PHONE SLAVE: A GRAPHICAL TELECOMMUNICATIONS INTERFACE

Chistopher Schmandt and Barry Arons
Architecture Machine Group, MIT, Cambridge MA 02139

ABSTRACT

A personalized integrated telecommunications management system is described; it is composed of both graphical and audio interfaces to a variety of message functions. This approach utilizes a general purpose computer with a color display and speech processing board to perform as a conversational answering machine, voice or text electronic mail terminal, and on-line telephone directory with repertory dialing.

OBJECTIVES

The varieties of function in telecommunication media are increasing as rapidly as our dependence on them, both in business and home environments. The multiplicity of telecommunication channels has resulted in an increasing number of components required for access and interaction with and through these media: computer terminals and modems for electronic mail, answering machines with their own particular remote control units, autodialers with directory memory, remote telephone control of home appliances, and voice storage and forward message systems.

The *Phone Slave* is an attempt to provide unified and multi-functional interaction with a variety of these new telecommunications components [1]. A key element is the merger of voice and text into a single message management system. Perhaps more important is the desire to create coherent interfaces to this multifaceted telecommunications environment; an interface which must be responsive, intuitive, and accessible both locally and remotely over the telephone.

Increasing availability and decreasing costs of several components suggest that a flexible and adaptive system can be created largely with software. General purpose microprocessors, powerful digital signal processing chips, and low cost color raster display technology may provide a viable environment for a sophisticated and intelligent implementation, which is capable of expansion rather than replacement with rapidly growing telecommunications tasks.

AUDIO MESSAGE COMPONENTS

Although this paper will focus on graphical components of the system, it is helpful to begin by describing the voice message taking process of the answering machine mode [9].

The *Phone Slave* takes voice messages by answering the phone and engaging the caller in a conversation. It plays speech segments and records replies to gather a series of message components, for example:

machine: *Hello, Barry's telephone speaking, who's calling please?*
caller: This is Ellen Robinson
m: *What is this in reference to?*
c: Coming to see your lab at MIT.
m: *He's not available at the moment, but he left this message:* <Barry's voice>I'M HAVING LUNCH WITH CHRIS, I SHOULD BE BACK AROUND 2:30 OR 3:00. *At what number can he reach you?*
c: 555-1234
m: *When will you be there?*
c: All day today until about 5:00 P.M.
m: *Can I take a longer message?*
c: I'd like to stop by ArcMac to see the phone project.
m: *I'll be sure he gets that. Thanks for calling. Goodbye.*
c: Bye.

This process increases the likelihood of obtaining a message, by immediately interacting with the caller instead of playing a low content *"I can't answer the phone now"* message. It further encourages the caller to leave a message, by guiding the conversation through short questions which invite short responses; taking advantage of the fact that people are very likely to to answer any specific question asked of them. Callers who are greeted by a conventional tape answering machine are frequently not ready to spout off their message, and instead hang up, mentally prepare a longer monologue, and call back.

Of course, this series of questions has a focus: to gather a complete and usefully segmented incoming message. This allows the machine's owner to make a query such as *"Who left messages?"*, the reply to which consists of playing the message segments of all callers identifying themselves. Similarly one may ask *"What's her number?"* or *"When should I call?"* and receive the proper answer in the caller's voice, or in the case of an electronic mail message, in a synthesized voice.

A further motivation for asking the caller to identify herself is the desire to identify the caller in order to convey an individualized outgoing recording. In the current implementation of the system, caller identification is performed by directing the incoming audio signal to a general purpose connected speech recognizer in parallel to recording the answer to the *"Who's calling please?"* question. Identification success ranges from excellent over local exchanges to abysmal on long distance calls [5, 7].

Once a caller is identified, the conversation takes a different tack; the caller is greeted by name, played a

personal outgoing recording, and told *"If you'd like to leave a (another) message, I'll record it now. Otherwise just hang up and I'll tell him you called (again)."* The owner may call in, hear the message, and record a new outgoing recording for this caller, making a dialog possible even though the parties never speak to each other directly.

