SimPhony — a voice communication tool for distributed workgroups

V Lakshmipathy and C Schmandt

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 IEEE802.11b wireless network, and it transitions between different communication styles as messages become 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. SimPhony was designed specifically for, and with critique from, technicians working in a large semiconductor fabrication plant.

1. Introduction

Whether it takes the form of formal meetings or informal hallway conversations, communication and co-ordination among members of a team is what allows work to get done. Studies show that informal group communication in the form of synchronous, face-to-face contact serves a variety of roles including co-ordination of tasks, collaboration, group building and social communication, and, depending on the job, this type of communication could take between 25% and 70% of the time spent in the workplace [1].

Work groups are now distributed over various locations and time zones, and successful teams at work often comprise members with heterogeneous but complementary skills. Communication is essential for the success of these work groups. Additionally, certain work situations require an unusually large degree of data transfer and co-ordination 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. These environments might have any number of restrictions in addition to having a highly mobile workforce and spatial constraints. The skill set of the team is often distributed geographically as well, requiring that workers in one area often need to consult with managers, specialists, doctors, or support teams in other locations. In addition to the resulting need for communication support, many of these work settings also require intensive hand-off processes from one shift to another; shifts often last up to 12 hours and the collective knowledge of the workers on shift must be transferred from one group to the next. In hospital intensive care, for example, the hand-off process occurs in time that overlaps between two shifts when each patient and the care that he has been given over the last 12 hours must be reviewed while patients continue to be cared for — hence the long working shifts.

communication and co-ordination among members of a team is what allows work to get done

Clean rooms in semiconductor fabrication facilities (fabs) also fall into this class of work environment. In addition to the previously mentioned restrictions, here the task of each individual worker is a critical part of the workflow of the entire group. However, what distinguishes these situations is that the need for co-operation is also imperative. In these scenarios, collaborating 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.

SimPhony is a group voice communication tool using conventional PDAs and voice over IP (VoIP) over WiFi wireless data networking. SimPhony was an attempt to apply our previous work with wireless VoIP and voice control of hand-held devices to prototype a more dynamic and flexible mobile communication solution in a fab, combined with the desire to exploit newly available wireless networking in that environment. What made the problem even more challenging, however, were the specific environmental constraints posed on the techs and the need for a communication system to fit within their current requirements. Nonetheless, we believe that the voice communication techniques developed for SimPhony could be used to enhance group communication in a wide range of workplaces.

Very recent commercial developments in VoIP, including WiFi, and forms of voice instant messaging (IM), as well as conventional and push-to-talk mobile telephony, support the perceived value of voice in mobile communication. SimPhony makes several contributions to this area:

- a voice (and multimodal) user interface to voice messaging,
- a means for automatically switching between asynchronous and synchronous communication,
- audio cues for participation in multiple 'chat' groups simultaneously,
- audio storage and catch-up to provide context when a group member rejoins a chat,
- a full telephone-based client for off-LAN participation.

In this paper, we first describe the fab workplace and its communication and computer facilities. We then discuss SimPhony, firstly from a functional point of view and then the various aspects of its multimodal user interface. Next we discuss how SimPhony could be used by fab technicians, and feedback received from them in our user-centred design process. Finally we will summarise some related work.

2. The setting — communication in the fab

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 and by the environment and clothing. Techs in the clean room are responsible for moving each lot (a set of microchips) through various machines, or 'tools', representing steps in the semiconductor manufacturing process. As most machines are now automated and computer controlled, the majority of techs are responsible for troubleshooting tools when they go down or when an error occurs in the process of fabricating a lot of chips.

Lots are easily susceptible to damage and as a result, clean rooms must be free of dust, debris and static at all times so as not to contaminate the microchips. Workers undergo a strict, lengthy, and costly cleaning process before entering the clean room and while inside are forced to wear 'bunny suits', shown in Fig 1, covering their clothing, shoes, hands, and head. These suits limit not only their ability to do any precision input, but also their peripheral vision and movement. The cleaning process makes it difficult for others to enter the fab and for workers to easily move in and out. In addition, the

Fig 1 Full body 'bunny suit' worn by techs while in the fab.

lighting in the clean room is either extremely bright or a yellow colour making vision even more difficult. Noise from the tools is often quite loud. Finally, materials brought into the fab must also undergo cleaning and as a result paper, unless it is laminated, is not allowed in the fab. This makes taking notes or documenting information difficult as computers are often located some distance from each tool.

For communication, clean rooms are outfitted with wired telephones near each tool set or group of similar tools located in proximity. When techs want to communicate with someone in another area or a manager outside of the fab, they must page that individual on an alphanumeric pager, give the person their contact telephone number and wait at the phone near their tool. In the meantime, they have 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. Furthermore, while they are communicating, they are most probably a few metres from their station (and in some cases also a few metres from a computer), preventing any further interaction with their tool. This not only prevents spontaneous interaction but inhibits even necessary contact.

