
NewsTalk

A speech interface to a personalized information agent

Jeffrey Alan Herman

B.B.A, The University of Texas at Austin, 1987

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

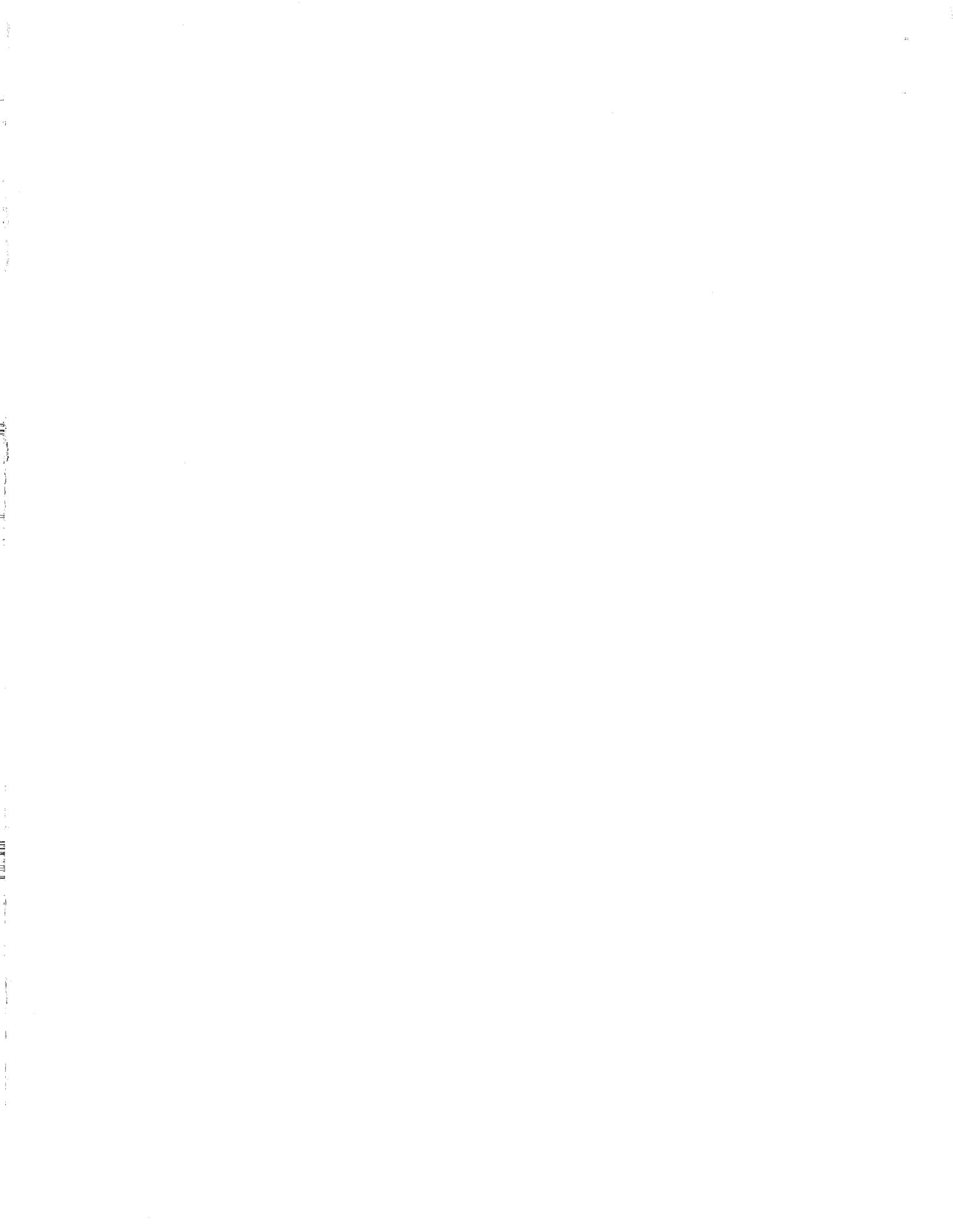
**Master of Science in Media Arts and Sciences at the
Massachusetts Institute of Technology**
June 1995

© Massachusetts Institute of Technology, 1995
All Rights Reserved

Author: Jeff Herman
Program in Media Arts and Sciences
May 12, 1995

Certified by: Ch Schmandt
Christopher M. Schmandt
Principal Research Scientist, MIT Media Laboratory
Thesis supervisor

Accepted by: _____
Stephen A. Benton
Chair, Departmental Committee on Graduate Students
Program in Media Arts and Sciences



NewsTalk:

A speech interface to a personalized information agent

Jeffrey Alan Herman

B.B.A, The University of Texas at Austin, 1987

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 12, 1995,
in partial fulfillment of the requirements for the degree of

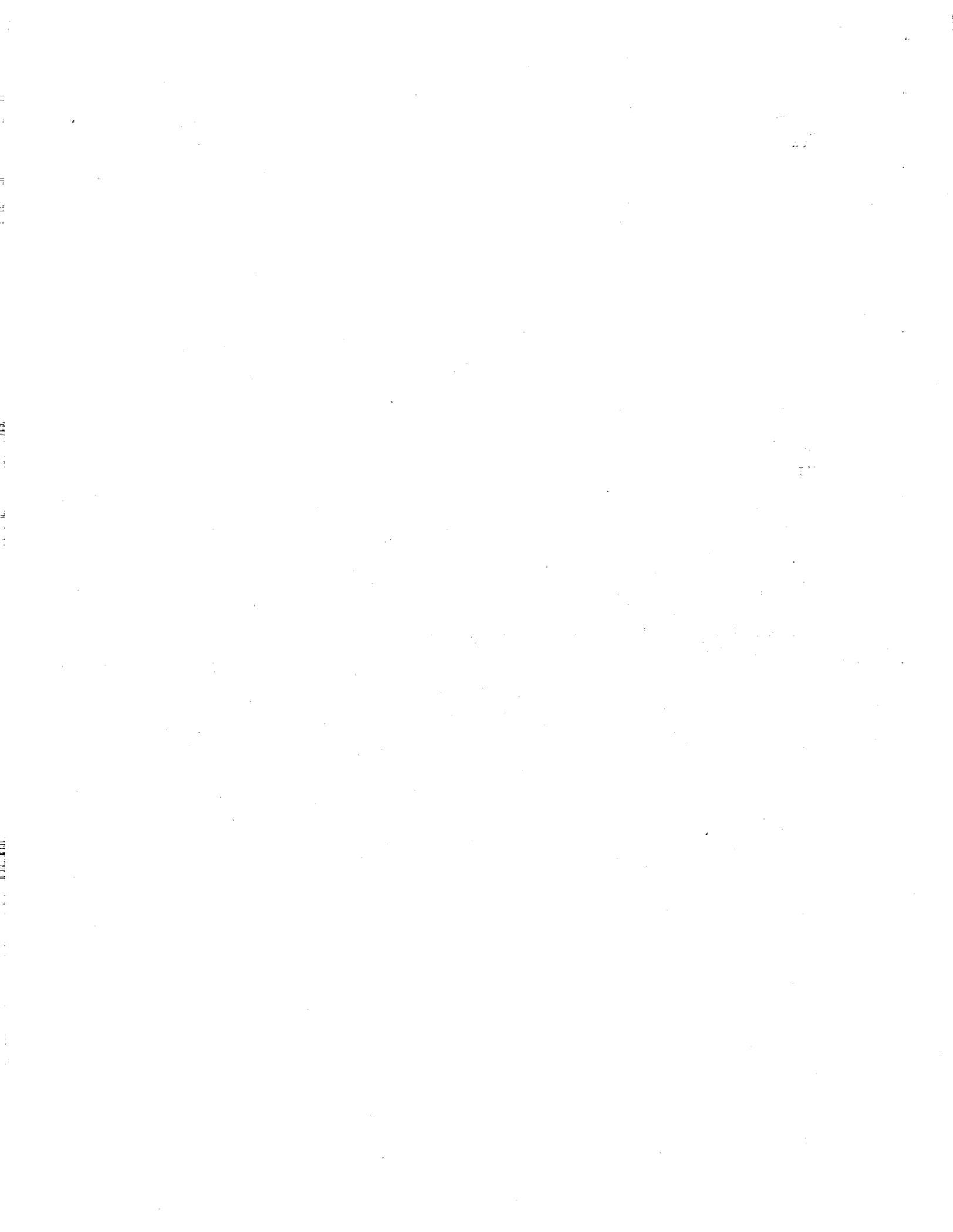
**Master of Science in Media Arts and Sciences at the
Massachusetts Institute of Technology**

Abstract: Personalized information agents have proven to be useful tools for a number of tasks. To date, these tools have required a person to use a visual display in the current situation. However, this reliance on the use of the eyes is inappropriate in common situations (e.g., while driving a car) and impossible for some people (those who are blind or visually impaired).

NewsTalk, a speech-only interface to the day's news, enables a person to access personalized information without the use of the eyes. The interface creates a spoken conversation metaphor, and makes use of speech recognition, synthesis, and digitized speech to enable communication. Qualitative and quantitative studies validate the utility of the system.

Thesis supervisor: Christopher M. Schmandt
Title: Principal Research Scientist, MIT Media Laboratory

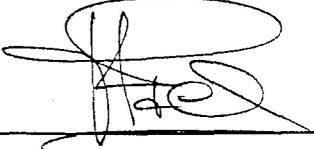
This work was sponsored by IBM and the MIT Media Lab News in the Future Consortium.



