

VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker

Lisa J. Stifelman*[†], Barry Arons*, Chris Schmandt*, Eric A. Hulteen[†]

*Speech Research Group
MIT Media Lab

20 Ames Street, Cambridge, MA 02139
617-253-8026, lisa@media-lab.mit.edu

[†]Human Interface Group/ATG
Apple Computer, Inc.
20525 Mariani Ave., MS 301-3H
Cupertino, CA 95014

ABSTRACT

VoiceNotes is an application for a voice-controlled hand-held computer that allows the creation, management, and retrieval of user-authored *voice notes*—small segments of digitized speech containing thoughts, ideas, reminders, or things to do. Iterative design and user testing helped to refine the initial user interface design. VoiceNotes explores the problem of capturing and retrieving spontaneous ideas, the use of speech as data, and the use of speech input and output in the user interface for a hand-held computer without a visual display. In addition, VoiceNotes serves as a step toward new uses of voice technology and interfaces for future portable devices.

KEYWORDS

Speech interfaces, speech recognition, non-speech audio, hand-held computers, speech as data.

INTRODUCTION

How can you capture spontaneous ideas that come to you in the middle of the night, when you are walking down the street, or driving in your car? Pen and paper are often used to record this information, but it is difficult to read and write while driving, and scraps of paper with notes can become scattered or lost. A portable computer can provide better organization, but it is impossible to carry a computer, type, and look at the display while walking down the street. Some people use microcassette™ recorders since voice is a particularly fast and easy way to record such information. However, one is left with a long linear stream of audio and cannot randomly access individual thoughts. In a study of microcassette recorder users, this lack of random access was found to be the user's worst frustration [5].

This paper presents a speech interface for a hand-held¹ computer that allows users to capture and randomly access

¹The term 'hand-held' is used to refer to the size of the device. The device may actually be something worn on a belt.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

voice notes—segments of digitized speech containing thoughts and ideas. The development of the VoiceNotes application explores: (1) the problem of capturing and retrieving spontaneous ideas; (2) the use of speech as data; and (3) the use of speech input and output in the user interface for a hand-held computer.

WHY VOICE?

With advances in microelectronics, computers are rapidly shrinking in size. Laptop computers are portable versions of desktop PCs, but the user interface has remained essentially unchanged. There are also a host of small specialized electronic organizers, travel keepers, and even pocket-sized PCs that present the user with a tiny display and a bewildering array of keys. As computers decrease in size, so does the utility of traditional input and output modalities (keyboards, mice, and high resolution displays). Functionality and ease-of-use are limited on these small devices in which designers have tried to 'squeeze' more and more features into an ever decreasing product size. Rather than simply shrinking the size of traditional interface elements, new I/O modalities must be explored.²

The work presented in this paper explores the concept of a hand-held computer that has no keyboard or visual display, but uses a speech interface instead. Information is stored in an audio format, as opposed to text, and accessed by issuing spoken commands instead of typing. Feedback is also provided aurally instead of visually.

Voice technology has been explored for use in desktop computers and telephone information systems, yet the role of voice in the interface for a hand-held device has received little attention. There are two important research challenges for this work: (1) taking advantage of the utility of stored voice as a data type for a hand-held computer while overcoming its liabilities (speech is slow, serial, and difficult to manage); (2) determining the role of voice in the user interface for a hand-held computer given the limitations in current speech recognition technology.

Research and experience using voice in user interfaces has revealed its many advantages as well as its liabilities.

²Small pen-based computers are an effort in this direction, but the interface is primarily suited for visual tasks.

Voice allows the interface to be scaled down in size. In the extreme case, the physical interface may be negligible, requiring only a speaker and microphone. Speech provides a direct and natural means of input for capturing spontaneous thoughts and ideas. In comparison to writing, speech can provide faster output rates, and allows momentary thoughts to be recorded before they are forgotten [8]. In addition, voice can be more 'expressive' than text, placing less cognitive demands on the communicator and allowing more attention to be devoted to the content of the message [3]. Voice as an access mechanism is also direct (the user thinks of an action and speaks it), and allows additional tasks to be performed while the hands or eyes are busy [11].

