

C. Schmandt and B. Arons

Architecture Machine Group, Media Laboratory, MIT
Cambridge, MA

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

I. INTRODUCTION

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. Some examples are 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* provides unified and multifunctional interaction with a variety of these new telecommunications components,¹ by merging voice and text into a single-message management system. A coherent interface to this variegated telecommunications environment 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 is now possible. General-purpose microprocessors, powerful digital-signal-processing chips, and low-cost color raster display technology provide a viable environment for a sophisticated and intelligent implementation. A system based largely of software is capable of expansion rather than replacement with rapidly growing telecommunications tasks.

The Phone Slave has been designed and developed in the MIT Media Laboratory on a super-mini (Perkin-Elmer 3230) development system. The target for commercial application is clearly a personal computer with a single signal-processing/telephone interface card.

II. 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 (a related publication emphasizes the audio interface¹⁰).

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: Hi, this is Allyson Haut.

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: 445-6441

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 Phone Slave 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 minimal content message of the form "*I can't answer the phone now.*" 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 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 simultaneous with recording the answer to the "*Who's calling please?*"

question. Identification success ranges from excellent over local exchanges to abysmal on long-distance calls.^{5,8}

Once a caller is identified, the conversation takes a different path; 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.

III. THE GRAPHICAL INTERFACE

A theme behind the graphical interface to recording, message retrieval, and call initiation functions is to provide a *consistent* and *unified* means of access to a complete telecommunications system. The Phone Slave differs from the latest generation of *teleterminals*^{2,3} 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 part of 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 a color monitor, 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,^{6,9} 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 to eliminate flicker.

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 the user touching another item on the screen. Text and voice messages are accessed through identical control functions; the distinction between message types is transparent to the user. Simple graphical symbols are used throughout for function icons and data presentation.

Each screen image 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, an 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 electronic Rolodex one may touch the handset icon for some directory entry, which brings up the keypad screen, takes the phone off-hook, and generates touch-tones. As the keypad screen appears, the person's name, location, and local time are displayed.

IV. MESSAGE ACCESS

Access to both incoming and outgoing messages is the domain of several message screens. The incoming message screen summarizes both voice and electronic 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 caller's reply to one of the answering machine's queries. Colored bars appearing in this matrix indicate the presence and length of associated recordings, color indicates the *current* message and which messages have already been heard (Fig. 1). A name in the caller column identified callers who were recognized from voice recognition or mail header fields. Multiple voice or mail messages from the same caller are grouped into contiguous sequences, as subsequent calls usually have common subject matter.

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. The telephone handset causes a switch to a dial screen with touch keypad. Other functions include pause/continue (a sound is repeated simply by touching it again) and delete message.

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 (Fig. 2). 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 gray to indicate that it has been heard.

A text message may be identified by the presence of a text label in the subject field; this is extracted from the

	Name	Subject	Time	Date	Message
Friday	Chris				
1:00 PM					
Yesterday		Japanese Video Tape			
9:15 AM					
Yesterday	Walter				
6:00 PM					
Today		Shadow Mask			
9:00 PM					
Today					
Noon					

FIG. 1. Message Screen. Incoming messages, represented by colored bars, are presented for voice or text retrieval. Callers identified by voice recognition are named. Icons across the bottom are directory screen transition, dial screen transition, delete message, and pause/continue.

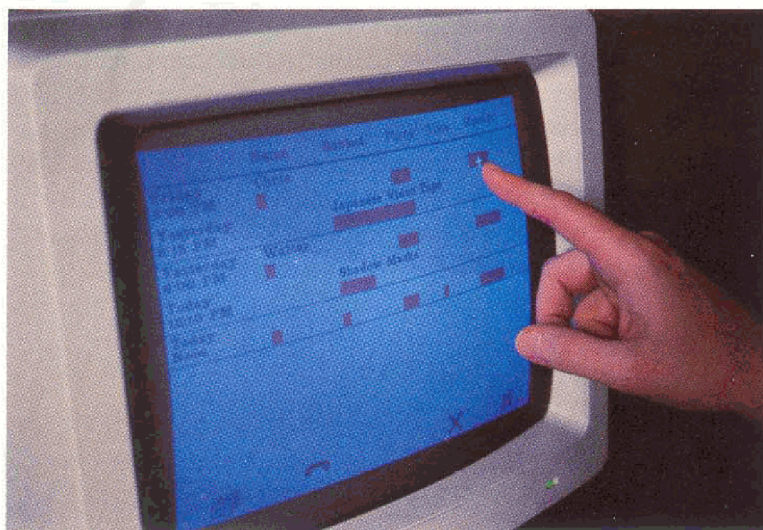


FIG. 2. Message Access. All messages are touch accessed: the column heading plays all relevant segments, the message bar plays a particular segment.