THE GRAPHICAL INTERFACE

The motivation for a graphical interface to recording, message retrieval, and call initiation functions is to provide a single consistent means of access. Although encouraged by the sudden appearance of a rash of *teleterminals* [2, 3], our work differs in several important aspects. The first is the above described audio message taking component, which imparts a new role to the terminal, and requires a specialized access method. The second is the use of color, both as a subcarrier of information and as a means of making interaction more pleasing to the eye. Finally, the integrated system is seen as a personal computer, which may be used for a variety of other, perhaps unrelated, audio or graphical tasks, rather than just a terminal with special telephony functions.

Graphic access to telecommunications functions is organized around a series of screen images, each serving to control a thematically related subset of tasks. Display is via an NTSC-compatible color frame store, with a transparent touch sensitive surface attached to the surface of the CRT [4].

A conventional NTSC display was chosen to highlight the consumer electronics potential for common displays and home computers. To accommodate this medium some constraints were placed on the graphics: text is displayed using grayscale fonts [6, 8], action occurs within television safe areas, highly saturated colors are avoided, and both horizontal and vertical lines are constrained to minimum widths greater than one pixel.

Several important design criteria were successfully incorporated in this interface. Immediate audio and/or graphical feedback is provided in response to any touch input. While performing any function, the system can always be interrupted by touching another item on the screen. Text and voice messages are accessed through the identical control functions, the distinction between message types is transparent to the user. Simple graphical symbols are used throughout for icons and display structure. The owner may dial a known caller by touching her name, eliminating the need to remember long digit strings.

Each screen provides a consistent subset of functions, i.e., dialing, directory, or message access; icons allow a change to any of the other screens. To maintain consistency, when use of one screen causes a shift in function, the appropriate screen appears for the new transaction. This new screen may, however, be primed for its task by the source screen. For example, while flipping through the Rolodex one may touch the handset icon for some directory entry (figure 1), which brings up the keypad screen simultaneous with taking the phone off-hook and generating touch tones. As the keypad screen appears, the person's name, location, and local time are displayed.

Message Access

Access to both incoming and outgoing messages is the domain of several message screens. The main message screen (figure 2) summarizes both voice and mail activity, and would ordinarily be displayed on return to one's desk. Each row represents a message, and each column a message fragment, i.e. the reply to one of the answering machine's queries. The colored bars appearing in this matrix indicate the presence and length of associated recordings. A name in the caller column indicates that the caller was identified. Multiple voice or mail messages from the same caller are grouped into contiguous sequences, as subsequent calls usually have common subject matter.

Touching a heading plays all the message segments associated with that particular row or column; this can be used, for example, to find out the subject of all the calls. Similarly, touching the bar in any message box plays that message segment; while playing, the bar changes color in sound sync and on completion becomes grey to indicate that it has been heard.

A text message may be identified by the presence of a text label in the subject field, as this can be extracted from the body of the electronic mail. When viewed, text is typeset in a grayscale font in a window on the right side of the screen, while the rest of the screen fades (figure 3). When accessed remotely by voice, this message would instead by played by the text-to-speech speech synthesizer [10].

Additional icons across the bottom of the screen provide access to other function screens. Touching the directory icon, for example, brings up the Rolodex screen with the currently active person's card on top. Other functions include pause/continue (a sound is repeated simply by touching it again), delete message, and create outgoing recordings.

Dialing Functions

Although conventional dial-by-number must be provided, the *Phone Slave* permits dial-by-name wherever possible. Outgoing calls are placed using a keypad screen (figure 6). The phone is taken off-hook by touching the handset which then raises and changes color. Numbers may be pressed digit-by-digit, as on a normal keypad, with the addition of a backspace button, as the number is not actually dialed until all digits have been entered. If a long distance call is placed, both the name and local time of the called area are displayed.

More likely, one will have entered the keypad screen by replying to a message from an identified caller or through the Rolodex directory (see below). In this case, the call is placed automatically with the recipient's name and status of any pending messages displayed.