Speech is a natural means of interaction, yet recording, retrieving, and navigating among spoken information is a challenging problem. Speech is fast for the author but slow and tedious for the listener [8]. When reviewing written information, the eye can quickly scan a page of text, using visual cues such as highlighting and spatial layout to move from one idea to the next. Navigating among spoken segments of information is more difficult, due to the slow, sequential, and transient nature of speech [1][12]. The research presented in this paper addresses the issue of how voice input can be used to record, retrieve, and navigate among segments of speech data using a hand-held device that has no visual display.

RELATED WORK

The work detailed in this section highlights key issues considered in the development of VoiceNotes including: the use of voice in hand-held environments, navigating in speech-only interfaces, and notetaking.

Degen added two buttons to a conventional tape recorder that allow users to 'mark' segments of audio while recording [5]. The audio data is then digitized and stored on a Macintosh® computer for review. In an evaluation of this prototype, users expressed the desire to customize the meanings of the marks, for more buttons to uniquely tag audio segments, and the ability to play back the marked segments directly from the device. VoiceNotes addresses these issues by using speech recognition and storage technology to allow users to create and name personal categories, and a user interface for direct entry, retrieval, and organization of audio data from a hand-held device.

Hyperspeech, a speech-only hypermedia system, addresses important design considerations for speech-only interfaces [1]. Hyperspeech provides the ability to navigate through a network of recorded speech segments using isolated word recognition. The Hyperspeech database was created and organized by the author of the system. In contrast, VoiceNotes is composed of information created and organized by the user. Additionally, voice notes are automatically segmented by the application, while the

Hyperspeech audio data was manually segmented by the author.

Notepad, a visual notetaking program, is a tool for "thought-dumping—the process of quickly jotting down a flood of fleeting ideas" ([4], p. 260). Cypher emphasizes the importance of allowing users to quickly record an idea with a minimum amount of interference. VoiceNotes, like Notepad, is intended to allow 'thought-dumping' so the interactions must be efficient—the 'tool' should not impede the user's thought process. However, the considerations for designing a voice interface are very different from those for a visual interface. While a visual interface can present information simultaneously in a multitude of windows, VoiceNotes must be more efficient in its presentation of speech data.

VOICENOTES

VoiceNotes is an application for a hand-held computer (Figure 1) that allows the creation, management, and retrieval of user-authored voice notes [15]. For example, "call mom to wish her happy birthday" can be recorded as a voice note. Voice notes can be categorized according to their content. For example, the note "call mom..." could be put into a category of notes named "phone calls."



Figure 1: Photograph of hand-held prototype.

Usage Scenarios

A demonstration of the VoiceNotes application in the context of how it might be used during the course of a user's day is provided in Figures 2 and 3.

Description of VoiceNotes

VoiceNotes provides a simple digital audio file system for organizing recorded segments of speech. The user can create lists of voice notes. Each list has a category name and a collection of associated notes. For example, the user might create a "things to do" list with associated notes such as "pay the rent," and "pick up the cleaning" (Figure 4). VoiceNotes allows the user to create multiple lists of notes.

Figure 4 shows additional examples of categories such as "calls" to make, "movies" to rent, and "groceries" to buy. These are just examples—the names of the categories are defined by the user.

The user selects the "Things to do" category:
 User: "THINGS TO DO"
 Hand-held: "Moving into *Things to do*"³
 "Pay the rent"
 "Pick up clothes from cleaners"

The user interrupts to add a note to the category:
 User interrupts: "RECORD"
 Hand-held: "Recording note"
 User: "Stop at the grocery store"
 Hand-held: "New note added"

Figure 2: Waking up in the morning, checking and updating the day's activities.

The user plays the list of categories:
 User: "CATEGORIES"
 Hand-held: "Moving into *Categories*"
 "Things to do," "Calls"

The user adds a new category called "Groceries":
 User interrupts: "RECORD"
 Hand-held: "Recording category"
 "Groceries"
 "New category added"

The user selects the "Groceries" category:
 User: "GROCERIES"
 Hand-held: "Moving into *Groceries*,
 list is empty"

The user adds notes to the "Groceries" category:
 User: "RECORD"
 Hand-held: "Recording note"
 User: "Milk"
 Hand-held: "New note added"
 User: "RECORD"
 Hand-held: "Recording note"
 User: "Orange juice"
 Hand-held: "New note added"

Figure 3: In the kitchen making breakfast, creating a grocery list.