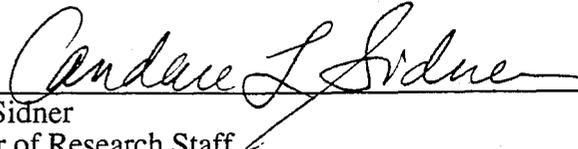
Thesis committee

Thesis supervisor: 

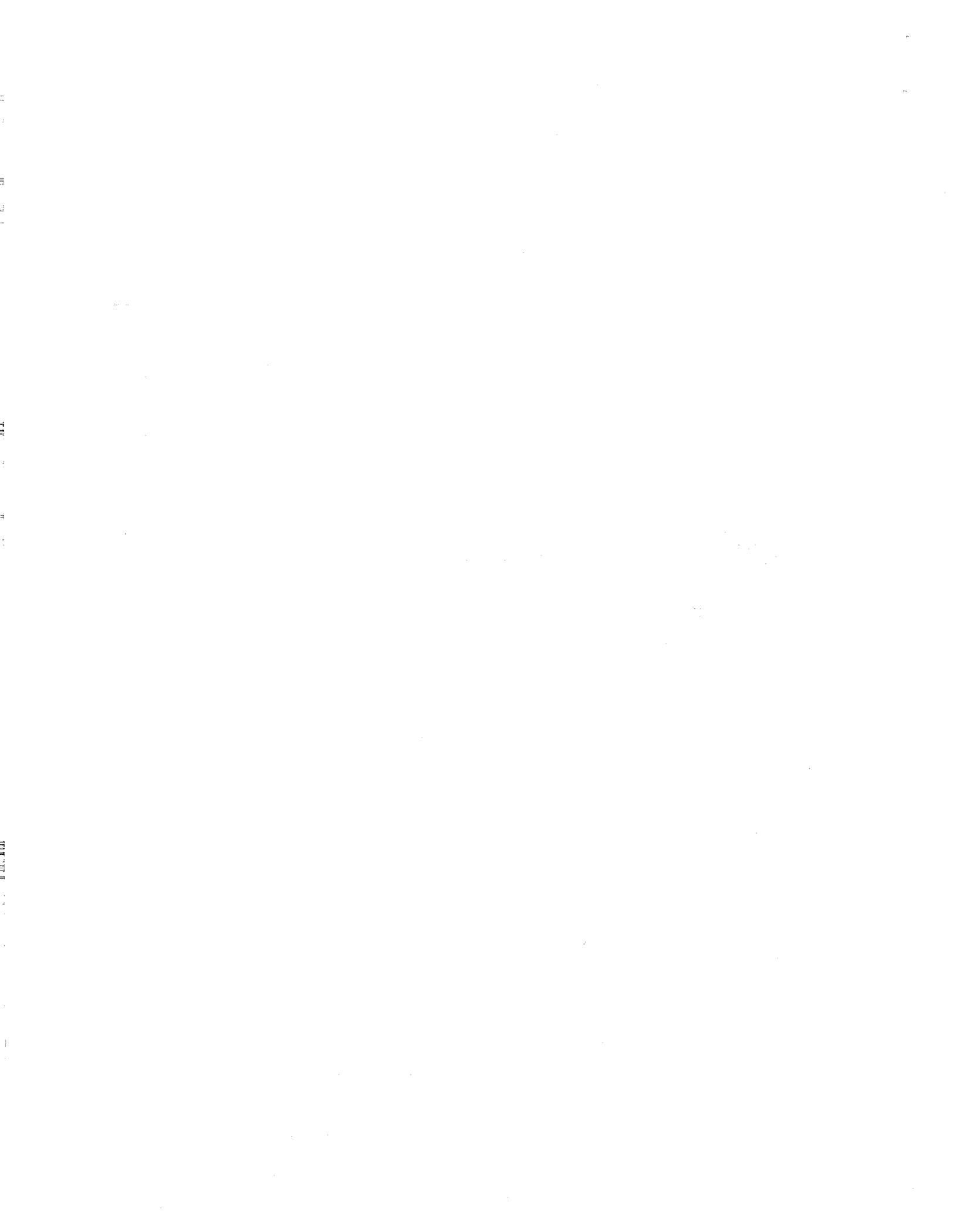
Christopher M. Schmandt
Principal Research Scientist, MIT Media Laboratory
Thesis supervisor

Reader: 

Pattie Maes
Assistant Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Reader: 

Candy Sidner
Member of Research Staff
Lotus Development, Corp.



Acknowledgments

Chris Schmandt, my advisor, contributed his considerable experience in and enthusiasm for speech user interfaces. His perspective and critiques added to what this thesis has become.

Pattie Maes and Candy Sidner, my readers, helped me to rethink how computers can assist us, and provided many useful comments and ideas.

Deb Roy, my officemate, was the first recipient of many of these ideas and always provided thoughtful feedback. In addition, he helped make the last couple of years in 353 fun. And to think it all happened because of the flip of a coin.

Matt Marx, my cohort in creating spoken conversation systems, was a constant source of information, utilities, advice, and humor.

Lisa Stifelman and Barry Arons were indispensable sources of ideas and encouragement regarding just about everything. Jordan Slott, Bruce McHenry, Sumit Basu, Henry Holtzman, and Steven Lee helped me in numerous important ways.

Steve Benton, Janet Cahn, Justine Cassell, Anil Chakravarthy, Debby Hindus, Eric Ly, Michelle McDonald, Ron MacNeil, Atty Mullins, Whitman Richards, Kris Thorisson, Chi Wong, and Yin Yin Wong provided valuable feedback on these ideas. Thanks to Gerard Salton, Chris Buckley and James Allan for providing me with SMART and answering my questions about it. The advice of Linda Peterson has been priceless.

Jack Driscoll of the Boston Globe, Bernie Gershon and Bob Donnelley of ABC News Radio introduced me to the day-to-day realities of the news business. Thanks also to the News in the Future Consortium, its director Walter Bender, and IBM for sponsoring this research.

For the past year and a half, I have spent my Wednesday evenings with Josh Bers, Sarah Elo, Earl Rennison, Deb Roy, and Nicolas St. Arnaud. We shared ideas, perspective, and camaraderie. It is inadequate to say I will miss them.

My friends beyond the lab, in particular Nicola, Barbara, Sarah, Grace, Nikki, Av, and Marge have been especially supportive and encouraging.

Finally, and most important, thanks to my mother and father, and my family.

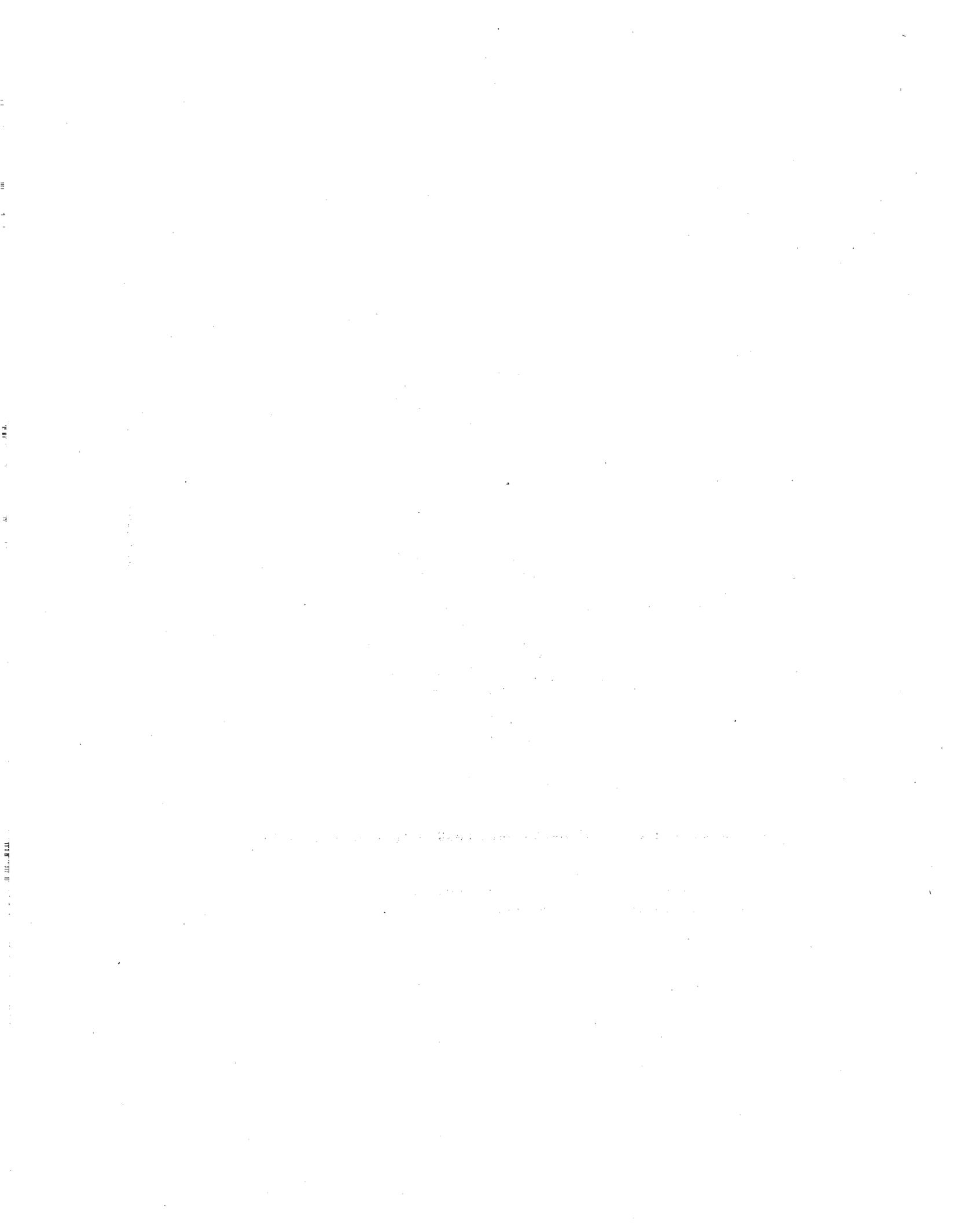


Table of Contents

1. Introduction	11
Motivations for this research	11
Contributions of this research	14
Overview of this document	14
2. Related Work in Speech User Interface Design	15
Conversational Desktop	15
Hyperspeech	15
VoiceNotes	16
Chatter	18
Voice Navigation System	18
SpeechActs	19
Wildfire	20
Comparisons and contrasts with NewsTalk	21
3. NewsTalk User Interface Design	23
Introduction	23
Transcript of a typical interaction	24
Creating mental models	28
The ups and downs of metaphors	29
Adding articles to a User Profile	31
Suggesting information	32
Confirmations and error correction	34
Dialogue states	39