Directory

Both calling and creation of personal recordings may be done through a directory screen (figure 4). Designed and functioning like a desktop Rolodex, cards bear names, addresses, and phone numbers. Additional text and symbols indicate whether the owner has any pending incoming or outgoing messages for that person, and the date and time she last called in. A portion of the alphabet may be selected by touching the index tab at the top of the cards. An individual person's card can be accessed by touching her name at the top, causing the cards that are obscuring it to be removed.

In addition to the standard screen transition icons for direct access to dial and message functions, several special icons appear on each person's card. Touching the telephone icon dials that person, while changing to the keypad screen. Other icons delete any pending outgoing recordings for that person or playback such recordings for review.

Outgoing Recordings

When a known caller is identified, she can receive several possible classes of recordings from the owner. A three level hierarchy is associated with these recordings; the highest level is a caller specific message. The system owner may create such a recording by touching the tape recording symbol on the appropriate directory card; a pending outgoing recording is indicated by a single reel of tape. The next level is for mailing lists or aggregations of people; any group member will receive the current recording for that list. The lowest level and most general type of recording is the currently selected default message. This can be chosen from a predefined list, or newly created at any time (figure 5).

CONCLUSIONS

The *Phone Slave* successfully demonstrates unified voice and gesture interaction with a diverse range of telecommunication tasks. The conversational answering machine is surprisingly effective at gathering coherent messages. The touch screen responsiveness and effective NTSC display resolution suggest a variety of consumer applications. The graphical interface design philosophy provides a cogent framework for future encroachment of personal computers on communications.

ACKNOWLEDGMENTS

This work has been funded by grants from Atari, Inc. and NTT, the Nippon Telephone and Telegraph Company. Speech synthesis hardware was supplied by Speech Plus. The authors also wish to thank Marc Spehlman for his dedicated work building the telephone interface hardware.

REFERENCES

1. Barry Arons. The Audio-Graphical Interface to a Personal Integrated Telecomunications System. Master Th., MIT, June, 1984.

2. D. L. Bayer and R. A. Thompson. An Experimental Teleterminal - The Software Strategy. *The Bell System Technical Journal 62*, 1 (January 1983), 121-144.

3. G. D. Bergland. Experiments in Telecommunications Technology. *IEEE Communications Magazine 20*, 6 (December 1982), 4-14.

4. *Graphical Digitizer Operating Manual.* Elographics, Inc., P.O. Box 388, Oak Ridge TN, 37830, 1980.

5. S. E. Levinson, A. E. Rosenberg, J. L. Flanagan. Evaluation of a Word Recognition System Using Syntax Analysis. *The Bell System Technical Journal 57*, 5 (May-June 1978), 1619-1626.

6. Nicholas Negroponte. Soft Fonts. Digest of Technical Papers, SID International Symposium, 1980.

7. A. E. Rosenberg and C. E. Schmidt. Automatic Recognition of Spoken Spelled Names for Obtaining Directory Listings. *The Bell System Technical Journal 58*, 8 (October 1979), 1797-1823.

8. Christopher Schmandt. Fuzzy Fonts: Analog Models Improve Digital Text Quality. Conference Proceedings, National Computer Graphics Association, 1983.

9. Christopher Schmandt and Barry Arons. A Conversational Telephone Messaging System. Digest of Technical Papers, IEEE International Conference on Consumer Electronics, 1984.

10. *PROSE 2000 Text-to-SpeechConverter User's Manual.* Telesensory Systems, Inc., 3408 Hillview Ave., P.O. Box 10099, Palo Alto CA, 94304, 1982.