The category name provides a method for organizing a collection of notes as well as a handle for accessing them. When the user speaks a category name⁴, it is simulta-

³Note that "moving into" is replaced by a 'list opening' sound effect if non-speech audio feedback is selected by the user. For the purposes of this demonstration, speech rather than non-speech feedback is used.

⁴A category name consists of a single short utterance.

neously recorded for playback and trained for speech recognition. Category names allow random access across lists; to select and play back a list of voice notes, the user simply speaks the category name. Since training the recognizer only requires a single utterance, the user's spoken category name becomes a voice command without a separate training process.

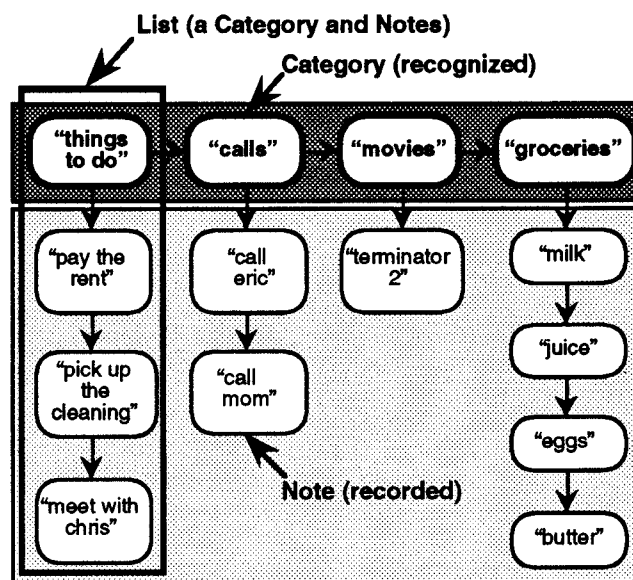


Figure 4: Sample VoiceNotes speech database.

The VoiceNotes user interface provides a simple set of voice commands for recording, navigating, and managing voice notes. Figure 5 lists some of the basic voice commands and their associated actions.

Command	Action
Play	Plays each item in a list
Record	Records an item at the end of a list
Stop	Interrupts the current activity
Next, Previous	Plays the next/previous item
Categories	Plays all of the categories
<Category name>	Selects a category and plays notes
Delete	Deletes the current item in the list
Undelete	Retrieves the last item deleted
Scan	Plays a portion of each item in a list
First, Last	Plays the first/last item in a list
Stop-listening	Turns recognition off
Pay-attention	Turns recognition on
Where-am-i	Plays the current category name
Move	Moves a note to another list

Figure 5: Basic voice commands.

In this design the "record", "next", "previous", and "delete" commands can apply to either voice notes or categories of notes depending upon the user's current position in the speech database.

In the first hand-held prototype, there were equivalent button controls for most of the voice commands. While the goal of this research has been to explore voice interfaces and applications in hand-held computers, there are cases in which button input provides a better, or more appropriate, interface.

Hand-Held Prototype

A prototype was developed to simulate the user interface experience with such a hand-held device (Figure 1). Although the prototype is tethered to a PowerBook™ computer, it allows exploration of the interface for a voice-controlled hand-held device that does not yet exist.

In the prototype, a Motorola® 68HC11 microcontroller was placed inside the shell of an Olympus® S912 microcassette recorder and interfaced to its buttons. The prototype communicates with a PowerBook through a serial connection to indicate button presses, and an analog audio connection for speech I/O. The original volume control is used for setting the speed of playback. Microphone input from the prototype is routed to the PowerBook for digitization and storage, and to a Voice Navigator™ for speech recognition.⁵

INTERFACE DESIGN ISSUES

The following key issues were considered during the initial phase of interface design, prior to user testing.