4. User Studies	45
Introduction	45
Descriptions of participants in User Study 2	46
Findings	48
Final impressions from users	54
5. Personalized Information Agents	57
Introduction	57
Related work	58
A personalized information agent	61
Quantitative evaluation of results	66
6. NewsBlast (World Wide Web user interface)	69
Introduction	69
User interface design description	70
7. Conclusions and Future Directions	75
Future directions	75
Conclusions	77
References	79
Appendix 1: NewsTalk Grammar	89
Appendix 2: User Study 1	93
Appendix 3: Materials for User Study 2	99
Appendix 4: Biographies of the committee	101

1. Introduction

Motivations for this research

Personalized information agents have proven to be useful tools for a number of tasks, including filtering news [Sheth 1993]. However, today one typically needs to see and attend to a computer's visual display in order to select and receive personalized information. Using speech, NewsTalk enables more people to access personalized information in more situations.

For sighted users, NewsTalk enables access to personalized information in eyes-busy/hands-busy situations such as driving, walking, or exercising.¹

For blind users², NewsTalk provides the advantages of a newspaper (random access) and, in comparison to current options³ (e.g., listening to the radio, television, or audio cassette tape⁴), more control over what news is received.

A user communicates with NewsTalk by speaking; NewsTalk communicates with a user by synthesizing text, or by playing digitized audio. A user can request news from any one of several news sections. The interface suggests articles to a user based on what it has learned implicitly about her news interests from the current conversation and previous conversations. Because speech is temporal, a user can ask the interface to repeat its most recent utterance, to talk faster or slower, to pause for any length of time, or to end any article at any time. Because misunderstandings are an inevitable part of speech communication, the interface creates a shared context, implicitly and explicitly confirms what it

¹One user studies participant used NewsTalk while rollerblading.

² There are 11.5 million people in the United States who are blind or visually impaired [WGBH 1993].

³See [WGBH 1993] for an excellent overview of the current news options available to people who are blind or, for some other reason, have difficulty using or getting access to conventional news sources.

⁴For example, Newsweek magazine is available on audio cassette tape from American Printing House for the Blind, Louisville, Kentucky. This service provides the control of an audio cassette (e.g., start, stop, fast forward and rewind), but a user receives the tape one week after the newsstand issue is published.

has heard, and asks for clarification when it lacks confidence that it has heard a user's utterance correctly.

Creating such a speech interface is a challenging design problem because: speech-only interfaces lack visible affordances⁵ and mappings,⁶ which serve as useful cues in on-screen interfaces; the limitations of current speech recognizers [Rudnicky 1994]; and the dynamic nature of the information used (i.e., the day's news). The news on any one day will cover a vast number of topics that could be described in countless ways.

Figure 1 is an overview of the complete system described in this thesis. The Filter Agent selects and prioritizes incoming news according to the User Profile, i.e., the articles a user has listened to in previous conversations. The Filter Agent is designed so that any number of interfaces may access its results. (Currently, NewsTalk is the only interface that sends its results to the User Profile.) In addition to NewsTalk, this thesis also describes NewsBlast, a World Wide Web interface to the same filtered news.

To determine whether NewsTalk and the Filter Agent provide a useful tool, a number of qualitative and quantitative studies were conducted. Three users used the interface approximately four times a week for a period ranging from two weeks to two months. Both quantitative and qualitative results show that NewsTalk and the Filter Agent provide information of interest to particular users. (For qualitative results, see chapter 4. For quantitative results, see chapter 5.)

⁵[Norman 1988, p. 9] defines *affordances* as "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could be used."

⁶[Norman 1992, p. 25] defines *mappings* as "the relationship between the controls and their results."

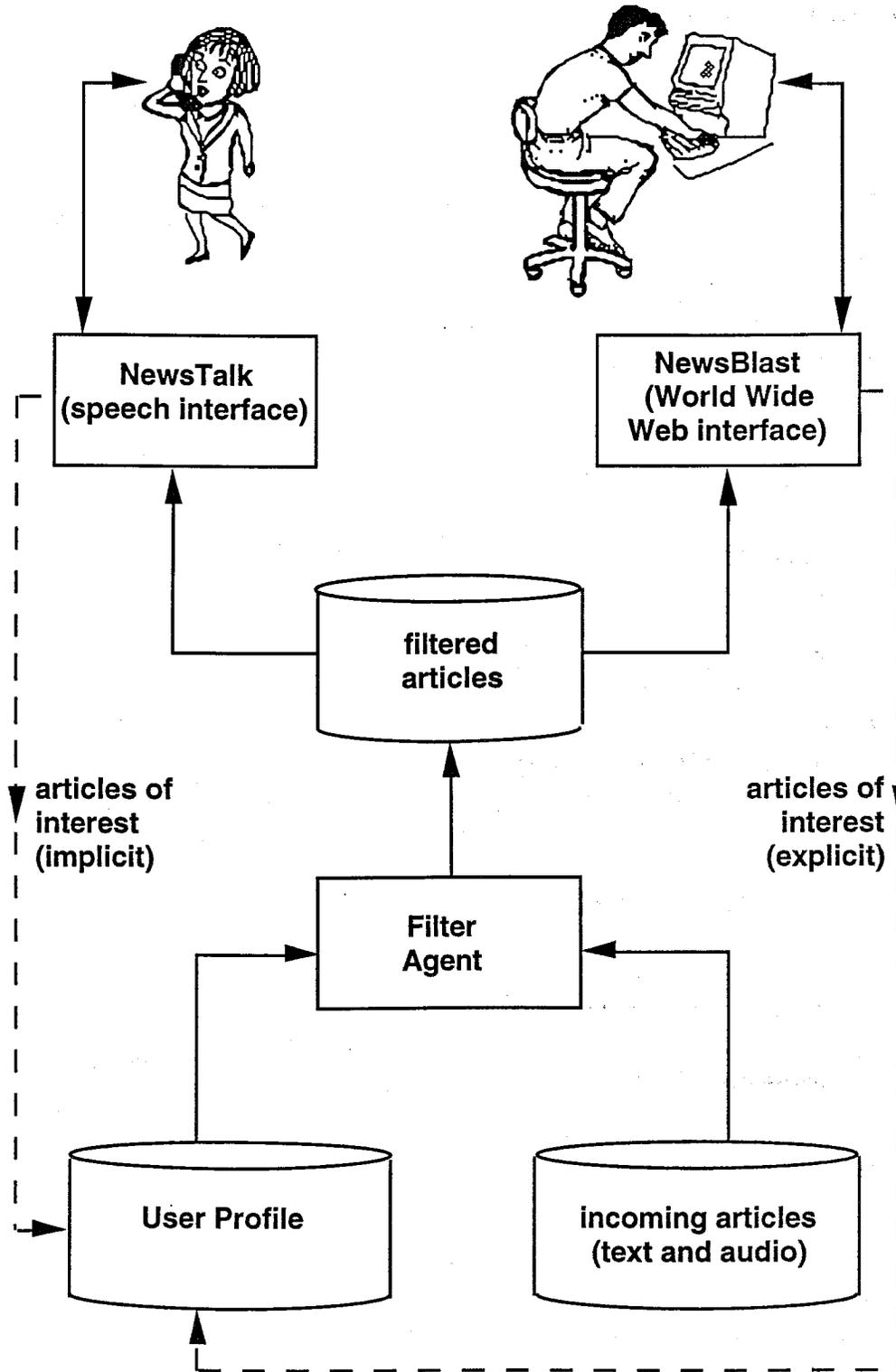


Figure 1: The components and data flow of the system.

Contributions of this research

- A speech user interface design and implementation that enables users to select and receive personalized news.
- A method where a user never explicitly judges an article, but the system adapts regardless, avoiding the “explicit judgment” problem (i.e., users too often forget to explicitly judge an article) found in previous research [Sheth 1993]. Quantitative results show the validity of this technique.
- A series of user studies that validate the interface design, provide evidence of its utility, and point out its limitations.

Overview of this document

Chapter 2, *Related Work in Speech User Interface Design*, highlights several related speech interface projects created over the last decade, including Conversational Desktop [Schmandt 1985], Chatter [Ly 1993], and SpeechActs [Yankelovich 1995a].

