

Not Just Another Voice Mail System

Lisa J. Stifelman
MIT Media Laboratory
20 Ames Street, E15-350
Cambridge MA, 02139

+1 617-253-8071
lisa@media-lab.mit.edu

L. Stifelman. Not Just Another Voice Mail System. *Proceedings of 1991 Conference*. American Voice I/O Society, 1991.

September 24-26, 1991; Atlanta GA. ppages 21-26.

Introduction

Telephone keypad interfaces are commonly used for interacting with voice mail systems whether one is at home or in the office. Typically, after new messages are heard by the recipient, they are transferred to an archived voice mailbox. Management of messages inside the mailbox is difficult due to the limited interaction available from a 12 button interface. Users often need to recall seemingly arbitrary assignments of functions to keys. An enhanced interface for the management of voice messages would enable users to send and receive voice mail from a personal workstation, thereby integrating speech communication into the common "desktop" interface. The voice mail system developed by the Speech Research Group at the MIT Media Lab provides such alternative methods of access and notification. The system is composed of three services: QDVM, XVMail, and PHONESHELL. QDVM is responsible for answering the telephone and for recording messages. Notification of new messages is accomplished graphically and by electronic mail. XVMail, an X Window System-based interface running on Sun SPARCstations, gives users the ability to play and record messages, as well as dial back the caller directly from the graphical workstation interface. Telephone access to voice mail is provided by PHONESHELL, a service that combines remote access to rolodex, calendar, and voice mail data.

Screen-Based Voice Mail

Alternate Method of Access

Screen-based access to voice mail has many advantages over the traditional telephone-based interface. While conventional voice mail systems require the user to dial in each time to receive new messages, XVMail's screen-based interface provides immediate access to one's messages. Each message is displayed in a window on the workstation along with the name of the caller.¹ Using a graphical interface, messages can be displayed in parallel (see Figure 1), while recordings must be presented serially over a telephone. A quick visual scan of the display is all that is necessary to locate a particular voice mail recording along with the date and time of delivery. Due to the slow and serial nature of speech, a visual interface provides a faster and less intrusive way to convey such information. In addition, the status of each message (new, viewed, deleted, or saved) can be determined at a glance and can be easily changed. For instance, the user may delete a message and then recover it using the mouse. Such operations are difficult or impossible using most telephone-based systems.

Random access is provided not only across voice messages but within each message as well. Recordings are displayed using a "Sound Viewer" widget² that is used for both playing and recording. As the selected message plays, a visual indicator advances, displaying how much of the message has been heard. Browsing through the recordings is made easier using the Sound Viewer, because the indicator can be stopped and dragged forward or back for easy and precise movement within the recording. A point in the sound can be marked and returned to over and over again. This feature is especially useful for retrieving a phone number contained in a message. Tick marks allow for the visual comparison of recording lengths between messages. This is important since length is often used as a criteria for deciding whether to listen to a particular message immediately or at some later time.

The Sound Viewer is also used for recording a new greeting. It provides a level of interaction which is not usually available for this task. The user has direct control over when to begin the recording, when to stop it, and can review any portion of the message as many times as desired. The greeting can be re-recorded until the user is satisfied with the result.

¹QDVM uses an ISDN phone system with calling party ID in order to identify the name and telephone number associated with each incoming call.

²A widget is an X Window System user interface component such as a scroll bar or button.

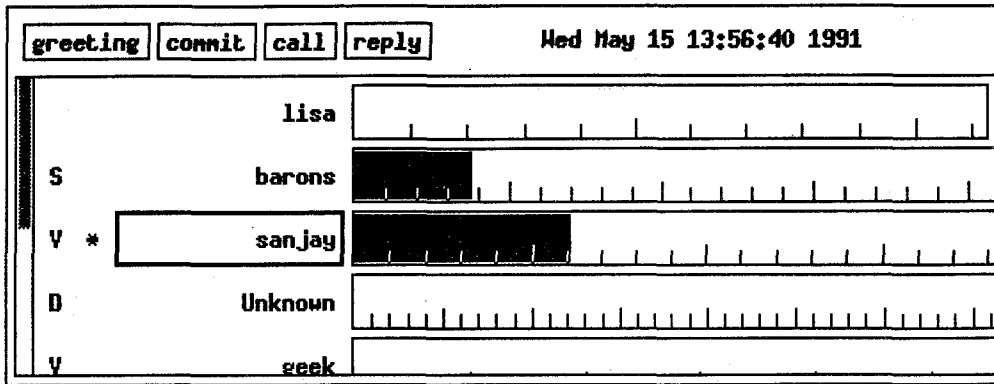


Figure 1: XVMail's screen-based voice mail display. Note that the date and time for the current message are displayed in the upper right hand corner. The black bars indicate how much of each message has been played. The letters "S", "V", and "D" indicate a saved, viewed or deleted status.