User-Definable Category Names

The most important use of speech input in the VoiceNotes application is for naming new categories and randomly accessing them. The user's ability to personalize the application by creating their own category names is essential. Since a category might contain the name of a friend, company, or an acronym, category names cannot come from a fixed recognition vocabulary. Users must be able to create these categories in real-time to support the spontaneous capture of information. Speaker dependent isolated word recognizers typically allow new words to be added to the recognition vocabulary in real-time based on acoustic data alone, whereas, some speaker independent recognition systems require a phonetic spelling of the word. Requiring users to spell or type in new words would defeat the premise underlying the use of voice (e.g., spontaneity, speed of entry). In addition, since the hand-held is a personal device, speaker independent recognition is not necessary.

Navigation

Since there is no visual display, users must be able to maintain a mental model of the VoiceNotes speech database. Voice notes are organized into a two-dimensional matrix (Figure 4), allowing the user to navigate within a particular list of notes, or between categories of notes. It was anticipated that users would have difficulty keeping track of their position in the speech database if the

organization of notes was too complex. Therefore, the database was limited to a one-level hierarchy (a category of notes cannot contain a sub-category).

While graphical hypermedia systems can show the user's navigational path visually, speech is transient and leaves no trace of its existence. Navigating between lists of notes and keeping track of one's position is simplified if there is always an active list. The current list position does not change unless the user explicitly issues a navigational command. It is important that the user feel 'in control' of the navigation, so automatic actions are avoided and commands are provided to give users complete control over their movement.

Voice and Button Input

The user interface for VoiceNotes combines multiple complementary input and output modalities. Combining voice and button input takes advantage of the different capabilities provided by each modality while allowing the limitations of one type of input to be overcome by the other. VoiceNotes can be operated using voice alone, buttons alone, or any combination of voice and button input. This flexibility is important, since the user's selection of how to interact with the application at any given time will be dependent on several factors.

The task. List selection by voice is direct, fast, and intuitive and gives the user control over the number of lists and the category names. Given a flexible number of lists, voice can provide a one-to-one correspondence between each list and the command for accessing it, while buttons cannot due to space limitations. However, buttons are better for tasks requiring fine control such as speed and volume adjustment since voice commands such as "faster, faster..." are awkward.

The context. The acoustic environment, social situation, and current user activity affect the choice of using voice or button input. Button input allows the hand-held to be operated when speech recognition accuracy is degraded due to background noise. Furthermore, button input supports the use of VoiceNotes in social contexts when it is inappropriate or awkward to speak aloud to one's hand-held computer. Alternatively, when the user's hands and eyes are busy (e.g., while driving), or vision is degraded (e.g., in darkness), voice input allows users to operate the application without requiring them to switch their visual attention in order to press a button.

Individual user preference. Some users may prefer to use buttons rather than speak to the computer while others may prefer to use speech input all the time.

⁵The Voice Navigator is a speaker dependent isolated word recognizer.

Speech and Non-Speech Audio Output

Speech and non-speech audio⁶ [7] output are the primary means of giving feedback to the user—they indicate the current state of the interface whenever the user issues a voice command or presses a button. Just as the combination of speech and button input provides the user with a richer set of interactions, the combination of speech and non-speech audio output is also powerful.

The type of feedback presented depends on the action being performed, the type of input used (voice or button), and the user's experience level and preference. Speech output is used, for example, to play back the contents of a voice note when it is deleted, while a page-flipping sound⁷ indicates movement between notes. Speech feedback is used more often in response to voice input rather than button input, since speech recognition is error prone and requires the system to provide evidence that the correct command was recognized [9]. However, too much speech output becomes laborious and slows down the interactions. For example, spoken feedback both before and after recording a note ("recording note". . . "new note added") is tedious when recording several notes in a row. Non-speech audio (i.e., a single beep before recording and a double beep after) is faster and less intrusive on the user's task.

Streamlining the Speech Interaction

In graphical interfaces screen real estate is the most limited resource, yet for speech interfaces, time is the most valuable commodity [14]. Feedback must be brief, yet informative, to conserve time and to reduce the amount of information that the user must retain in working memory [17]. Audio output must be interruptible at all times—VoiceNotes provides the ability to jump between notes on a particular list, between different lists, or to stop playback at any instant. According to Waterworth, "If he can stop the flow, obtain repeats, and move forwards and backwards in the dialogue at will, he can effectively receive information at a rate that suits his own needs and capabilities" ([17], p. 167).