If a problem occurs in the fab and a tech is forced to troubleshoot, it is often a problem that has occurred in the past and another tech or manager experienced with the problem or who is the resident expert on that tool might be available. If the problem is new, documenting the problem and its solution is helpful for techs in a later shift as it is bound to occur again. This accumulated knowledge must somehow be aggregated and documented without the use of paper and while continuing to process semiconductor lots at each tool and must eventually be passed down from one shift to another or from expert to novice. Although computers and terminals do exist in the clean room, finding the time to sit and document procedures or 'cheat sheets' is rare. Although files, manuals, and procedure logs are aplenty in the fab, adding knowledge to organisational memory or tips for other techs is, however, still difficult. To address this issue, techs in some fabs are now being given PDAs which allow them to take notes either on screen or using the voice recorder, transfer this data between one another using infra-red (IR) beaming, and synchronise it with data from their personal computers or from the Web. The storage capacity of these devices allows them to accumulate and disperse knowledge more quickly and easily while continuing to remain at their designated tool. This does not yet address the communication problem, however.

SimPhony leverages this existing new infrastructure, of PDAs in addition to the IEEE802.11x wireless network now being introduced in some fabs, to facilitate opportunistic interaction and communication among members of a distributed workgroup. 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.

Although trials have not been conducted in the intended environment, we describe a specific situation that occurs in the fab and theorise how and where SimPhony could replace and improve upon the present phone and pager communication system being used today. Prospective 'customers' in the fab were consulted at several stages in the design of SimPhony and offered input in the form of critiques of the usage scenarios at each stage of design.

3. The SimPhony system

Although SimPhony allows its users to send voice or text messages to individuals or user-defined chat groups, voice is the primary communication medium. In asynchronous, voice instant messaging mode, a user records a message for another user and upon completion sends it to the other. If a user logs out before listening to a message or is logged out when the message is sent, the messages will have to be accessed through the message archive. The message will most likely be 'semi-synchronous' where the delay is small and the recipient of the message listens to it immediately upon arrival (making the delay equal to the time of composition plus the time before the recipient listens to the message). Synchronous mode allows users to talk in a full-duplex mode (much like a telephone conversation). Voice instant messaging is convenient, particularly for people who are busy, because it gives the recipient notification of a message but allows them to listen to it when convenient.

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 instant messaging 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 their 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. People often find themselves sending e-mails back and forth when they are 'on-line' at the same time and do not realise that for some tasks, having a telephone conversation might be faster and more efficient. This feature hopes to bypass this type of 'mode inertia' by making communication quick and efficient.

a key contribution is automatic transitioning between voice communication modes

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, after hearing briefly what is being discussed, the interrupting session were to be more interesting, the user can choose to tune in. This preview is much like hearing a conversation as its members pass by. When the speakers are in range, the listener 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.

Because of the distributed nature of SimPhony, the various functional components, shown in Fig 2, are all separate but interdependent entities. SimPhony acts primarily as a

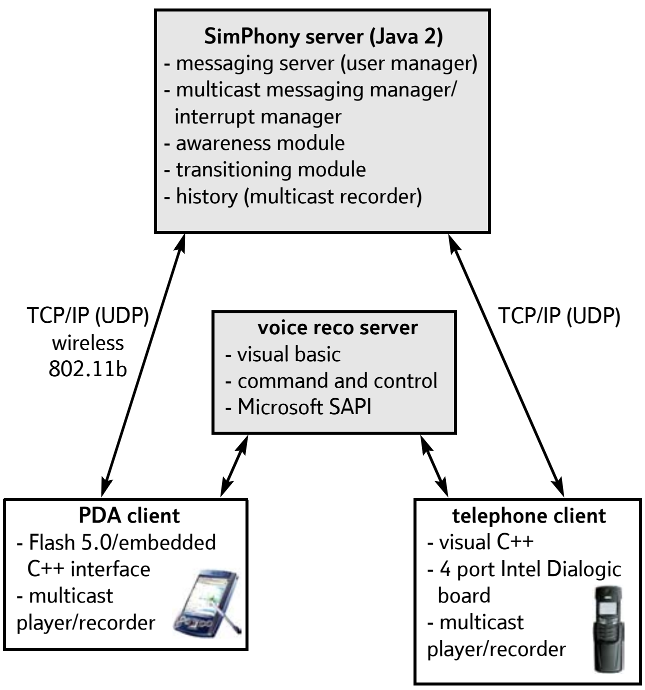

Fig 2 A diagram showing various SimPhony modules, their function, and relationship.

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 behaviour of the client. Most of the behaviour 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.