Alternate Methods of Notification

Voice mail systems traditionally inform users that new mail has arrived using a light on the telephone or a special dial tone heard only when the user picks up the handset. XVMail offers alternate means of notification via graphics and electronic mail. When a message is received, the voice mail icon highlights and the new recording appears at the bottom of the window (see Figure 2). The user can now access the message without shifting his/her attention away from the computer to the telephone. Along with iconic notification, the user also receives a short electronic mail message. This is particularly useful for users who log in from a remote location to read their electronic mail when traveling, as it saves an unnecessary "check-in" phone call if no voice messages have arrived.

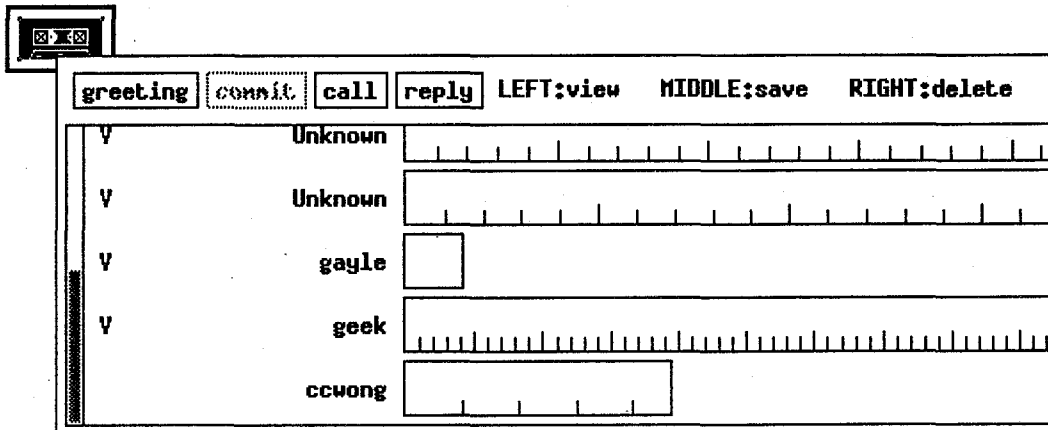


Figure 2: Iconic notification of new voice mail. Note that one can determine that the longest message is from geek due to the number of tick marks. The correspondence of mouse buttons to functions is displayed as well.

Management of Messages

Graphical access to voice mail allows for better management of one's messages. Retrieval of archived messages using a telephone-based interface is cumbersome since messages are normally ordered sequentially by time. The messages are assigned numbers, and there is no cue as to their contents, requiring the user to listen to each one in order until the desired one is found. This process is typically so awkward and time consuming that it renders saving messages almost useless. XVMail provides the ability to name and store each message, allowing voice files to be manipulated like those containing text and/or graphics. Now that messages can be stored outside of the voice mail system, the user can attend to new messages without constantly being distracted by the old ones.

Call and Reply Options

XVMAIL also has the advantage of providing easy-to-use call back and reply mechanisms. After listening to a message, the user can select "call" which causes the caller's phone number to be dialed using XPHONE, a desktop telephone dialer [Schm89]. This eliminates telephone number look up and dialing. If a voice message reply is desired, the "reply" option will cause a graphical recording window to be displayed with the recipient's name already entered (see Figure 3). Using a microphone connected to the Sun SPARCstation's audio input jack, a message can be recorded and easily reviewed using the Sound Viewer widget. This operation has the same effect as leaving a voice message over the telephone except that it is recorded directly from the workstation. The ability to deliver a voice mail reply is convenient for times when the user knows in advance that s/he does not need to speak with the person directly.