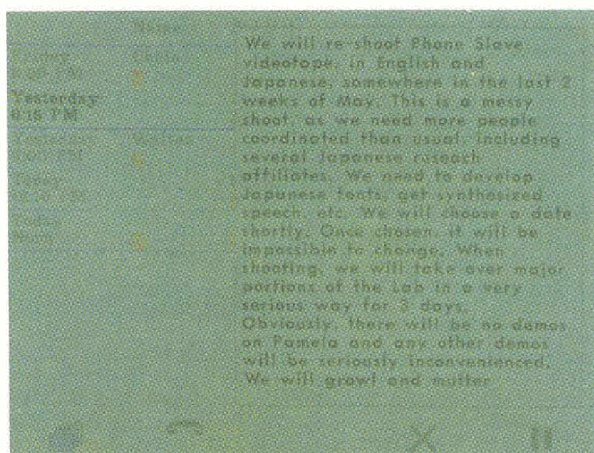


FIG. 3. Electronic Mail. A selected text message is displayed in a window; the dog-eared corner is touched to view the next page.

heading of incoming 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 (Fig. 3). Long text messages can be paged through by touching the page corners. When accessed remotely by voice, this message would instead be played by the text-to-speech speech synthesizer.⁷

V. DIRECTORY

Both calling and creation of personal recordings may be done through a directory screen (Fig. 4). This screen is based on a Rolodex metaphor; cards bear names, addresses, and phone numbers. 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.

Each card shows the person's name, the date they last called, their electronic mail address, and has fields for both home and work numbers. Of these numbers, one is highlighted in red, depending on time of day based on their area code, and can be auto-dialed by simply touching the handset icon on the top of the card. A card may also have a

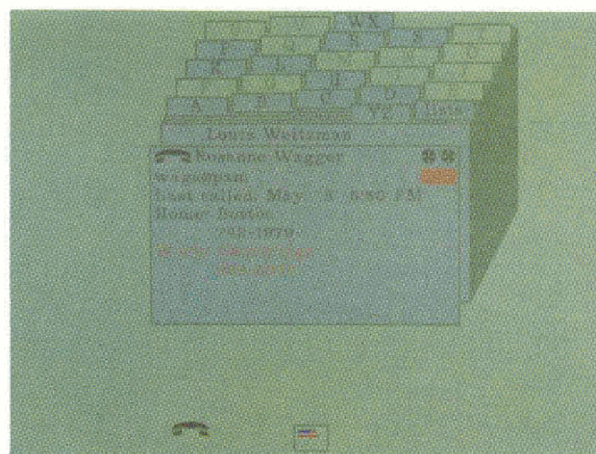


FIG. 4. Directory Screen. A Rolodex directory allows touch access to alphabetized cards, caller information and personalized messages, and the other function screens.

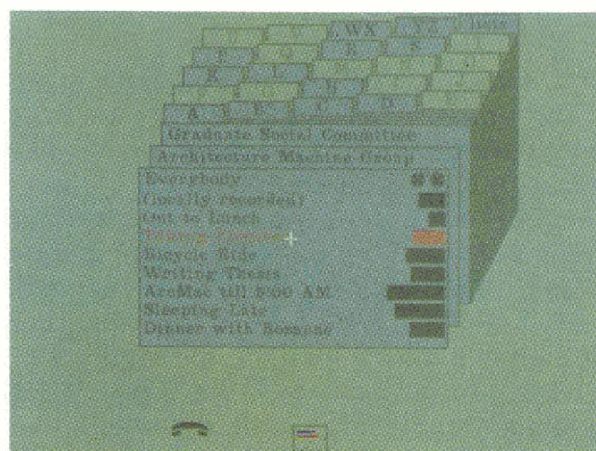


FIG. 5. Outgoing Message Card. Pre-recorded default messages may be selected or new messages recorded.

red message bar, indicating the presence of a personal outgoing message; if touched, that message will play. The tape recorder icon on the upper right of the card allows recording of a new personal outgoing message.