Chapter 3, *NewsTalk User Interface Design*, is the heart of this thesis. It uses transcripts from actual conversations between a user and the interface in order to explain the value and rationale behind the interface design. A number of specific design elements are discussed, including a spoken conversation metaphor, a physical space metaphor, implicit and explicit confirmation, and error correction.

Chapter 4, *User Studies*, explains the methodology and results of a series of user studies conducted before, during, and after designing NewsTalk.

Chapter 5, *Personalized Information Agents*, explains how the system selects and prioritizes incoming news articles according to each user’s interests. This chapter also presents previous, related work in information filtering and interface agents.

Chapter 6, *NewsBlast (World Wide Web user interface)*, discusses the on-screen interface to the same filtered information available in NewsTalk. In addition, this chapter briefly discusses the World Wide Web, and related work in online newspapers.

Chapter 7, *Conclusions and Future Directions*, points out several areas where the work described here can be continued and improved upon.

2. Related Work in Speech User Interface Design

This chapter describes work related to the NewsTalk user interface. All of the projects described use speech to retrieve information and share similar goals:

- They rely on speech input, either because it is efficient for the speaker [Gould 1978; Martin 1987], or because speech is useful in eyes-busy/hands-busy situations [Martin 1987].
- They propose methods to mitigate the drawbacks of speech output (it is slow, sequential, and transient [Muller 1990]), which makes it more tedious for the listener [Gould 1978].
- Several projects investigate how to give feedback to a user without the use of a visual display.

Conversational Desktop

Conversational Desktop [Schmandt 1985] was a demonstration system that performed a wide variety of tasks to facilitate workgroup communication, including: announcing a caller's identity; placing telephone calls; taking voice messages; recording a reminder; playing a reminder based on related events; scheduling meetings; and accessing current traffic information.

Conversational Desktop accepted continuous speech input via telephone or microphone, and played synthetic and digitized speech. In addition to the speech interface, it also included a touch-sensitive graphical interface.

Its dialogue system and parser were designed to pick out the key words in phrase fragments (e.g., if a user trying to schedule a meeting said "Barry, Friday, at 2") in order to determine the meaning of the phrase. Also, the system could engage a user in a subdialogue in order to obtain missing information or to correct a recognition error [Schmandt 1986].

Hyperspeech

Hyperspeech [Arons, 1991a; Arons 1991b; Arons 1994] is a speech-only hypermedia system that enables a user to navigate through an author-defined network of digitized speech segments (or nodes).

The speech nodes are excerpts from interviews with human-computer interaction researchers. The researchers' responses to a common list of questions were automatically recorded⁷ and manually categorized into summary and detail nodes. Hypermedia links, manually assigned by the investigator, connected logically related comments.

In the Hyperspeech system⁸, a user listens to a speech segment (or interrupts it) and follows a link by saying the appropriate command. The link types in Hyperspeech include:

- a *name* link takes a user to a node of a particular speaker. For example, if a user says "Minsky" then the system will play a related comment spoken by Marvin Minsky.
- a *control* link (caused by a user saying "browse", "scan", "more", or, "continue") enables a user to specify the level of comments desired.
- a *dialogical* link (caused by a user saying "supporting" or "opposing") enables a user to go to a node that supports or opposes the viewpoint espoused in the current node.
- a *control* link (caused by a user saying "return" or "repeat") enables a user to go to a previous node, or to hear the current node from the beginning.
- a *help* link (caused by a user saying "help" or "options") enables a user to hear a description of the current location, or a list of utterances that are legal in the current context.

VoiceNotes

VoiceNotes [Stifelman 1992; Stifelman 1993] is a speech and push-button interface that enables a user to create, organize, manage, and retrieve voice notes. A *voice note* is any user-specified utterance: it may be a thought, an idea, or a reminder. The user assigns each note to a user-defined category.

⁷A "tele-marketing style" application placed a phone call to each researcher, stated each question, and digitally recorded the researcher's response.

⁸Hyperspeech uses an isolated word, speaker-dependent speech recognizer. The system uses digitized speech to present the researchers' comments, and synthesized speech to present system prompts or questions. The preferred input device is a head-mounted, noise-canceling microphone.

For example, the note “call Lisa” might be assigned to a category named “phone calls.”

A hardware prototype was created by adding a microcontroller to the inside and buttons to the outside of a microcassette recorder. The prototype is attached to a laptop computer by a serial cable (to communicate button presses) and an analog connection (for speech I/O).

Speech is a useful medium for the task of quickly noting thoughts, ideas, and reminders because the typical person can talk faster than she can write or type [Gould 1978]. Also, a user can use such a device while her hands and eyes are busy (e.g., while driving or walking).

In addition, VoiceNotes⁹ has advantages over conventional tape recorders. A conventional tape recorder enables only recording, erasing, and linear access, but with VoiceNotes a user can:

- create new categories (by saying “record” and the name of the new category);¹⁰
- navigate quickly from one category to another (by saying the category name), or from one note to another (by saying “next”, “previous”, “first”, or “last”);
- insert a new note at any location in the current category (by saying “record”);
- delete the current category or note (by saying “delete”);
- move a note from one category to another (by saying “move” and the name of the destination category).

After issuing a command, a user receives feedback through speech (in part, recorded previously by the user) and non-speech audio (such as a page flip to indicate movement between notes).

⁹VoiceNotes uses Voice Navigator, a speaker-dependent, isolated word recognizer.

¹⁰The system records the utterance and uses it to train the speech recognizer. In the future, when the system recognizes the user saying that category, it plays the recording as implicit confirmation and takes the user to that category.

Chatter

Chatter [Ly 1993] is a speech-only interface that supports tasks such as managing e-mail messages, sending voice mail messages, and determining where a specific person is at the current time. Chatter¹¹ uses the Grosz-Sidner discourse model [Grosz 1986] to enable a user to interrupt a task, perform another, and then to pick up the interrupted task.

It also uses memory-based reasoning [Stanfill 1986] to learn what a user typically does with messages from a specific person; after communicating a message from that person, Chatter suggests that the user carry out that action. For example, if a user typically saves messages from Eric Ly then, after listening to a message from him, Chatter will ask, "Save it?"¹²

Note that the more accurately the system predicts and suggests the next action then the lower the user's cognitive load, the less the user needs to say, and the fewer opportunities there are for a recognition error.

Voice Navigation System

Another speech-only interface, Voice Navigation System (VNS) [Sparks 1994], gives users driving directions in the Denver, Colorado, area. A dialogue management system uses dialogue states and an inheritance hierarchy in order to accomplish several goals, including: enabling users to provide several utterances in succession, or to provide them in different orders; enabling meta-dialogues (such as confirming or repeating what was last said); making the implementation of the system easier and more efficient.

Sparks began the project by conducting a Wizard of Oz study.¹³ Study participants used cellular phones to interact with a "wizard" in order to navigate to specific locations. Sparks noted that these interactions were consistent in their sequential and hierarchical structure, "due largely to the structure of the task."

¹¹Like NewsTalk, Chatter uses Dagger, a speaker-independent, continuous speech recognizer from Texas Instruments, and DECTalk, a text synthesizer from Digital Equipment Corporation (DEC).

¹²Other reply types are reply, forward, and delete.

¹³For information on how to conduct a Wizard of Oz study, see [Gould 1983].

In VNS, the dialogue manager determines the current dialogue state and responds to input in an appropriate manner. It consists of two components: a sequential dialogue plan and hierarchical dialogue states.

The dialogue plan is a default path of dialogue states. This default path is altered when appropriate, e.g., when a user has already accomplished a required sub-task, or when a user requests help. Each instance of a dialogue state inherits its properties¹⁴ and behaviors¹⁵ from a class within a hierarchy.

Sparks concluded that encapsulating dialogue states' properties and behaviors in an inheritance hierarchy has "proved to be an effective way of conceptualizing and organizing the structure of task-oriented dialogues." He suggested that these same techniques "may prove useful for the design of other types of interfaces as well."

SpeechActs

SpeechActs [Yankelovich 1994a; Yankelovich 1994b; Yankelovich 1995a] includes a speech-only interface to a variety of data, including e-mail, calendar, weather, and information on publicly-traded stocks. Users access the same data on-screen (using existing applications) and over the telephone (using SpeechActs).

Early in the design process, the design team conducted a field study¹⁶ in order to observe people who called their secretary to verify or change information in their on-screen calendar. The design team observed that the grammar and vocabulary the telephone conversants used to refer to the information was "quite different" than the terms used in the screen interface. As a result, the designers decided not to translate the screen interface

¹⁴Examples of inherited properties include: 1) the task to be performed, 2) conditions that must be satisfied for the state to be complete, 3) meta-dialogues that may be invoked, 4) expectations of the user input that may occur, and 5) prompts or other information that may be provided to a user.

¹⁵Examples of inherited behaviors include the ability to: 1) send a message to a user, 2) accept input from a user, and 3) schedule primary or sub-dialogue states.

¹⁶For information on field studies, see [Holtzblatt 1993].

exactly, but instead to use the capabilities of the existing screen applications as a functional specification for the speech interface.¹⁷

Wildfire

Wildfire [Wildfire 1994] is a speech-only interface you use via the telephone to: listen to your voice mail; place telephone calls; specify criteria to screen your incoming calls and route your outgoing calls; specify criteria for reminders; and schedule meetings.

