes a reaction that is elucidating in its own right. These gauges tended to lead people who normally might dominate a conversation to speak less and quiet members to offer input more frequently. That participants felt a need for fairness or balance in a conversation and worked to institute such a balance is incredibly interesting. Venturing a guess, this may be related to maintenance of an image. Not only is fairness a quality valued by society, but it is more critically so when others
know that the speaker knows that they are taking up more than their share of the time.

A facet of the paper that would have been interesting to develop relates to the use of context in the interpretation of silence and interruptions. While it is true that context can pin down the meaning of such cues most precisely, analysis of patterns might lend some insight as well. Parsing an entire conversation, it might be possible to recognize a particular person’s backchannel feedback and differentiate it from interruption. If that person reasonably consistently provides such feedback, any instance of this pattern could be flagged with this possible meaning. Increasing the sample size to multiple people across multiple conversations would lend even more credibility to this approach. Some commonalities - say, the most frequent format for an attention backchannel - might reveal themselves if this were attempted.

Also intriguing are the participants’ commentaries on the Conversation Clock. That a longer history was assumed to have a greater effect on behavior reveals the common impression that more time will necessarily allow for greater change. This is not necessarily the case; indeed, analysis of the conversations showed no significant correlation. Even more interesting is the comment about training oneself to avoid saying ‘umm’ or pausing. This reveals a novel application for the Conversation Clock. The ability to see the rhythm and flow of a conversation includes the ability to compare one’s own performance against another’s. This type of visualization would be invaluable as a supplement to tutoring in a foreign language. People who normally have difficulty getting just the right sound would have a visual benchmark for comparison. Different languages have different inherent rhythms; visualizing the rhythms could transform a technically precise speaker into a much more natural sounding speaker. Learning accents, dialects and more; countless uses can be imagined. Here, the rain falls mainly on the plane of the conference table.