In addition, it is valuable to provide users with interactive control of the rate of playback. There are a range of techniques for time-compressing speech without changing the pitch (summarized in [2]). VoiceNotes allows the speed of playback to be increased up to several times the speed of the original recording. Research suggests that a speed up of more than two times the original rate presents too little of the signal in too little time to be accurately perceived [10]. However, comprehension of time-compressed speech increases with practice and users tend to adapt quickly [16]. VoiceNotes allows users to dynamically adjust the speed of playback in order to browse a list, speeding up during some portions and slowing down when reaching a note of

interest. In addition, users can select a fixed rate of playback that they find comfortable for normal listening.

ITERATIVE DESIGN

The initial design of the VoiceNotes interface described above was developed through an iterative design process. Each aspect of the interface, especially navigation and feedback, went through many changes prior to user testing.

Moded vs. Modeless Navigation

The first VoiceNotes interface was moded—only a subset of the voice commands was valid at each point in the interaction. For example, when the last note on a list was played, the system would return to a 'top level' mode, causing users to lose their position in the speech database. The user was essentially 'dropped off the end of the list' and commands like "next" and "previous" were no longer valid.

The interface was redesigned in an attempt to create a modeless interface and to simplify navigation. In this design, all commands are always valid. When the last note is reached, if the user says "next" the system responds "end of list," and retains the user's position on the last item. Now, the user can issue commands like "next" and "previous" without fear of 'falling off the end of the list'. In this way, the beginning and end of a list act as 'anchors' for navigational control instead of drop off points.

Distinct Feedback

There were several problems with the initial design of the feedback provided by VoiceNotes. One problem was that feedback for different voice commands was not distinct, making it ambiguous as to whether or not a command was correctly recognized. For example, when selecting a category (e.g., "things to do") or saying "where-am-i", the system played the category name in response to both commands. Another problem with the feedback for selecting a list was that merely echoing the category name did not indicate any movement from one list to another.

In order to address these problems, the response to each command was made distinct and the feedback for selecting a category was changed to indicate movement ("moving into *things to do*"). Once this change was made, however, the feedback became too wordy. Therefore, an option was added to allow "moving into" to be replaced by a shorter duration sound effect (auditory icon).

USER TESTING

An informal user test (of the type described in [13]) was performed to help further refine the initial design of the VoiceNotes interface. The goal was to observe users to determine those aspects of the interface with which they had the most difficulty; particularly, how well users could navigate the speech database, given the structure shown in Figure 4. In addition, we solicited their initial reactions to the application.

⁶VoiceNotes uses mostly *auditory icons*, everyday sounds used to convey information about computer events [7].

⁷This is an example of an *auditory icon*.

Method

Six participants, three male and three female, used VoiceNotes to perform an inventory task and were subsequently interviewed. Each subject used VoiceNotes for a one hour period. The tests were video taped for later analysis. One of the participants used a microcassette recorder extensively at home and in the car for recording things to buy, videos to rent, books to read, etc. Another participant was considering buying a microcassette recorder to help keep track of personal information. None of the participants had ever used a speech recognizer before. Participants were told to 'think out loud' as they performed the different tasks [6].

First, each user trained the speech recognizer⁸ on the voice commands (Figure 5). Next, the user was briefly instructed on VoiceNotes operations. Following training, the user walked around an office⁹, performing an inventory task of several cubicles. The user created a category for the name of the person occupying the office and a note for each piece of equipment contained in the office. While taking inventory, users were interrupted occasionally and asked to create and add items to a grocery and a to-do list. The user was free to use either voice or buttons for any task.

Observations

Performance varied widely across the users tested. Some users learned the application very quickly and had few problems performing any of the tasks, while others struggled throughout the test. Several problems with the interface design were consistently observed during the testing.

Navigation. Users sometimes lost track of their position in the VoiceNotes speech database. This often occurred when selecting a category of notes, after which, the notes in the category would automatically begin to play. This automatic playback was unexpected and made the user feel out of control of the interaction. While some used the "where-am-i" command to determine their location, most wanted some kind of visual indication on the device of the current list and note.