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 personal message for a specific caller. The next level is a mailing list or aggregation of people; any group member will receive the current recording for that list if they have no personal outgoing message pending. These messages are recorded or selected from special *group* cards that appear at the front of the Rolodex.

The lowest level and most general type of recording is the currently selected default message. This can be chosen from a predefined list by touching it on its Rolodex card (Fig. 5). The generic message may be recorded at any time, either locally with the touch screen or remotely over the phone.

VI. DIALING FUNCTIONS

The Phone Slave permits dial-by-name wherever possible, although conventional dial-by-number is also

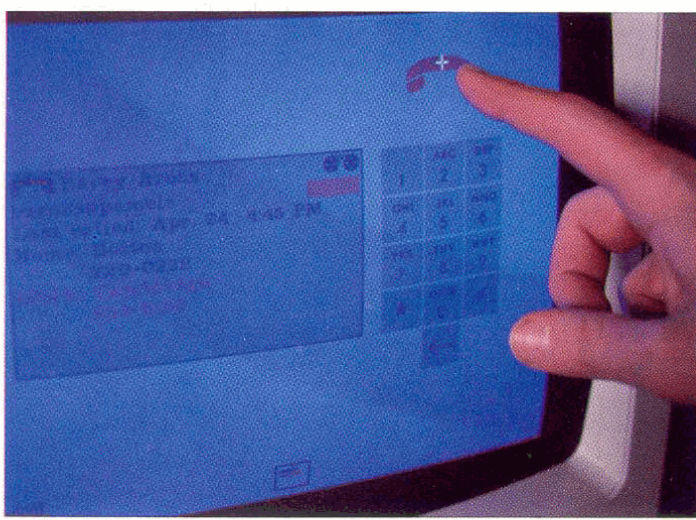


FIG. 6. Keypad Screen. The number on the card is dialed by taking the phone off-hook with a touch.

supported. Outgoing calls are placed using a keypad screen. The phone is taken off-hook by touching the handset which then raises and changes color. Numbers are pressed digit-by-digit, as on a normal keypad; a backspace button has been added for correction, and 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. The user hangs up by touching the raised handset icon (Fig. 6).

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

VII. CONCLUSIONS

The Phone Slave successfully demonstrates unified voice and gesture interaction with a diverse range of telecom-

munication 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 application of personal computers to communications for both home and business environments.

VIII. ACKNOWLEDGMENTS

This work has been supported by a grant from NTT, the Nippon Telegraph and Telephone Public Corporation. Speech synthesis hardware was supplied by Speech Plus. The authors also wish to thank Marc Spehlmann for his dedicated work building the telephone interface hardware.

REFERENCES

- ¹ B. Arons, "The Audio-Graphical Interface to a Personal Integrated Telecommunications 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, pp. 121-144 (January 1983).
- ³ G. D. Bergland, "Experiments in Telecommunications Technology," *IEEE Communications Magazine*, **20**, 6, pp. 4-14 (December 1982).
- ⁴ Graphical Digitizer Operating Manual, Elographics, Inc., P.O. Box 388, Oak Ridge, TN, 37830 (1980).
- ⁵ S. E. Levinson, A. E. Rosenberg, and J. L. Flanagan, "Evaluation of a Word Recognition System Using Syntax Analysis," *The Bell System Technical Journal*, **57**, 5, pp. 1619-1626 (May-June 1978).
- ⁶ N. Negroponte, *Soft Fonts*, Digest of Technical Papers, SID International Symposium, 1980.
- ⁷ PROSE 2000 Text-to-Speech Converter User's Manual, Speech Plus, 461 North Bernardo Ave., Mountain View, CA 94043 (1982).
- ⁸ A. E. Rosenberg and C. E. Schmidt, "Automatic Recognition of Spoken Spelled Names for Obtaining Directory Listing," *The Bell System Technical Journal*, **58**, 8, pp. 1797-1823 (October 1979).
- ⁹ C. Schmandt, *Fuzzy Fonts: Analog Models Improve Digital Text Quality*, Conference Proceedings, National Computer Graphics Association, 1983.
- ¹⁰ C. Schmandt and B. Arons, "A Conversational Telephone Messaging System," *IEEE Transactions on Consumer Electronics* CE-30, 3, pp. xxi-xxiv (August 1984).