A user initiates a task, and Wildfire requests the information necessary to complete it. For example:

User: Wildfire?
Wildfire: What can I do for you?
User: Remind me to call.
Wildfire: Call whom?
User: Bill Warner.
Wildfire: When?
User: Tuesday.
Wildfire: At what time?
User: 9.
Wildfire: AM or PM?
User: AM.
Wildfire: About what?

¹⁷[Nye 1982] came to a similar conclusion. For an alternative approach, see [Mynatt 1994].

User: Brainstorm agenda for
 Thursday's strategy
 meeting.¹⁸

The Wildfire voice is recorded audio (rather than synthesized text). Wildfire is currently in beta testing and is being targeted toward business customers. Each business that buys Wildfire receives one or more Pentium-based servers (running the UNIX operating system) that include add-in cards for speech recognition, audio record and playback, and connections to the local telephone network. Wildfire uses a speaker-independent speech recognizer by default, but it enables speaker-dependent recognition for user-specific vocabulary and for users who have little success using the speaker-independent recognizer.

Comparisons and contrasts with NewsTalk

These projects compare to and contrast with NewsTalk in several important ways.

Like Chatter, SpeechActs, and VNS, NewsTalk creates a spoken conversation metaphor to communicate with a user. (The next chapter discusses the advantages and potential problems created by such a metaphor.) And, like SpeechActs, VNS, and VoiceNotes, NewsTalk benefits from extensive user studies.¹⁹

But NewsTalk is unique in that its data is organized and prioritized not by a user (as in VoiceNotes) or the system designer (as in Hyperspeech, VNS, and others), but instead by outside editors and an adaptive software component (referred to as the Filter Agent). The fact that new information is arriving daily from outside sources, that this information covers a very large number of topics, and that the information is prioritized by a filter agent without human intervention presents several challenges in speech user interface design. These issues are discussed in the next chapter.

¹⁸This speech is recorded, not recognized or transcribed.

¹⁹See Chapter 4 for information on NewsTalk user studies.

3. NewsTalk User Interface Design

“Designing human-computer experience is ... about creating imaginary worlds that have a special relationship to reality—worlds in which we can extend, amplify, and enrich our own capacities to think, feel, and act.”

—Brenda Laurel [Laurel 1991]

Introduction

A user communicates with NewsTalk by speaking to a speech recognizer.²⁰ She may use a telephone, or a microphone attached to a computer workstation. The interface communicates with a user by synthesizing text articles and digitizing audio broadcasts.²¹

The system’s goal is to provide audio news of particular interest to the current user. To accomplish this, the News Collection program collects text articles daily and audio broadcasts hourly from a number of sources. The News Assignment program assigns each item to one or more news sections: Top Stories, National, International, Local, Business, Technology, Living, or Sports. Each of these news sections is divided into two lists—the editors’ list and the user’s list (see Figure 2).²²

²⁰The system uses Dagger, a speaker-independent, limited vocabulary, continuous-speech recognizer created by Texas Instruments.

²¹NewsTalk has access to a number of audio news broadcasts. For example, ABC News sends via satellite 10-15 digitized speech segments each hour from 5am to 10pm. The segments are accompanied by a text transcript, which is used for filtering. The segments are typically sound bites from a figure in that day’s news—anyone from the President of the U.S. to an astronaut on the space shuttle.

²²For more information on collection and assignment processes, see Chapter 5.

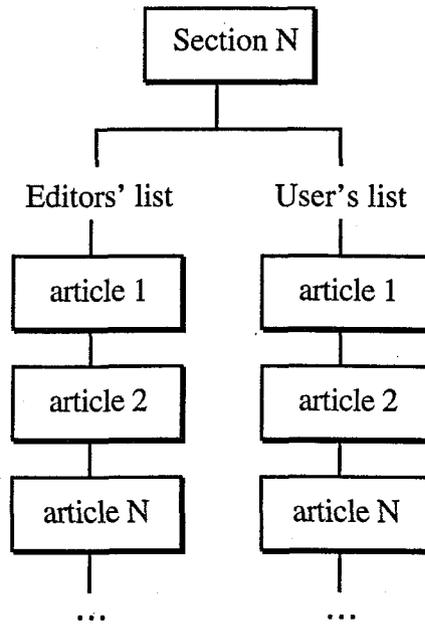


Figure 2: Software components assign and prioritize new articles in one of several sections, and the editors' list or the user's list within those sections.

- The editors' list contains the top stories in a specific section as determined by news editors. These are the articles that one would typically see on the front page of a newspaper, or hear at the beginning of a television or radio newscast. Editors' lists typically have ten to twenty articles.
- The user's list contains the top stories in a specific section as determined by the Filter Agent. The Filter Agent selects articles for this list by comparing the incoming articles with articles the user has shown a particular interest in in the past. User lists may have zero to twenty articles.

When a new user begins using NewsTalk, she has articles in the editors' lists only; the system knows nothing about her news interests (in other words, her user profile is empty) and, as a result, her lists are empty.

Transcript of a typical interaction

Below is a transcript of a dialogue between a user and NewsTalk:

Host ²³ :	Welcome to NewsTalk. What is your first and last name?	NewsTalk begins the dialogue by prompting the user for her name.
Jane:	Jane Smith.	Jane identifies herself and NewsTalk retrieves her filtered news.
Host:	Hi Jane.	NewsTalk implicitly confirms the name it has recognized.
Host:	Your top sections today are: Top Stories, Technology, and Business. What section do you want to go to first?	Based on what the Filter Agent knows of Jane's interests relative to the day's news, NewsTalk suggests the most relevant sections for her. It then prompts her for the section she would like to begin with.
Jane:	Technology.	Jane specifies a section.
Host:	In technology, do you want your list or the editors' list?	NewsTalk implicitly confirms the section ("technology"), and prompts her for additional information: because the section contains two lists, one created by news editors and the other created by her Filter Agent, her specification is incomplete.
Jane:	My list.	She requests the list of articles the Filter Agent has selected specifically for her in the section Technology.

²³NewsTalk uses one synthetic voice (the News Reader) to read articles and another synthetic voice (the Host) for all other utterances. If you are familiar with the DECTalk text-to-speech synthesizer, you may recognize the News Reader as Huge Harry and the Host as Perfect Paul.

Host:	Technology headlines in your list. Let me know which one I should read.	NewsTalk implicitly confirms the section (technology), the list (the user's list), and reminds her to respond to the headlines presented.
Host:	Microsoft rivals cheer rejection of settlement.	NewsTalk reads the headline of the most relevant headline in her list in Technology, and waits for a response.
Jane:	Read it.	She accepts the article.
News Reader:	OK, here's the article. (The News Reader reads the entire article.)	The News Reader voice implicitly confirms that it will read the article, and then reads it.
Host:	On a related note, Microsoft Prepares Windows 95.	NewsTalk notes her implied interest in the previous article and presents the article most related to the previous one.
Jane:	Skip it.	She rejects the article.
Host:	Ireland Architectural Archive Now On-line.	NewsTalk offers the next most relevant headline in her list and waits for a response.
Jane:	Go to National.	She implicitly rejects the headline and specifies a new section.
Host:	Did you say International?	The recognizer reports that it heard "International", but its confidence is below the threshold set by the interface. As a result, NewsTalk explicitly confirms what the recognizer reported. Note that it doesn't repeat the entire utterance, but just the key word(s) (in this case, "International").

Jane:	No.	Jane rejects the interface's hypothesis.
Host:	My fault. What section do you want to go to?	Given Jane's response, the interface realizes that an error has been made, accepts the blame for it ("my fault"), and prompts her for the section name.
Jane:	National.	She repeats the name of the section she wants.
Host:	National headlines in your list.	NewsTalk implicitly confirms the new section and assumes that she wants to continue getting articles in her lists.
		Note that, unlike the previous introduction to a new section, NewsTalk doesn't say "Let me know which one I should read"; it assumes that she has learned she needs to respond to headlines it suggests.
Host:	House Passes Foreign Policy Bill.	NewsTalk offers the most relevant article in her list.
Jane:	Yes.	She accepts it.
News Reader:	(The News Reader reads the article)	Note that, unlike the previous introduction to an article, NewsTalk doesn't say "OK, here's the article"; it assumes that she has learned that when she hears the News Reader' voice that the interface is reading an article.
Host:	Baseball Moves Forward with Replacement Players.	NewsTalk offers the next article in her list.
Jane:	Good-bye.	She wants to end the dialogue and the telephone call.

Host:	Did you say you're hanging up?	NewsTalk explicitly confirms that she wants to hang up.
Jane:	Yes.	She confirms.
Host:	OK Jane, good-bye.	NewsTalk implicitly confirms and hangs up. The dialogue and the telephone call end.

This transcript points out a number of design elements that NewsTalk utilizes in order to provide the most useful news to a user in the most convenient way. Among these are: a spoken conversation metaphor, a physical space metaphor, implicit and explicit confirmation, error correction, and a Filter Agent. The next chapter covers the Filter Agent; the remainder of this chapter describes the other design elements in detail.

Creating mental models

“A person forms an internal, mental model of themselves and of the things and people with whom they interact. These models provide predictive and explanatory power for understanding the interaction.”
—Don Norman [Norman 1986a]

Useful mental models aid learning—a person can use a mental model to determine what behavior is appropriate in a new situation. The user's model of a system does not have to be (and, in most cases, should not be) what the system is doing, but rather what is going on in the user interface—its context, objects, agents, and activities [Laurel 1991].