Despite our efforts to create a 'modeless' interface, users still perceived the interface as moded (users referred to 'category' and 'notes' modes). Since the record, delete, previous, and next commands were overloaded (used for both categories and notes), users were often confused as to whether they were operating on categories or notes. When playing back the list of categories, some users stopped when they heard the category they wanted and attempted to record a note. Since they didn't first move into the list, their 'note' was actually interpreted as a new category. When asked to add a new category, users would often say "new list" instead of "record". These 'modes' also negated

the benefit of list selection by voice, since most of the users thought they had to be in 'category' mode in order to select a new category.

Interruption. A related problem was that users were unable to determine how to interrupt the speech output. During the user test, interruption by voice input was not enabled, although the ability to interrupt was available using the stop button. Users had a bias towards using voice to interrupt—when attempting to interrupt the user said "stop" rather than using the button. One user said, "I interrupt people that way [with voice], so why shouldn't I be able to interrupt this machine the same way."

Voice Input. VoiceNotes always listens for voice input, under the assumption that this allows more spontaneous use of the application. However, this makes it difficult to determine when the user is speaking a command (the system must differentiate between background noise and voice commands). Therefore, VoiceNotes remains silent unless a word is correctly recognized. During testing, if the user spoke a command and VoiceNotes did not respond, rather than repeat the command, users waited for a response, thinking that the system was still processing the input or busy performing the task. This caused the user to become confused and frustrated. Furthermore, background conversation often falsely triggered playback, making the user feel out of control because the device appeared to be operating spontaneously. Users expressed concern over the embarrassment that would be caused if this happened during a meeting or when talking to one's supervisor.

Interviews

At the end of the test, participants were interviewed about their difficulties with the user interface, their preference for voice or buttons, and their potential use of the application.

Feedback. Users perceived the interface as overly talkative or wordy—partly due to problems with interrupting the output, and due to the feedback initiated by falsely triggered recognitions. One user wanted the ability to turn the speech feedback off or select an alternative method of response.¹⁰

Voice vs. Buttons. When performing tasks, users employed both voice and button input. Users who obtained poor recognition results simply used buttons instead. Furthermore, during the test there was often background noise (e.g., a printer) that interfered with recognition, and users similarly compensated for this. When asked which input modality was preferred, some users said they would prefer voice if it was reliable enough, but all the users tested said they wanted both voice and buttons for communicating with the device.

Potential Use. All but one user said that they would like to use a hand-held device for creating personal voice notes. In

⁸Users were prompted to speak each word in the VoiceNotes vocabulary one time.

⁹The device was used under realistic ambient noise conditions.

¹⁰Non-speech feedback was not available during testing.

addition, some wanted to use the device to listen to voice mail and electronic mail messages while driving.

Implications for Redesign

The information gathered during user testing uncovered aspects of the VoiceNotes interface requiring further design development.

Navigation. One solution for addressing the user's confusion between operating on categories versus notes, is to provide separate commands for each (e.g., "new category", "new note"). Another solution is a one-to-one mapping between categories and buttons on the device. A visual indicator for each category could also help users keep track of their position.

Interruption and Voice Input. Although the ultimate goal is to allow users to pick up the device and speak a command immediately, an alternate approach must be taken due to problems with speech recognition in noise. One solution is to provide a 'push-to-talk' button. This approach also provides a consistent mechanism for interrupting the VoiceNotes speech output.

Feedback. The type of feedback (i.e., primarily speech or primarily non-speech) and amount (i.e., verbose or terse) used by VoiceNotes should be user definable. The perception of VoiceNotes as 'wordy' indicates the need to make these customization capabilities easily accessible to the user.

Voice vs. Buttons. When asked whether they would use the device if only one input modality was provided, the users consistently responded that they wanted both voice and buttons. This reinforces our original assumption about the value of offering both of these input modalities.