Each user has a client that typically runs on a PDA. These clients contact each other through the server, but send messages directly peer-to-peer using IP multicast. The centralised server, however, also listens to these conversations and records them. When a new user joins an existing group, he may optionally ‘catch up’ by listening to the recent previous discussion.

Based on the fab application scenario, most users are expected to run clients on their PDAs, which are also being used for other purposes. However, techs in the fab sometimes need to contact other experts who are outside. If they were on the same corporate LAN, they might also use PDAs, or could run the same software on a desktop computer. In practice, some are off site, and so an additional client was built to operate over ordinary telephone lines. An Intel Dialogic D/41JCT-LS 4-port telephony board is used as a gateway between the phone and a telephone client application, written in C++ and running on the machine containing the board. This module is based on the same underlying code as the PDA client; however, it limits its functionality to recording and playback of messages.

4. The SimPhony user interface

The primary motivation when designing SimPhony’s user interface was to create a flexible, simple-to-use, yet powerful communication system for distributed workgroups, one that would not lack any of the conveniences and transparency of the systems currently in use. Furthermore, by providing location independence, multiple modalities, multiple metaphors for communication and the ability to manage several audio channels almost simultaneously, we believed SimPhony would ultimately provide a more useful system. Building this system on the PDA platform currently in trial use in the fab allows users to have access to their data as well as their communication tool in one device, simplifying their need for more tools and streamlining data and voice into one storage intensive, mobile platform. Figure 3 shows SimPhony’s user interface on the PDA.

The system allows for distributed one-to-one and one-to-many communication in a variety of styles from voice instant messaging to full duplex audio. It allows users to create lists of ‘buddies’, similar to commercial IM clients to whom they have quick communication access and a higher level of communication awareness through the system. Figure 4 shows SimPhony being used in several modes.

Fig 3 SimPhony user interface shown on a PDA.

The system can be accessed by three interfaces, two on the PDA and one on a landline or mobile telephone, to allow users access from outside the immediate fab environment. A group can actually consist of several telephone users (as a call centre might be), and connecting to that group might simply alert the first available member of the group instead of having to access one at a time. The flexibility of providing one relatively new modality for access, the PDA, and one older modality for access, the telephone, allows for varying construction of the group hierarchy and poses several different design challenges, some of which are addressed in this system and some of which are proposed as future work.

4.1 Voice

Many of the workgroups mentioned in section 1 operate in an environment of constant challenges and interruptions. Members are often in transition between one location and the next, or engaged in a task that requires a high amount of attention or, at the very least, both hands. A voice-controlled interface (one accessible solely with voice commands) works well for such environments because it allows the user to multitask by using voice to control the system while keeping hands and eyes on their primary task. There are many considerations that need to be taken into account when designing a speech interface for such a communication system. We discuss several challenges that we encountered during the design of SimPhony and that motivate our solutions.

Using speech to control a voice communication device is problematic because the device must know when it is being given a spoken command, as opposed to just transmitting what is recorded by the microphone. The ‘push-to-talk’ button must have two functions or modes, one to transmit a system command and one to transmit a voice message.

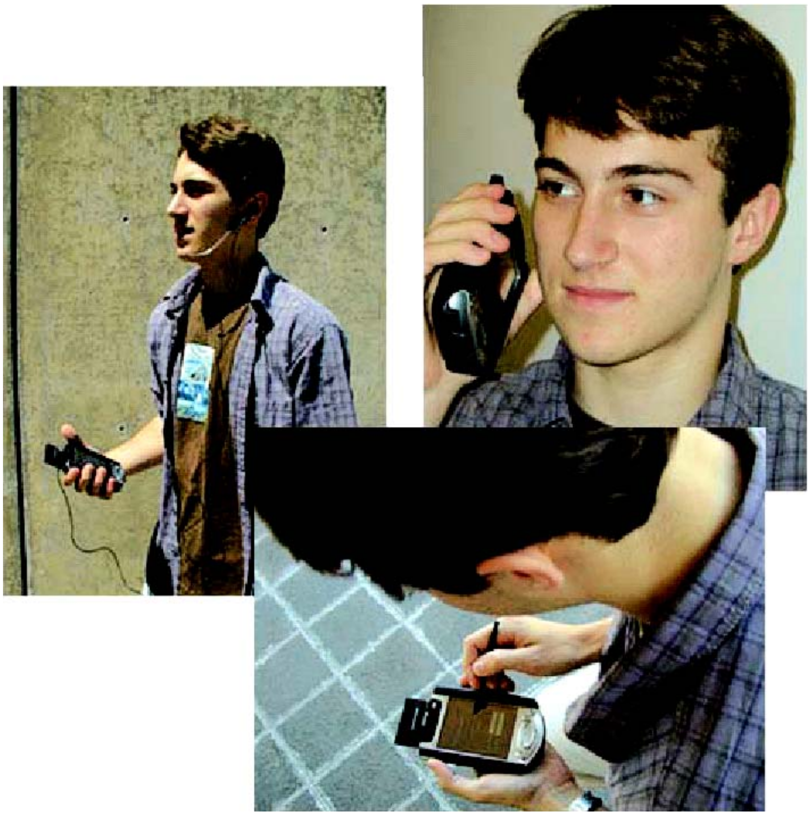

Fig 4 SimPhony used in a variety of styles from audio only, to audio and visual, to visual only.