Interface designers can aid a user in creating useful mental models by creating metaphors. The next section discusses metaphors and how they are applied in the NewsTalk user interface.

The ups and downs of metaphors

“A metaphor is an invisible web of terms and associations that underlies the way we speak and think about a concept. It is this extended structure that makes metaphor such a powerful and essential part of our thinking. Metaphors function as natural models, allowing us to take our knowledge of familiar, concrete objects and experiences and use it to give structure to more abstract concepts.”
—Tom Erickson [Erickson 1990]

Metaphors are a fundamental part of our language, thoughts, and actions [Lakoff 1980]. The purpose of a metaphor in a user interface is to provide a cognitive aid that “anchors users’ understanding of the computer to something with which they are already familiar” [Mountford 1990].²⁴

A well-chosen and well-designed interface metaphor clarifies the abilities and constraints of the system, and creates coherence, i.e., ensuring that all of the elements in the interface “go together in natural ways” [Laurel 1991]. However, when a metaphor doesn’t match a user’s expectations in significant ways (what [Laurel 1991] has described as a “cognitive train-wreck”), a user is likely to have trouble applying it [Owen 1986; Erickson 1990].

A well-chosen interface metaphor enables a person to overcome what [Norman 1986b] has labeled the *gulf of execution* and the *gulf of evaluation*. The gulf of execution is when a user has difficulty translating her intentions into the language of the interface. The gulf of evaluation is when she has difficulty evaluating the interface’s response and determining if she is closer to her goal.

The NewsTalk user interface uses two main metaphors, spoken conversation and physical space, to enable users to form a useful mental model of the interface, and to overcome the gulfs of execution and evaluation.

²⁴One of the most familiar interface metaphors is the desktop [Malone 1983], used in the XEROX Star [Smith 1982; Bewley 1983], and the Apple Lisa and Macintosh [Apple 1987].

Interaction and the spoken conversation metaphor

People are well-versed at participating in conversations, and NewsTalk exploits this common skill. By modeling a human-human spoken conversation, the system creates a conversational *common ground* [Stalnaker 1978] and uses this shared context to communicate its abilities and constraints. It obeys the conversational protocol of a series of speaking turns [Sacks 1974] in the form of adjacency pairs [Schegloff 1973].²⁵ Also, it relies on the tendency of human conversants to be cooperative [Grice 1975].

Potential pitfalls of a spoken conversation metaphor

Despite the advantages of a spoken conversation metaphor, a user may still have difficulty using such a system: human-computer conversation is significantly more brittle than human-human conversation.

For example, when conversing with a person, you have a great deal of flexibility in the words and syntax you use and the topics you discuss [Reichman 1986]. In contrast, NewsTalk recognizes only utterances it finds in a pre-defined list,²⁶ and it can discuss only the news of the day. In addition, you can't talk to the interface while it is talking to you.²⁷

However, these limitations are mitigated by people's tendency to design their utterances with the abilities of their conversant in mind [Clark 1983], even when the conversant is a computer [Richards 1984; Brennan 1990]:

“The way a conversational partner represents itself and the style in which it responds influence how a user designs utterances for that partner.” [Brennan 1990]

For example, when a conversant asks a questions, the other conversant tends to respond using the same syntax and vocabulary [Levelt 1982].

²⁵For example, a question and an answer are one adjacency pair.

²⁶See Appendix 1: NewsTalk grammar.

²⁷In studies, users commonly believed that they could talk to the interface while it was talking to them. Of course, this is a technique they are familiar with from human-to-human conversation. The current system employs no echo-cancellation and, as a result, the recognizer is turned off while the synthesizer is talking. Talking to the interface during this period is fruitless.

NewsTalk exploits this tendency by using in its questions the same syntax and vocabulary that it expects to receive in a user's response:

Host: What section do you want
 to go to?

Jane: Go to Technology.

Navigation and the physical space metaphor

When users traverse information in a graphical user interface, it is not uncommon for them to become disoriented, i.e., lose track of where they are, and be unable to determine where they can go and how they can get back to information they have already seen [Conklin 1987].

This problem is potentially even more common in a speech user interface because such an interface lacks any visual cues or maps of the information. NewsTalk needs to enable a user to form an accurate mental model of how the information is organized, but without the aid of a visual representation.

In NewsTalk, users navigate from section to section and from list to list with the aid of a physical space metaphor constructed by the interface's utterances, and by the consistent use of spatial metaphors in a user's legal vocabulary and syntax.

For example, when the interface needs to know what section a user would like, it asks which section the user wants to "go to". A user may respond with the name of a section, or a longer utterance that uses the spatial metaphor (e.g., "go to technology"). Further, when a user returns to a section she has already been in, the interface implicitly confirms this with an appropriate prompt (e.g., "Back to technology").

Adding articles to a User Profile

For each user, the system creates a User Profile. The profile contains the text of articles in which a user has shown interest.

The interface determines "interest" in an article if one of two criteria is met: (1) a user listens to the entire article, or (2) she listens to three or more paragraphs of the article. These criteria are based on the fact that people tend to spend time on only those activities that provide a utility to them.

Note that users only imply interest in an article, they never explicitly judge an article. This approach has three significant advantages. First, users can focus on the task of receiving information (the reason they are using the interface) rather than on judging articles. Second, users can receive more information in the same amount of time, since they never have to take the time to give explicit feedback an article. And third, the feedback from a user is not dependent on her remembering to give feedback. (In previous research, users have often forgotten to give feedback and, as a result, the system has less information to learn from. See [Sheth 1993].)

Of course, there will be instances when a user listens to an article that does not match her interests. However, the affect of any one article on the Filter Agent is not significant since her User Profile contains all of the articles she has listened to. It is only when a number of User Profile articles refer to a specific topic that the Filter Agent will select new articles that refer to that topic.

When the interface determines that a user is interested in the previous article, it attempts to continue the theme of that article by suggesting the article most related to it. This topic is taken up in the next section.

Suggesting information

In graphical user interfaces, the most limited resource is the physical space used to display the interface; in speech user interfaces, the most limited resource is time [Rudnicky 1991]. To make the most efficient use of time, NewsTalk uses three strategies to suggest information of interest to the current user. All three strategies are based on what the system has learned about a user, either in previous conversations or in the current conversation.

The overview

At the beginning of each dialogue, after a user has identified herself, the interface suggests the news sections which are the most relevant to the current user (as determined by the user's Filter Agent).

Host: **Your top sections
today are: Top Stories,
Technology, and
Business. What section
do you want to go to first?**

This overview acts as a sign post (i.e., “try these directions”), reminds the user of the legal vocabulary (i.e., the section names), and establishes a shared context [Clark 1992] for the conversation.

Suggesting relevant articles (the user’s list)

As discussed briefly at the beginning of this chapter (and in more detail in Chapter 5), the system assigns each incoming news article to one of eight news sections (e.g., technology) and to one or both lists within a section (the editors’ list or the user’s list). The user’s list is the set of articles chosen by the Filter Agent for the current user.

A user might choose her list instead of the editors’ list because she has developed sufficient confidence in the Filter Agent to select articles of interest to her. Alternatively, she may choose the editors’ list because she is interested to hear the major articles of the day.

Suggesting related articles

If a user shows particular interest in the current article, the interface responds by suggesting the article most related to it.²⁸ For example:

Host:	Clinton Visits Boston.	The host suggests a headline.
Jane:	Read it.	She accepts the headline.
News Reader:	(The News Reader reads the article.)	
Host:	On a related note, Clinton Fights Tax Bill in Congress.	The interface notes her interest in the previous article and, rather than reading the next article in the current list, it suggests the article most related to the previous one.
Jane:	Read it.	She accepts the related article and the conversation continues.

²⁸For information on how “relatedness” is determined, see Chapter 5.

Rather than constantly relying on a user to issue a command, the interface makes suggestions if it has reason to believe the user will want to take that action [Ly 1993]. By suggesting a related article, the interface is acting similar to a newspaper or TV news report in which editors' try to group related articles (by space and time, respectively).

Confirmations and error correction

People often misunderstand each other, but are adept at the steps necessary to correct errors: realizing an error may have occurred, entering a sub-dialogue to correct the error, and then continuing on with the conversation. Similarly, errors and misunderstandings are inevitable in human-computer communication, and the interface must be designed so that one or both parties in the conversation can realize an error has occurred and make corrections [Lewis 1986].

[Schmandt 1994] assigns speech recognition errors to one of three categories:

- *Rejection errors* occur when a user speaks an utterance from the recognizer's grammar but the recognizer doesn't hear (and, as a result, doesn't report) an utterance.
- *Substitution errors* occur when a user speaks an utterance from the grammar, but the recognizer reports it as a different utterance.
- *Insertion errors* occur when extraneous sounds (such as a user's breathing or noise in the environment) are mistakenly recognized and reported as an utterance from the grammar.

An interface could explicitly confirm each user utterance, but this would quickly become tiresome for a user. Instead, [Kamm 1994] recommends that confirmation be commensurate with the cost of the action which would be effected by the recognized utterance. More specifically, [Yankelovich 1995a] recommends the following criteria: if the recognized utterance will present data, then the interface should verify the utterance implicitly. If the recognized utterance will destroy data or set in motion future events, the interface should verify the utterance explicitly.

Implicit confirmation

NewsTalk implicitly confirms²⁹ a user's utterance using one of two methods: by speaking with a different synthetic voice or by using the same words a user spoke in her utterance.

