SimPhony: a voice communication tool for distributed workgroups

Vidya Lakshmipathy
MIT Media Lab
20 Ames Street E15
Cambridge, MA 04105
vidya@media.mit.edu

Chris Schmandt
MIT Media Lab
20 Ames Street E15
Cambridge, MA 04105
geek@media.mit.edu

ABSTRACT

Communication is vital in any workplace. However, as workers become less tied to their desktops and computers, the need to provide them with a mobile communication tool that adapts to their work environment becomes more necessary. This paper describes SimPhony - a mobile, voice-controlled, voice communication system built on a PDA and designed specifically for distributed workgroups. SimPhony supports one-to-one or one-to-many communication with voice instant messages or synchronous audio transmitted over an 802.11b wireless network, and it transitions between different communication styles as messages becomes more frequent. The SimPhony interface looks much like an instant messaging client but is accessible through a voice or visual interface on the PDA or a voice interface accessible by any telephone.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work, Asynchronous interaction, Synchronous interaction, Organizational design

General Terms

Design, Human Factors

Keywords

Mobile communication, voice user interface, voice instant messaging, VoIP, group communication

1. INTRODUCTION

Work groups are distributed over various locations and time zones, and successful teams at work often comprise members with heterogeneous but complementary skills. Additionally, certain work situations require an unusually large degree of data transfer and coordination *between* co-workers in addition to the main set of tasks for each worker. These situations might occur in hospitals between doctors and nurses, in buildings between support staff or in factories on production lines.

Clean rooms in semiconductor fabrication facilities (fabs) also fall into this class of work environment. In this environment, collaborating and cooperating with other individuals who are engaging in their own primary task is critical to one's individual progress. Collaboration becomes a task in itself, one that must be performed simultaneously with several others. Several technicians (techs) are often called on to play two roles or must act as a bridge between two teams. Shift changes between one 12 hour group and the next are often very costly downtimes because of the abrupt disruptions in the flow of information and consequently, the flow of work.

2. THE PROBLEM

Communication in a fab is notoriously difficult because in addition to the time, space, and information flow restrictions discussed above, there are additional restrictions posed by the clean room environment and clothing. Technicians (techs) must wear full body "bunny suits" which restrict their movement, vision, and ability to do precision input. In addition, the types of materials brought into the clean room are restricted, limiting the ways in which information can be recorded and transferred.

For communication, clean rooms are outfitted with wired telephones near each tool set or group of similar tools located in proximity. When a tech wants to communicate with someone in another area or a manager outside of the fab, he or she must page that individual on an alphanumeric pager, give the person his contact phone number and wait at the phone near their tool. In the meantime, the tech has no indication of the whereabouts of that person or how long it will be before that person returns their call. If the issue is a question about the tool, the downtime might be valuable and might further delay the process at other tools, becoming very costly. This often prevents spontaneous interaction and often restricts necessary contact.

3. THE SIMPHONY SOLUTION

What motivated the SimPhony project was the need for a more dynamic and flexible mobile communication solution in a fab. Although the scope of the communication tool proposed in this paper is actually much greater, we use the small user group of fab techs to make specific assumptions about the design and get feedback about the system. The restrictions of their environment make them a particularly focused and interesting group on which to base specific design assumptions.

The SimPhony system uses voice over IP and multicast protocol to transmit synchronous and asynchronous voice over the 802.11b wireless network. The system allows users to define or join multiple groups with whom they can communicate simultaneously or in a "call center" fashion. Users can send voice instant messages to other users or groups or connect and transfer synchronous audio. In addition, they can connect using the same interface to users outside the wireless network over the telephone.

A key contribution of SimPhony is automatic transitioning between voice communication modes. If two (or more) users send each other messages in quick succession, SimPhony automatically switches them to full-duplex synchronous ("telephone") mode. This was done after observing that when IM'ing with text, a session often has long periods of inactivity, followed by a brief flurry of messages, and if message traffic is high, it is convenient to be able to bypass the push-to-talk feature. This feature also adapts to the work style of the user by modifying the demand of the mode to her availability for communication. A user engaged with a difficult task might be

slow to respond to incoming voice messages; with this feature, a user who responds promptly is assumed to be available for communication and is transitioned into a more demanding style. Often users get "stuck" in the initiating communication mode and sometimes the mode is not the most efficient for the task at hand. This feature hopes to bypass this type of "mode inertia" by making communication quick and efficient.