Visual feedback (pop-up boxes or system messages which indicate the state of the system) was not a solution to distinguish the two states because it would require that the user look at the screen to understand the system's state; either audio or tactile feedback was required. Since all of the commands were short, we chose to make the command mode push (and hold) to talk. To issue commands to the system, the user must push the button on the side of the PDA and hold it as the command is given. Releasing the button signals the system to process the command and results in either audio feedback that the appropriate action has been completed or in an error sound. When sending voice or to record a message for a buddy, firstly audio feedback is received that the system is ready to record and, instead of pushing and holding the button while talking, the user simply pushes the button once and the system begins recording and/or transmitting.

This constitutes software 'push and lock' and the recording state is toggled; one additional press terminates recording. In the case of the voice message, completion of recording also sends the message to its recipient. In the case of a full-duplex connection, it simply ends the transmission and the user must press again to begin transmitting once more.

The most important functionality, that of messaging, is what is best done using the voice commands (spoken commands which allow the user to send and listen to messages on the system); a number of maintenance commands are not voice enabled in order to keep recognition accuracy high within a small vocabulary. Names of all the users were also added to the speech recognition vocabulary.

4.2 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 complementary audio cues, a tone with a rising pitch for log-in and a tone with a decreasing pitch for log-out, which indicate successful action. When another user logs in, their buddies receive an alert, either the sound of a knock at the door or a sound unique to that user so others know the identity of the new user without looking at the screen. When a group session becomes active or a voice message is received, different activity-indicating tones alert the user. A chime followed by 10 seconds of live audio from a group indicates a session becoming active and another chime indicates that a voice message was received.

4.3 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 organisational 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'; users logged into the system and users who are off-line are displayed in different colours indicating their differing states. Groups, in addition to individuals, can also be listed on the buddy list. Although the current interface does not display the names of the individuals who make up the group, future versions may easily have this accessible to users. Like individuals, currently active groups, meaning that one (or more) individual is sending a message to or chatting in that group, are also shown in white.

From the main screen shown in Fig 5, users can perform several actions or configurations. To create a group, a user selects the check box next to the names of as many buddies as wanted and names the group. Screen menus support all of the actions accessible by voice commands. Clicking on a buddy's name pops up a menu that allows the user to text, voice message or directly connect to another buddy or a group. When the user records a message or chats with a buddy/ group, the screen reflects the active state by showing a box with the name of the person or group. Beginning and ending recording is indicated with a tone, in the case of a voice message, and with a flashing LED on the PDA. In a synchronous connection, a tone is used only when the session begins and ends but transmission is indicated by the same flashing LED. 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. Alternatively, a voice command will accomplish the same task.

Fig 5 The main screen of the SimPhony user interface.

4.4 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 at a desk and phone or on the road and only accessible by mobile phone 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 without waiting to hear menus. Touch tones may be used in addition to voice commands; this is most useful in particularly noisy environments.

A user specifies their name to log in, and employs the same commands used with the PDAs to navigate. As with the PDA client, a message can be recorded for any other user using the 'voice message' command and a synchronous connection to any other on-line user or group can be made with the 'connect to' command. Users can also listen to any messages that have been recorded for them while on-line or off-line. Further development will allow groups and other users to place calls to clients normally connected over the telephone if a message is urgent. Clients will receive a telephone call which automatically connects them to the group which tried to reach them. If any user is unavailable, one of the members of the conversation can leave a voicemail message for them.

5. SimPhony in the fab

The following usage scenario is based on observations of real activities in a fab. The traditional pager and telephone communications system is replaced with SimPhony, but much of the terminology reflects current work practice utilising telephone conference bridges.

Lane is a tech who has worked at the Intel fab in New Mexico for over ten years. Over his career, he has seen his job become more focused around pinpointing and fixing tools in the clean room. During one of Lane's 12-hour shifts, a mission-critical application fails, and subsequently, this results in the failure of the automation system. Lane's first job is to find out whether this is an automation problem, a process that automates the movement of lots from one tool to the next, or a problem with the individual tool.