NewsTalk uses one synthetic voice (the News Reader) to read articles and another synthetic voice (the Host) for all other utterances. In user studies, users quickly learned that the News Reader voice (which is deeper than the Host's voice) means that the system is reading an article. As a result, if a user rejects a headline but then hears the News Reader's voice, she knows that the system has made an error.³⁰ For example:

Host:	Clinton Visits Boston.	The Host offers Jane a headline.
Jane:	No.	Jane rejects the headline.
News Reader:	President Clinton came to Boston today, seeking support for ...	The recognizer reports that Jane said "Yes" (a substitution error), and the interface directs the News Reader to read the article.
Jane:	(Presses a touchtone on her telephone.)	Because she hears the News Reader voice, Jane realizes that NewsTalk is reading the article despite her rejection of it. She presses a touchtone to interrupt the news reader. ³¹
Host:	The House Passes Welfare Bill.	The Host offers Jane the next headline.

²⁹Implicit confirmation is one type of "feedback", what [Norman 1988] describes as "sending back to the user information about what action has actually been done, what result has been accomplished."

³⁰See chapter 4 for more advantages of using more than one voice.

³¹While listening to an article, a user can press any touchtone, which causes the News Reader to stop reading the article and the Host to read the next headline. Interruption could also be provided using speech, but that capability is not implemented in this system.

In all other situations, the interface implicitly confirms a user's utterance by using the same key words a user spoke. For example:

Jane:	Go to National .	Jane asks for the National section.
Host:	In the International section, do you want your list or the editors' list?	The recognizer reports "International" (a substitution error) and the interface implicitly confirms this.
Jane:	Go to National .	Jane realizes an error has been made and she corrects the interface.
Host:	In the National section, do you want your list or the editors' list?	The recognizer correctly reports "National", the interface implicitly confirms this, and the conversation is back on track.

Because the interface implicitly confirmed what the speech recognizer reported, Jane was able to detect the error and correct the interface. Note that, in this case, the interface didn't realize an error had been made, and it was only by some combination of luck and Jane saying the phrase differently that she got to the section she wanted. The next section deals with the case when the interface believes the recognizer did not correctly report a user's utterance.

Explicit confirmation

NewsTalk uses explicit confirmation in two instances: when the recognizer is uncertain that it is correctly reporting what a user said, and when a user wants to hang up.

When the speech recognizer hears a user utterance, it reports several items of information to the interface, including the utterance (as text) and a

measure of its confidence³² (as a number) that it is accurately reporting what the user said. When the recognizer reports a confidence below a set threshold, the NewsTalk interface will confirm the user's utterance explicitly:

Host:	What section do you want to go to next?	The interface prompts the user for a section name.
Jane:	Go to Sports .	She requests sports.
Host:	Did you say Sports ?	The recognizer reports that it heard "Sports", but its confidence is below the threshold set by the interface. As a result, the interface explicitly confirms what the recognizer reported.
Jane:	Yes.	Jane confirms the explicit confirmation.
Host:	Sports headlines in your list.	The interface implicitly confirms the section it heard and the conversation continues.

Also, the interface always explicitly confirms when the recognizer reports a Sign Off utterance (e.g., "Good-bye"). Obviously, it's important to confirm these Sign Off utterances explicitly because, if the recognizer has made a substitution error, the interface will prematurely hang up on the user.

Context-free confirmations

It's crucial that all confirmations be unambiguous to a user regardless of what she thinks the current context is. For example, the original interface replied "That's all?" when the speech recognizer reported that a user had

³²Not all recognizers report their confidence. Dagger (from Texas Instruments), the recognizer used by NewsTalk, does.

spoken a Sign Off utterance. However, in User Study 2 the following exchange took place:

Jane:	Go to the National section.	Jane specifies a new section.
Host:	That's all?	The speech recognizer mistakenly reports a Sign Off utterance, and the interface asks the user to confirm she is hanging up.
Jane:	(Pause.) Yes (tentatively).	Jane is a bit confused, but assumes that the meaning of NewsTalk's previous utterance is "The National section is the only thing you want?". She confirms. ³³
Host:	OK Jane, good-bye. (It hangs up.)	NewsTalk accepts the "confirmation" and hangs up.
Jane:	What happened?	Jane is frustrated and confused about why NewsTalk hung up.

As a result of this type of error, the interface was changed: all context-dependent utterances were replaced with context-free utterances. (In this case, "That's all?" was replaced with "Are you hanging up?".)

Accepting the blame for errors

[Schneiderman 1980] has observed that the tone of system error messages often causes a user to believe that she has made an egregious error and that, as a result, she is incompetent. Lewis points out the frustration these types of error messages can cause users:

³³The user reported this state of mind in a de-briefing session after the study.

“Failures to understand are commonplace and normal. Conversation is riddled with speech errors, from incomplete sentences to erroneous choice of words. But certainly we do not expect the people with whom we talk to respond to our speech errors with: *‘Your sentence was not grammatical. Say it again. (But this time do it right. Please.)’*” [Lewis 1986, p. 413]

NewsTalk attempts to avoid such an outcome: when the interface realizes an error has been made, it accepts the blame for it [Marx 1995]. This is important because it puts the interface in a subservient role and makes the user feel more in control. For example:

Host:	Welcome to NewsTalk. What is your first and last name?	
Jane:	Jane Smith.	
Host:	Did you say Jeff Herman?	The recognizer reports “Jeff Herman”, but with low confidence, so the interface asks for explicit confirmation.
Jane:	No.	Jane rejects the interface’s hypothesis.
Host:	My fault. What is your first and last name?	The interface recognizes that an error has been made, accepts the blame for it (“my fault”), and repeats the request for the necessary information.
Jane:	Jane Smith.	She repeats her name.
Host:	Hi Jane. What section do you want to go to first?	The interface implicitly confirms and the conversation continues.

Dialogue states

In NewsTalk, dialogue states are internal representations of stages of the conversation. These states follow the sequence of typical conversations and are designed to be transparent to a user. The properties of a NewsTalk dialogue state are:

- legal grammar subsets. Subsetting the grammar enables more accurate speech recognition because the recognizer has fewer legal utterances to choose from.
- interfaces utterances, including utterances for error correction.
- conditions which must be met before leaving the dialogue state. For example, the interface must know a user's name before it can leave the Greeting dialog state.
- the dialogue state that follows the current one.

Figure 3 shows the legal paths through a conversation with NewsTalk. For example, all conversations start in the Greeting state and progress to the Section state. Once a user specifies a section, she may be asked which list she would like. The List dialog state leads to the Headline dialog state. In response to a headline, a user can listen to an article (the Article dialog state), sign off, or choose another section. At any time, a user can utter one of the meta-communicative acts (such as asking the interface to repeat an utterance, to talk faster or slower, to pause, or to correct an error).

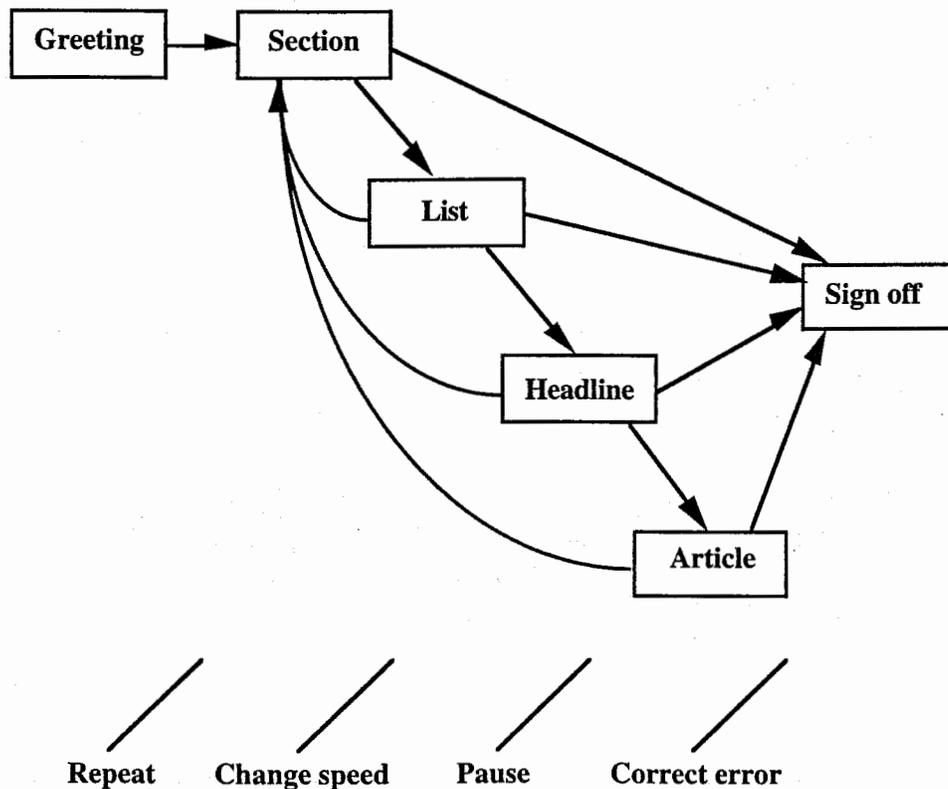


Figure 3: NewsTalk's finite state diagram shows the dialogue states and the legal transitions between them. The dialogue states listed at the bottom of the figure are always available.

The specifics of most of the dialog states have been covered in previous sections. The next section covers the remaining dialog states, meta-communicative acts.