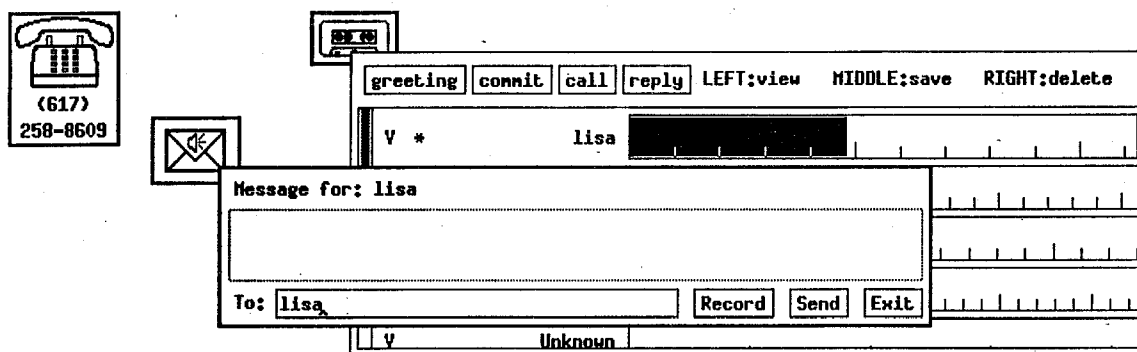


Figure 3: Selecting "call" causes XPHONE to dial the caller's number. Selecting "reply" causes a voice message recording window to be displayed.

Remote Access to Integrated Services

In addition to the graphical access provided by XVMAIL, PHONESHELL supports remote access to voice mail and three other services—DIAL-BY-NAME, ROLOTALK, and CALTALK [Schm90]. DIAL-BY-NAME provides the ability to dial a user by entering his/her name using the telephone keypad. ROLOTALK and CALTALK use text-to-speech synthesis to speak a user's rolodex and calendar information over the telephone. An extra layer of processing is performed before text is sent to a DECTalk speech synthesizer in order for data such as street and electronic mail addresses to be pronounced correctly. Synthetic speech is also used throughout the system for instructional prompts.

The combination of these services leads to many advantages over those systems which provide voice mail access in isolation. For example, an interaction between voice mail and the rolodex is needed when the user receives a message and wishes to phone the person in response. While on the same call, ROLOTALK can be selected and used to retrieve the caller's phone number and dial it. Interaction between voice mail and the calendar is convenient for a message that needs to be attended to on a particular date. For example, a message regarding a meeting planned for October 17 can be transferred into the calendar for that date, saving the user from any transcription. This is possible because audio is a data type that is supported by the graphical interface to the calendar [Schm90].

Combined Use of Screen- and Telephone-Based Voice Mail

Voice Memos

The combination of screen- and telephone-based voice mail leads to a new application—personal voice memos or audio to-do lists. When away from work, there is often the need to record a reminder to oneself of some matter to attend to upon returning to the office. Voice is a fast and easy way to record such a reminder. Users can record

memos remotely via the telephone and retrieve them using the screen-based interface when at the office. These memos will not be played when accessing voice mail remotely and therefore will not interfere with the retrieval of regular incoming messages. Personal voice memos are compelling because it is the visual interface which provides a reminder to review these recorded items. Such a feature would be useless if only telephone access were provided since the memos would become lost in the voice mailbox along with other saved messages. The word "Memo" is put in place of the caller's name so these special messages can easily be identified (see Figure 4).

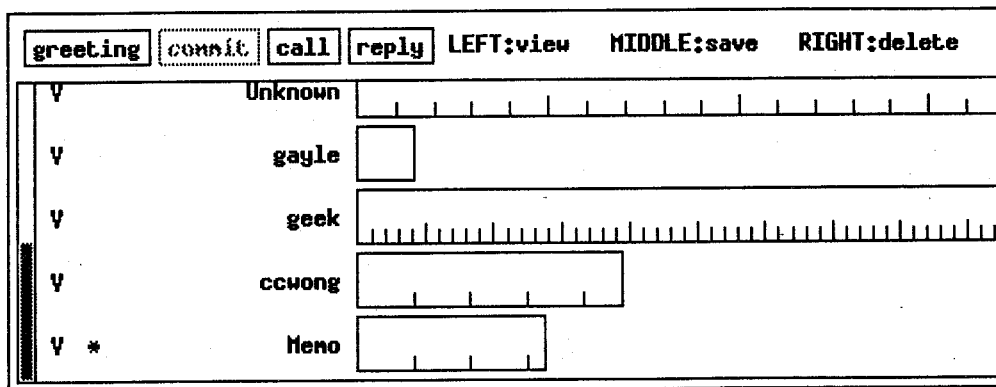


Figure 4: Voice memos recorded using the telephone interface are displayed using the screen-based interface.