To simplify this process, SimPhony allows him to conference with other techs in the fab at that time, or with the other techs at that tool set. He uses the command 'connect to ToolGroup' and asks everyone at his tool set whether they are experiencing the same problem. Several people respond saying that they too are having similar problems. He discusses with them exactly what they are experiencing so he can later relay this to others. In a short time, they come to the conclusion that this is in fact an automation problem because nobody seems to be having specific problems with their tools. Without SimPhony, this might have taken them hours to figure out and communicate to one another by paging each other back and forth.

Lane then uses the command 'connect to ASC' to report their problem to a member of the automation support centre (ASC). Often the ASC personnel take reports from several techs and try to gather as much information as possible. While talking with Lane, Joan, a member of the ASC decides to conference with other techs as well. She uses the command 'connect to LithoToolSet' to speak with all of the techs at Lane's tool set. With a more unified communications system, she can speak with several techs at once and get a more detailed description of the problem.

Jane is unable to troubleshoot the problem on her own so she declares a code yellow (an alert to all of the fab to let them know that part of the manufacturing process has gone down). Normally, at this stage, Joan would need to send a text message to all automation personnel informing them that they need to dial into a phone bridge where they can discuss the problem and devise a solution. With SimPhony, she is able to 'connect to main bridge' and speak with everyone at once. All personnel in that group will receive an audio alert indicating the group has become active and if they are in another conversation, they will be interrupted with 10 seconds of audio from the main bridge group and can switch to the group since it is important. Group members might be in the fab using a PDA or at a desk and receive a telephone call connecting them to the active group. Joan can see which members are on-line and which are off-line and know how many people should be present in the group.

Joan, the crisis manager in this situation, describes the problem and its impact to the members of the group. At this point, Matt, the manager on call (MOC) connects to Stefan, the fab sweep co-ordinator (FSC) and asks him to prepare a team to do a physical sweep of the fab. Since both Matt and Stefan heard the complete description of the problem from Joan, they are able to find the most skilled people in that area to make up the team. Stefan creates a sweep team of 4-5 members and forms a group using SimPhony's screen interface. He then connects to the sweep team that he has just created to inform them of their tasks. Stefan assigns each member of the team to an area of the fab. The team gathers information at each of their areas. The members are constantly able to discuss issues among themselves by messaging the sweep team group or remaining connected in their ongoing conversation. Stefan is able to learn from these discussions and ask specific questions throughout the sweep without having to wait for the team to dial back on to the phone bridge that would normally be created for this communication task. Audio interruptions inform members who are communicating in another group about activity on that channel.

Periodically, Stefan connects to the main bridge group to update the participants of this group about the sweep. He receives interruption from the sweep team group if he is needed in that conference. Simultaneously, Matt joins a group of production managers who also want to be updated about the sweep. He too is still monitoring the main bridge group and receiving interruptions when activity occurs in the group. Finally, the sweep team is able to solve the problem and Stefan asks them to process the wafers on the tools again. If anyone on the main bridge has any further questions or requests, they can easily ask Matt or Stefan directly to relay this to the sweep team. This pass down of instructions can occur almost instantaneously using SimPhony and does not require a complicated series of paging and waiting to receive telephone calls.

communications which normally take several minutes to perform can occur quickly

At this point, the sweep team reports a few remaining problems to Stefan. Stefan is not able to solve the problem himself so he uses the command 'connect to lithoexpert', a person he has nicknamed because of his expertise in troubleshooting automation at the lithography machines, to help solve the problem. Stefan reports these remaining issues back to the main bridge. Finally, the issues are resolved, and the sweep team makes a final sweep of the fab. The sweep is successful so the groups go idle once more until the next problem arises.

The advantage of SimPhony is that communications which normally take several minutes to perform can occur quickly, with no waiting time. The process of paging someone and waiting for them to call back has been reduced to a one step process of connecting to that individual or leaving them a detailed voice message to which they can respond at their convenience. In addition, groups of people can communicate either constantly or intermittently over long periods of time without having to play 'phone tag' to get in touch. Liaisons between groups can monitor the activity of several groups and have high-level awareness of when the groups are active and how much communication is occurring. Individuals can join a group and reach several people at once instead of having to disseminate information one at a time or through phone bridges where it is never completely clear who is present. It is also easier for managers to monitor multiple working groups simultaneously

6. User feedback

SimPhony was designed to facilitate communication in the semiconductor fab plant. Initial design proposals were based on evaluation of the role of newly introduced wireless PDAs in the fab. SimPhony then went through a series of build-critique-design cycles, with critique coming from those on or familiar with the fab floor. Strict restrictions about materials brought into the clean rooms made it difficult to test the prototype on short notice. As a result, the most effective feedback came primarily from techs role-playing the user scenario described in the previous section and the designers and colleagues refining the design through simple quality assurance methods. A more thorough review will be proposed from the lessons learned from the simple user interface and technical feedback gathered for the first prototype.