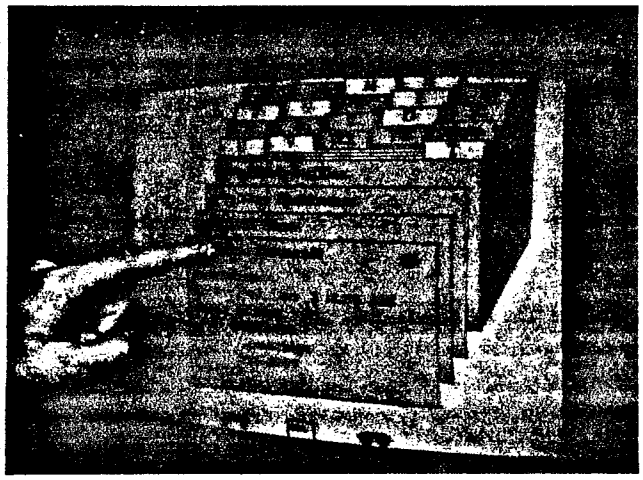

Figure 1: Close-up view of touch access to the directory.

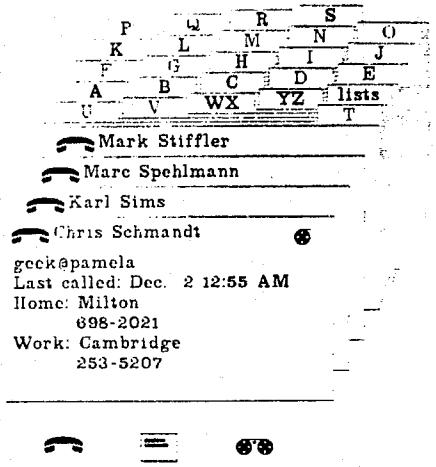

Figure 4: Directory screen with Rolodex cards.

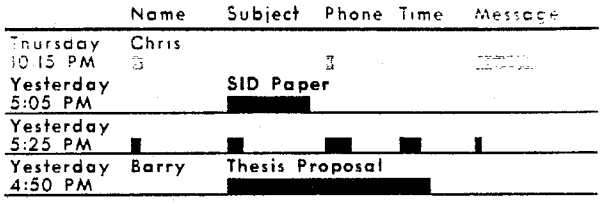

Figure 2: The main message screen.

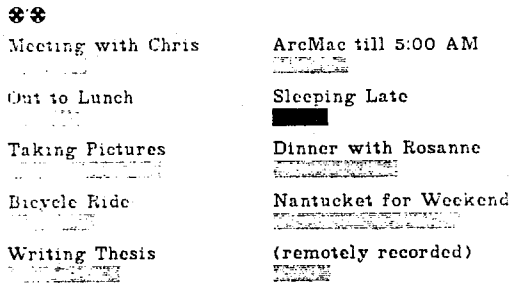

Figure 5: Default message selection screen.

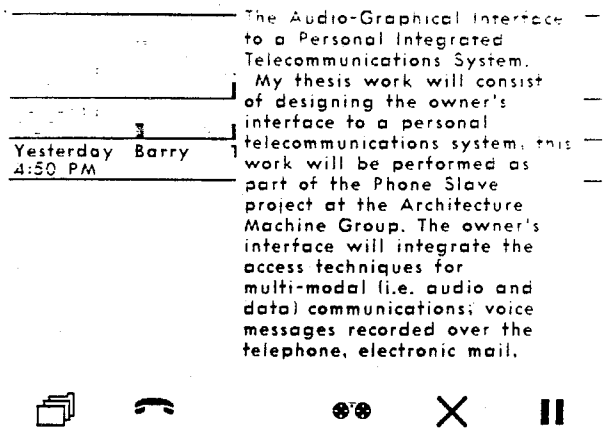

Figure 3: Message screen with text mail message.

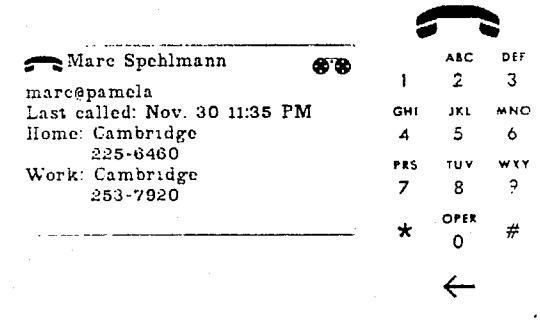

Figure 6: The keypad screen with a directory card.