CONCLUSIONS

In developing VoiceNotes, many lessons were learned that are applicable to other speech and small computer interfaces:

- The use of multiple input and output modalities (in this case voice and button input, speech and non-speech output) combines the capabilities of each, while allowing limitations of a particular modality to be overcome.
- In speech interfaces like VoiceNotes, time is a valuable commodity. Feedback must be as brief and responsive as possible, audio output must be interruptible at all times, and dynamic control over the rate of playback should be provided. Furthermore, despite the best attempts to design informative, unambiguous, and brief feedback, it is important to allow users to customize both the amount and type of system feedback.
- Voice input was found to be especially valuable for categorizing and randomly accessing information (in this case, small segments of digitized speech).
- Navigation in speech-only interfaces remains a challenging design problem. Audio feedback must provide a sense of movement when navigating. Navigational 'anchors' must represent the limits of the information space, helping users to keep track of their position and maintain control over their movement.

This work has explored the use of voice, both as the data and the access mechanism in the user interface for a hand-held computer. In addition to addressing the problems of capturing and retrieving spontaneous ideas, VoiceNotes serves as a step toward new uses of voice technology and interfaces for future portable devices.

ACKNOWLEDGMENTS

We would like to thank several people for their contributions. The user testing was done in collaboration with Richard Mander, Jesse Ellenbogen, Eric Gould, and Jonathan Cohen; Andrew Kass and Nicholas Chan contributed to the software development; Derek Atkins and Lewis Knapp developed the hand-held prototypes.

This work was sponsored by Apple® Computer, Inc.*

REFERENCES

1. Arons, B. Hyperspeech: Navigating in speech-only hypermedia. In *Proceedings of Hypertext '91*, pp. 133-146. ACM, 1991.
2. Arons, B. Techniques, perception, and applications of time-compressed speech. In *Proceedings of AVIOS '92*, pp. 169-177. American Voice I/O Society, 1992.
3. Chalfonte, B.L., Fish, R.S. and Kraut, R.E. Expressive richness: A comparison of speech and text as media for revision. In *Proceedings of CHI '92*, pp. 21-26. ACM, 1991.
4. Cypher, A. The structure of users' activities. In Norman, D.A. and Draper, S.W., editors, *User Centered System Design*, chapter 12, pp. 243-263. Lawrence Erlbaum Associates, 1986.
5. Degen, L., Mander, R. and Salomon, G. Working with audio: Integrating personal tape recorders and desktop computers. In *Proceedings of CHI '92*, pp. 413-418. ACM, 1992.
6. Ericsson, K.A. and Simon, H.A. *Protocol Analysis*. The MIT Press, 1984.

* Apple, the Apple logo, and Macintosh are registered trademarks of Apple Computer, Inc. PowerBook is a trademark of Apple Computer, Inc.

7. Gaver, W.W. The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4(1):67-94, 1989.
8. Gould, J.D. An experimental study of writing, dictating, and speaking. In Requin, J., editor, *Attention & Performance VII*, pp. 299-319. Lawrence Erlbaum, 1978.
9. Hayes, P.J. and Reddy, D.R. Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies*, 19:231-284, 1983.
10. Heiman, G.W., Leo, R.J., Leighbody, G. and Bowler, K. Word intelligibility decrements and the comprehension of time-compressed speech. *Perception and Psychophysics*, 40(6):407-411, 1986.
11. Martin, G.L. The utility of speech input in user-computer interfaces. *International Journal of Man-Machine Studies*, 30:355-375, 1989.
12. Muller, M.J. and Daniel, J.E. Toward a definition of voice documents. In *Proceedings of COIS '90*, pp. 174-182. ACM, 1990.
13. Nielsen, J. Usability engineering at a discount. In Salvendy, G. and Smith, M.J., editors, *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, pp. 394-401. Elsevier, 1989.
14. Rudnicky, A.I. and Hauptmann, A.G. Models for evaluating interaction protocols in speech recognition. In *Proceedings of CHI '91*, pp. 285-291. ACM, 1991.
15. Stifelman, L.J. VoiceNotes: An application for a voice-controlled hand-held computer. Master's Thesis. Massachusetts Institute of Technology, 1992.
16. Voor, J.B. and Miller, J.M. The effect of practice upon the comprehension of time-compressed speech. *Speech Monographs*, 32:452-455, 1965.
17. Waterworth, J.A. Interaction with machines by voice: A telecommunications perspective. *Behaviour and Information Technology*, 3(2):163-177, 1984.