The techs' primary need was a lightweight communications system by using which they could communicate with others both in the fab as well as outside, without having to go through the two-step process of paging that person and waiting for a call-back. They wanted a high-level overview of the progress of lots and what was going on in the fab, as well as a tool which allowed them to further promote group work and streamline the process of transferring information from one shift to another. Ideally that communication functionality could reside on a device that they already had and they could use it to send and receive data from one another. SimPhony focuses on the need for a lightweight communication device yet it hopes to be a flexible enough architecture so that it may someday be expanded to include data as well.

Voice instant messaging was the primary candidate for communication in the fab because of the physical restrictions of their environment and the fast paced, constantly changing nature of their work. Users preferred a lightweight system to send messages back and forth as opposed to a walkie-talkie system where everyone on the same band heard each other all the time. Although this system might have been a cheaper, simpler solution, it did not include some of the awareness and grouping features available with SimPhony. How much it might be used in a work environment remains to be seen, but techs felt that this might be a useful tool.

Secondly, the ability to communicate with more than one person at a time appealed to the techs. Their work rarely leaves them in isolation so being able to consult with other techs in the clean room and other people outside of the clean room was a high priority. They requested the further ability to add someone in to an ongoing conversation. This is a feature easily added into a second generation of SimPhony. Creating groups of people whom they could message as one and talk to together related closely to their need to keep groups aware of the overall activity and work flow inside the fab.

One of the techs commented that they had tried several tools, like instant messaging, to help improve communication between team members. However, they had to adapt their work styles to use the messaging system as opposed to the system adapting to their environment. He felt that without trying SimPhony in their natural environment, it would be difficult to tell whether the behaviours of the system did in fact adapt to their work styles; however, he said that the type of system which used communication behaviour to adapt to work style would probably be more successful than one which simply required users to adapt to it.

To get a real sense for how well SimPhony works within the fab, techs would have to use the system in a long-term trial in both their day-to-day work and perhaps during a drill or some other activity in which their communication behaviours could be controlled and compared with another tool. Such a test is difficult to do in a lab or another external environment because the communication differs so much from one day to the next. The goal of this work was to build a sufficiently satisfying initial prototype to evaluate it as a candidate for such deployment, and we expect SimPhony to undergo such a test in the future.

7. Related Work

Approaches to designing communication tools for work groups vary, but there are some common design principles that might emerge to inform a group voice messaging and communication system. Ethnographic studies of workplaces often look for communication behaviour and patterns of specific types of workgroup, and design systems to address each group's needs. Systems like Media Spaces [2] and VideoWindow [3] provide open synchronous audio and video links between two remote locations. Studies found that although both systems prompted brief social interactions, neither prompted the type of interaction that would have occurred with similar, face-to-face sightings.

Although text messaging and group chat applications in the workplace have yet to gain critical mass, they are a mechanism for opportunistic interactions which work well if users are able to overcome the assumption that those communication modalities are used specifically to 'goof off' [4]. Systems, like Hubbub [5], which support awareness, opportunistic interactions and mobility, show that the right type of notification of availability can have a lot of influence on encouraging opportunistic interactions. Hubbub uses sound icons triggered when bubs (people on the buddy list) become available or come 'on-line'. These individual icons not only give users some idea of who is available, they also allow the user to contact that person using a simple, lightweight interaction style. These sound icons and the lightweight communication mechanism work well for mobile users because instead of passively indicating the state of a co-worker, they push this information on the user, reminding them of others' presence.

Cellular radios or mobile push-to-talk communication services are also changing the metaphors for mobile voice communication. The Direct Connect service by Nextel offers such a service nationwide to over 11 million customers and allows groups of up to 25 people to communicate using a push-to-talk metaphor on their mobile phones [6]. This feature is becoming so popular that companies like Sprint are also offering similar services. Because of the many affordances provided by this medium, there are a wide variety of interaction styles and behaviours exhibited by users [7]. Unlike other messaging media (both text and audio), push-to-talk services not only provide a rich, cotemporaneous medium in which turn-taking is enforced by the technology (i.e. not a full-duplex channel), they also have low production, start-up, delay and reception costs experienced with other media.

Communication systems that use voice only in environments where users are usually behind a desk or carry a mobile device are known to be useful because they allow for a large amount of multi-tasking and can be used in an eyes or hands free mode. Systems like Thunderwire and Somewire [8] look at audio-only spaces, the types of interactions supported in these environments, and the types of interface needed to manage these spaces. Results of evaluating this space showed that a high quality audio space was a natural mode of interaction between participants. These were open, always-on systems, which would have benefited from an indication of who was present in the audio space and a simple way to move in and out of the space as the work environment demanded. Users of this system felt that they would be more comfortable with conducting interactions over the audio space if they knew who was present.