Future Work

Additional features to be added to this system involve dynamic updates to information based on the user's current activity. Activity information would be especially useful for determining the appropriate greeting. Since changing one's greeting is often a cumbersome process, users rarely do it. One way to handle this problem is to allow users to record and name multiple greetings and choose between them. Phone Slave [Schm84b, Schm85a], an intelligent answering machine, employed such an approach. This solution is still limited, however, in the level of specificity that can be achieved. It is doubtful that the user would change his/her greeting for every movement within the office because this would be too time consuming. The Conversational Desktop [Schm85b] and the PX Telephone Attendant [Kame90] used calendar data to automatically update the user's greeting. This approach can be taken one step further. The Speech Research Group has developed an Activity Server [Mana91] which can determine, for example, if the user is on the telephone, alone in his/her office, or in a meeting.³ Information provided by the Activity Server can then be applied to dynamically update the user's voice mail greeting. In addition, greetings could be personalized so that different messages could be left depending on the caller, the time of day, and the user's location.

Information about the user's current location would also be useful for determining when voice mail should answer a call. If one's location is in close proximity to the phone, the user may want to allow more rings before forwarding to voice mail. If the user is out of the office, s/he may want voice mail to answer immediately. Style of notification is location dependent as well. If the user is in the office, graphical notification seems most appropriate while electronic mail would be most useful when the user is located remotely.

The telephone-based service, PHONESHELL, is being expanded as well. In addition to voice mail, remote access to electronic mail [Schm84a] will also be provided. In most cases, a terminal is needed to read electronic mail when out of the office. PHONESHELL will provide convenient remote access to one's electronic mail messages as well as the ability to send voice replies.

³The user's physical location within the building is supplied by an Active Badge system.

Conclusions

Screen-based voice mail makes many of the features described in this paper possible. These kind of preference settings and personalizations are most easily made using a direct manipulation interface and would be difficult to implement with a telephone-based system alone.

Interfaces need to be designed for the environment in which they will be used. While a telephone interface to voice mail is adequate for remote access, it is not appropriate for office use. A direct manipulation interface provides superior interaction to a 12 button keypad. The combination of screen- and telephone-based interfaces provides the optimal interaction for both local and remote access to voice mail.

Acknowledgements

This work was partially funded by Sun Microsystems. Portions of this system were developed by Barry Arons, Debby Hindus, Sanjay Manandhar, Chris Schmandt, and Chi Wong. Barry Arons and Chris Schmandt provided valuable input for the content of this paper.

References

- [Kame90] R. Kamel, K. Emami, and R. Eckert. PX: Supporting voice in workstations. *IEEE Computer*, August 1990.
- [Mana91] S. Manandhar. Activity Server: You can run but you can't hide. In *Proceedings of the Summer 1991 USENIX Conference*. Usenix, 1991.
- [Schm84a] C. Schmandt. Speech synthesis gives access to an electronic mail system. *Speech Technology*, 2(3):66-69, 1984.
- [Schm84b] C. Schmandt and B. Arons. A conversational telephone messaging system. *IEEE Transactions on Consumer Electronics*, CE-30(3):xxi-xxiv, August 1984.
- [Schm85a] C. Schmandt and B. Arons. Phone Slave: A graphical telecommunications interface. *Proceedings of the Society for Information Display*, 26(1):79-82, 1985.
- [Schm85b] C. Schmandt, B. Arons, and C. Simmons. Voice interaction in an integrated office and telecommunications environment. *Proceedings of 1986 Conference*. American Voice I/O Society, 1985.
- [Schm89] C. Schmandt and S. Casner. Phonetool: Integrating telephones and workstations. In *Proceedings of GLOBECOM '89*. IEEE Communications Society, November 1989.
- [Schm90] C. Schmandt. Caltalk: a multi-media calendar. *Proceedings of 1990 Conference*. American Voice I/O Society, 1990.