If a user is currently in a chat and another group becomes active or another buddy tries to chat with him, he will be interrupted with 10 seconds of audio from the interrupting session, during which time he can switch to the new session. This "preview" of the newly active session is meant to serve as a topic indicator; if the user is more interested in the interrupting session after hearing briefly what is being discussed, she can choose to tune in. This preview is much like hearing a conversation as its members pass by. When the speakers are in the range of the listener, she can interject or attend to the conversation. Once the speakers have passed by, they are no longer in range and neither the listener nor the speakers can communicate.

3.1 The User Interface

The system can be accessed by three interfaces, two on the PDA and one on a landline or mobile telephone.

3.1.1 Voice

The primary challenge to using voice commands in the SimPhony system was in trying to distinguish voice commands made to the system from voice messages being recorded or transmitted to another user. We chose to make the command mode a push (and hold the record button) to talk. In the record mode, however, the button acts more as a toggle and is pressed once to begin recording or transmission and pressed again to end it. In both cases, the user receives audio feedback so that she knows without looking that the system is ready to record when the button is pressed. A number of short commands allow the user basic functionality of the system from sending or retrieving voice messages to connecting or switching to active groups.

3.1.2 Auditory Feedback

Because the majority of SimPhony's features are usable in a voice-only mode, it is important to provide the user with auditory feedback. Many of the audio cues given to the user represent the state of the system. When the user logs in and out, there are complimentary audio cues, a tone with a rising pitch for login and a tone with a decreasing pitch for logout, which indicate successful action. When a user logs in, a group session becomes active or a voice message is received, different activity indicating tones alert the user. A chime tone followed by 10 seconds of live audio from a group indicates a session becoming active. The 10 seconds of live audio is meant to act as a "topic indicator" for the newly active group. Although the current version of SimPhony uses standard audio cues for group activity and new messages, many of these tones can be individualized so the user has a better idea of who is trying to communicate.

3.1.3 Visual

Although the visual interface for SimPhony is used less frequently, it is important because it allows users to perform some of the high level organizational and navigational tasks inherent in a complex system. The visual interface looks much like those of today's commercially available instant messaging clients. Users can log in and view their "buddy list" and see some high level presence information about their buddies. Groups, in addition to individuals, can also be listed on the buddy list.

The visual interface also consists of the LEDs on the PDA which flash to indicate that the system is recording or transmitting audio. When a user receives a voice message, a message box pops up with the name of the sender. Clicking the audio icon on the box plays the voice message.

3.1.4 Telephone

The telephone interface to SimPhony is slightly simpler, allowing only the most basic but vital functionality. The telephone interface was designed to allow team members outside of the network to access SimPhony. Users dial into SimPhony and use voice commands to navigate through the menus and prompts. Expert users can navigate quickly through the system, using the same voice commands available on their PDA, without waiting to hear the menus. Redundancy is created by allowing touch tones in addition to voice commands to navigate through the system - this prevents any frustration with voice recognition that might occur in noisy environments.

3.2 Architecture

SimPhony acts primarily as a connector between individuals using a PDA or a phone. The intelligent part of the system allows for the connection to take many forms based on the restrictions, needs, and communication behavior of the client. Most of the behavior of the system is controlled and maintained by the server while the clients deal with audio recording and playback and maintain the interface between the system and its users. The PDA is used for its processing power, the high power microphone and speaker, and its ability to work on the wireless network and interact with other machines on this network. The platform is extensible and can also be used to send and receive or share documents, spreadsheets, or other data from other users. The typical client, on a PDA, is written in embedded C++ with a Flash interface. Audio is streamed directly, peer-to-peer, using multicast UDP/IP once the correct IP address is attained from the server. The telephone client uses the same server as the PDA, however, it uses an Intel Dialogic D/41JCT-LS 4-port board to create a gateway between the phone and a telephone client application, written in Visual C++, running on the machine containing the board. This interface is limited to recording and playback of messages. The voice recognition takes place on the server side and is written in Visual Basic using the Microsoft Speech API.

The centralized server also listens to conversations and records them. When a new user joins an existing group, he may optionally "catch up" by listening to the recent previous discussion. The server is written in Java and runs on a Windows PC or laptop. It coordinates communication stores basic user data.