Several audio systems designed at MIT show unique ways of grounding an interaction or allowing for a variety of styles of interaction using one system. Talking in Circles [9] allows for a clever mix between a visual and audio interface. Participants in an audio chat are represented by a circle that they draw on the screen interface and move between conversations represented by groups of circles. The volume of participants that are farther away in the virtual space is reduced to provide audio parallelism of the visual environment. This unique mix of a visual and audio environment allows for peripheral spatial awareness of others in the space. Such feedback allows users to see activity in a space without having to listen to it. The TattleTrail [10] audio chat system also developed at the MIT Media Lab uses the IP-based network to store the audio chat and allows users to browse and catch up with audio conversations, using variable rate speech time compression, much like they can with instant messaging histories. Interfaces like this have shown that a GUI is not necessary to manage communication in an audio-only interface [11].

One of the first systems that tries to dynamically change an audio space in response to participant behaviour, in order to more accurately model real group conversation, is the Mad

Hatter system [12]. This system tries to take into account the fact that the conversational floor changes frequently among members of a 'gelled social group' (e.g. two participants split off to form a new conversation). By detecting different conversations and dynamically modifying the audio presented to the participants, making their current conversation louder than the conversations of others, this system hopes to reinforce the dynamic configuration of conversations in audio spaces.

Communication interfaces already in use by particular workgroups, like the Voice Loops [13] system, are interesting because they are developed internally by users on a needs-driven basis. Not only does this make the interface specific to their work environment, but in many cases, users have become trained to interact on a very expert level with these systems. NASA uses the Voice Loops system in air traffic and space mission control. Voice loops are push-to-talk audio spaces organised in a hierarchical structure, much like the teams at the control centre; there are rigid social norms on the order and style of speech turns, based in part on normal mission chain of command. Engineers and technicians usually monitor more than one channel simultaneously and are able to extract the information relevant to them.

Communication systems on the market today, like the Vocera Communication Badge [14], allow for distributed, wireless communication among two or more co-workers. Vocera is a badge with a microphone, a (push-to-talk) button, and WiFi networking. When users push the button, the microphone activates and the system uses word-spotting, going into speech input mode when it hears its name. The system then allows users to connect to a group or individual with a synchronous voice connection or leave a message if the person is unavailable. SimPhony provides for different modes of communication — voice instant message and synchronous push-to-talk. Vocera also allows for outgoing calls via a PBX interface, but apparently outside callers do not have access to the normal system voice commands and functionality. Other key differences are SimPhony's automatic transition from asynchronous to synchronous communication, a means to monitor multiple 'channels' at once, and operation as just another application on existing PDAs that might already be in use.

8. Conclusions

Although this is only the first iteration and initial reaction to SimPhony, much can be taken from this design and the initial reactions of users. Contextual design methodology prescribes that it is impossible to design for a specific user group without deeply involving yourself and educating yourself about that user group's work style and needs. Learning so much about group work and a manufacturing technician's job and the requirements not only of each individual but of the group as a whole served as great inspiration when designing features for SimPhony. That focus on the user's scenario seems to have paid off particularly in the style of the system and the various metaphors it uses for communication.

Several features discussed when designing SimPhony were not included; many are already implemented in other systems. Integrating text messaging, file transferring, and document sharing might be useful for the techs and other workgroups. Buddies can be given nicknames by the individual user under the current system, but also allowing standard nicknames for people who fulfil a specific role during all of their shifts, or are always in a certain location, might also be helpful. For example, if there is always a mechanic who works during every shift, messaging mechanic will get whomever is the mechanic working during that shift (Vocera supports this). WiFi-based location detection might localise a user to a particular tool; calling that tool might get you the tech located closest to it or perhaps the one who is listed on the time-table as being assigned to that tool. In the clean room, all techs must use two hands when carrying lots from one location to another. If their gloves were to have sensors at the fingertips, the system might be able to guess when the techs were busy transporting expensive lots and not disturb them.

SimPhony provides for different modes of communication

Visualising each conversation by showing a timeline and the activity of each group member over the course of that timeline might allow users to more knowledgably browse the audio of past conversations. Sorting previous conversations based on their participants, their length or their subject matter might further enhance the audio as valuable data for other techs or incoming workers on the next shift. Some groups might be public and allow anyone to join and participate in the discussion while others might be private and be only used to send specific types of information to participants or used only during emergencies. Some techs were wary of having all of their conversations recorded for fear that the audio might be used to monitor their work or judge their productivity. The organisation of SimPhony might be changed so that only the conversations that all members feel might be relevant to others are recorded. This would prevent techs who are having personal conversations from being penalised or make them afraid to communicate but would still allow consenting techs to record conversations for reference.

At present, SimPhony just touches the surface of what can and will be exciting in the world of voice messaging. As further iterations of the design progress, we hope the system will reach its potential in becoming more useful to distributed workgroups than the walkie-talkie or telephone.

Acknowledgements

The authors would like to thank Matt Hoffman of the Speech Interface Group for his work on the telephone interface of SimPhony. Tom Gardos, PI Natural Dialog Interfaces Project, of Intel provided the fab as a problem space and gave us a great deal of feedback and much of the materials needed for the prototype. Without him, this project would have never moved forward. Jenna Burrell of Intel Research, People and Practices provided much of the invaluable background research and study of the fab environment. Michael McGrath and Joe Butler, Automation Systems Engineers in Intel's e-

Manufacturing Team in Leixlip, Ireland gave a great deal of feedback on the design of the system and informed many of its features. Additional thanks go to Intel for funding this research, members of the Speech Interfaces Group for their feedback and support, and our extremely helpful anonymous reviewers.

References

1 Whittaker S, Frohlick D and Daly-Jones O: 'Informal workplace communication: What is it like and how might we support it?', Proceedings of Human Factors in Computing Systems, pp 131—137 (1994).

2 Mantei M, Baecker R, Sellen A, Buxton W, Milligan T and Wellman B: 'Experience in the use of Media Space', Proceedings of the Conference on Computer Human Interaction, pp 203—209 (1991).

3 Fish R, Kraut R and Chalfonte B: 'The VideoWindow system in informal communication', Proceedings of the Conference of Computer Supported Co-operative Work, pp 1—12 (1990).

4 Herbsleb J, Atkins D, Boyer D, Handel M and Finholt T: 'Introducing instant messaging and chat in the workplace', Proceedings of the Conference on Computer Human Interaction, pp 171—178 (2002).

5 Isaacs E, Walendowski A and Ranganthan D: 'Hubbub: A sound-enhanced mobile instant messenger that supports awareness and opportunistic interactions', Proceedings of the Conference on Computer Human Interaction, pp 179—186 (2002).

6 Nextel Push to Talk Services (July 2003) — http://www.nextel.com/services/directconnect/ptt_overview.shtml

7 Woodruff A and Aoki P M: 'Media affordances of a mobile push-to-talk communication service', Computing Research Repository (CoRR) Technical Report arXiv:cs.HC/0309001 (2003) — http://www2.parc.com/csl/projects/audiospaces/publications.htm

8 Hindus D, Ackerman M, Mainwaring S and Starr B: 'Thunderwire: A field study of an audio-only media space', Proceedings of the Conference of Computer Supported Co-operative Work, pp 238—247 (1996).

9 Rodenstein R and Donath J S: 'Talking in circles: Designing a spatially grounded audioconferencing environment', Proceedings of the Conference on Computer Human Interaction, pp 81—87 (2000).

10 Schmandt C, Kim C J, Lee K, Vallejo G and Ackerman M: 'Mediated voice communication via mobile IP', Proceedings of the Conference on User Interface Software Technologies, pp 141—150 (2002).

11 Singer A, Hindus D, Stifelman L and White S: 'Tangible progress: Less is more in Somewire audio spaces', Proceedings of the Conference on Computer Human Interaction, pp 104—111 (1999).

12 Aoki P M, Romaine M, Szymanski M H, Thornton J D, Wilson D and Woodruff A: 'The Mad Hatter's cocktail party: A social mobile audio space supporting multiple conversations', Proceedings of the Conference on Computer Human Interaction, pp 425—432 (2003).

13 Watts J, Woods D, Corban J, Patterson E, Kerr R and Hicks L: 'Voice loops as cooperative aids in space shuttle mission control', Proceedings of the Conference of Computer Supported Co-operative Work, pp 48—56 (1996).

14 Vocera Communications Data Sheet (2004) — http://www.vocera.com/pdf/Voc_DS_052604.pdf

---

Vidya Lakshmipathy graduated from MIT in 2004 with an MS in media arts and sciences from the Media Lab's Speech Interface group.

While at MIT, she developed SimPhony, a mobile voice messaging and group communication application. Her prior work includes TalkBack, a conversational answering machine which allows recipients of messages to effectively 'converse' with their voicemail. Her background is in Human-Computer Interaction and she received her BS in symbolic systems from Stanford University. She has also worked in design at Siemens AG in Munich and SAP Labs in Palo Alto.

Chris Schmandt directs the Lab's Speech Interfaces group. His work is related to the enhancement of personal communication and virtual social awareness, as well as to conversational user interfaces. His group uses speech technologies and portable devices to enhance human communication and to make digitised audio more useful as a data type; its focus is on developing novel applications, user interfaces, and services to exploit computer speech processing for interacting with and through computers far removed from keyboards and monitors. He received his BS in computer science and his MS in visual studies, both from MIT.