Meta-communicative acts

Meta-communicative acts [Sparks 1994] enable a conversant to perform a variety of actions to maintain and control the dialogue itself, such as confirming or repeating what was last said, repairing misunderstandings, and soliciting or providing help or further instructions on how to proceed.

In NewsTalk, these acts enable a user to ask the interface to repeat the previous utterance, to change the speaking rate of the synthesizer, to pause the interface, and to correct errors.³⁴

Repeat

When conversants have trouble understanding what was just said, they often ask a conversation partner to repeat it. The speaker is likely to rephrase the utterance and/or decrease her rate of speaking.

NewsTalk incorporates the latter approach when a user requests a repetition (by saying "Repeat that") of what was just said. This capability was one of the most welcomed features in user studies, although users did not use it often. When this option became available in the interface, users reported feeling more in control because they knew they had more than one chance to understand what the synthesizer said.

Change speed

People often have trouble understanding synthetic speech when first exposed to it. However, their ability to understand such speech improves significantly with minimal exposure to it [Schwab 1985].

Since people's ability to understand synthetic speech varies, it's important to give users control over the speech rate. In NewsTalk, users can change the speaking rate of the synthesizer by saying "talk slowly" (which sets the synthesizer to 180 words per minute), "talk fast" (230 wpm), and "talk very fast" (280 wpm). In user studies, more experienced NewsTalk users tended to listen to news at one of the faster rates.

³⁴Error correction was covered in a previous section.

Also, when a user finishes a conversation, NewsTalk saves the current speech rate in order to set it again automatically in the next conversation with that user.

Pause

At any point in a conversation, a user can cause the interface to pause, i.e., to stop talking and to stop listening, by saying "Stop talking". This capability is useful when a user wants to take a break without hanging up (for example, to take a call on another line). It has also proved useful during user studies in which the investigator, a user, and the system are in a conference call; a user can pause the interface in order to make a comment or ask the investigator a question.

Comparing speech input and IVR systems

Now that the NewsTalk user interface design has been described, it's useful to compare the advantages and disadvantages of speech input to that of interactive voice response (IVR) systems. IVR systems are currently the most common way that people use a telephone to access information stored on a computer. For example, existing IVR systems enable people to access news [Irish Times 1994], movie time and locations, and train schedules. Such systems allow a person to communicate with a computer by pressing one of the twelve DTMF³⁵ touchtone keys on a telephone keypad.

IVR systems

Touchtones were designed to be unambiguous input to a device. This is a distinct advantage over speech input, which is prone to recognition errors. Further, many people have experience using an IVR system and, as a result, are accustomed to listening to a menu of options and pressing the appropriate touchtone. In contrast, many people have little or no experience speaking to a speech recognizer: they may have trouble determining the appropriate speaking style they should use and determining the limitations of the recognizer.

Also, pressing touchtones may be a more appropriate action in certain social situations (e.g., while attending a meeting) and it may be faster [Stifelman 1992]. Pressing touchtones is more private for a user than speaking: users of a speech recognition-based e-mail reader report some discomfort using it at a public payphone [Yankelovich 1995b].

³⁵DTMF stands for dual tone multi-frequency signal.

A number of development tools exist that make it easier to design and implement a touchtone interface, and the hardware necessary to run a touchtone system is much less expensive than that required to do speech recognition.

But interfaces that rely on touchtones have significant problems in usability, availability, and accessibility. [Yankelovich 1995a] describes current IVR systems as “often characterized by a labyrinth of invisible and tedious hierarchies which result when menu options outnumber telephone keys or when choices overload users’ short-term memory.” [Engelbeck 1989] recommends that “when helps are an issue (i.e., novice users are important), menus should be kept to four or fewer choices.” Also, many telephones do not have touchtones. [Schalk 1992] reports that 30% of calls placed in the U.S. are from rotary dial telephones.

Finally, one of the main advantages of audio information is that you can listen to it while accomplishing other tasks. Touchtones mitigate this advantage to a large extent, because they require the use of the hands and, often, the eyes. In the NewsTalk user studies users reported using NewsTalk’s speech interface while driving, walking, and rollerblading, and expressed a desire to use NewsTalk while accomplishing other tasks around the house, including preparing meals, washing dishes, and getting dressed.³⁶

Speech input

Many of speech input’s advantages and disadvantages have been discussed previously in this thesis, including the advantages of: hands-free/eyes-free use and a familiar metaphor (spoken conversation); and the disadvantages of recognition errors, learning the vocabulary and syntax, and learning the appropriate speaking style.

Speech input is natural for the speaker, and it allows a user to combine utterances rather than working through a hierarchy [Scharf 1994]. Also, speech input enables a one-to-one mapping between a user’s action and the system’s response and, in contrast to a touchtone interface, the number of legal utterances at any one time is not limited by physical space. However, adding utterances to a speech recognizer’s grammar comes at a cost: more legal utterances in the grammar tends to cause more speech recognition errors.

³⁶See the section *Using NewsTalk while involved in another task* in chapter 4, User Studies.

In addition to grammar size, the accuracy of the speech recognizer is dependent on a number of other factors, including variations in the user population (including familiarity with the interface and different voice patterns), microphone quality and placement, ambient and channel noise, and low signal bandwidth [Schmandt 1994]. Also, recognition results tend to be more accurate when a user is speaking on a land-line telephone rather than a cellular phone.

Finally, new devices will appear on the market in the next 5-10 years that are so small they will not have room for the 12 key touchtone keypad. For these devices, speech will most likely be the primary interface.

4. User Studies

*"The primary test of a user interface is its success with users."
—[Apple 1987]*

Introduction

By involving users in all stages of the process, designers are more likely to create products based on users' abilities, tasks, and environments rather than products based on underlying system mechanisms and assumptions [Ericsson 1984]. By observing users' reactions to a proposed design, a designer can:

- learn about user's abilities, environments, and tasks;
- determine how a design can be improved;
- foresee how a design may positively or negatively affect a user;
- brainstorm possible solutions with users.

User studies were conducted before and during the design of this thesis as an iterative part of the design process.

In the first study, four pairs of participants conducted telephone conversations about the day's news. As part of the scenario, one person in each pair was arbitrarily designated the assistant; the other was the manager. The assistant's goal was to provide the manager with the most useful news available from two newspapers. The manager's goal was to receive the most useful news from the assistant. The purpose of this study was to determine how managers would ask for information, how assistants would communicate it, and what process the assistant would use to determine the news interests of the manager.

The second study had two phases. In the first phase, several novice users interacted with an early version of the interface in three conversations via a telephone. The goal in this phase was to determine the parts of the design that aided or hindered a user who was unfamiliar the interface.

In the second phase, three of the participants continued using the interface approximately four times a week for a period of one week to two months. The purpose of this phase was to determine how well the interface served the needs of these users as they became more familiar with it.

The methodology and findings of the first user study (person-to-communication) are in Appendix 2. The remainder of this chapter the methodology and findings of the second study.

Descriptions of participants in User Study 2

Before discussing the results of this study, it's important to know about the individuals who participated in it, including: their interest in news; their abilities; their experience with computers, in particular speech recognition and speech synthesis; and the circumstances in which they used NewsTalk.³⁷

Long-term users

- **Beth** is blind and is a consultant on accessibilities in her 30s. She uses a portable computer with speech synthesis at home and at work. She always carries a number of electronic devices with her: a Braille 'n Speak³⁸ to take notes, a dictionary-thesaurus, and a watch that synthesizes the current time. She has an alarm clock with speech synthesis.

She listens to radio news 2 1/2 hours each day. Each week she listens to Newsweek magazine on audio tape.

She used NewsTalk via telephone at her home and in her office three to four times a week for a period of two months. Typically, these conversations lasted 20-30 minutes. Prior to these studies, she had no experience talking to a speech recognizer.

- **Nancy** is a graduate student in transportation studies in her 20s. She uses a computer at home and at work.

She reads the New York Times, the Wall Street Journal, and the Boston Globe. She listens to National Public Radio (NPR) and the BBC.

³⁷All of the participants names have been changed to protect their privacy.

³⁸A Braille n' Speak [Blazie 1991] is a lightweight (approximately one pound), battery-powered notetaking device for the blind. A user enters text by typing on a seven key chorded keyboard, and the device can read the text aloud using speech synthesis.

She used NewsTalk via telephone at her home and in her office. She used NewsTalk for a period of one month. Prior to these studies, she had no prior experience listening to synthesized text or talking to a speech recognizer.

- **Richard** is a researcher at MIT. He has extensive experience using and creating programs for computers, in particular programs for visual user interfaces. He had a moderate amount of experience listening to speech synthesizers, and he has extensive experience using speech recognizers.

He used NewsTalk while driving in his car to and from work for a period of one week. His conversations lasted approximately 15 minutes.

Short-term users

- **Diane** is a market researcher in her 30s. She uses a computer at work, but does not own one at home. She reads a newspaper 3 times a week and listens to radio news 90 minutes a day. She used NewsTalk twice as part of the initial phase of this study.
- **Charles** is a graduate student at the MIT and is in his 20s. He has extensive experience using and creating programs for computers. He currently gets his news from an online newspaper, which he reads twice a day. He used NewsTalk twice as part of the initial phase of this study.
- **Ted** is blind, in his 30s, and works in public relations. At work he uses a computer with speech synthesis. He always carries a Braille 'n Speak with him to take notes and to keep his calendar. He has extensive experience listening to synthetic speech, but no experience talking to a speech recognizer.

He listens to radio news one hour per day, and a friend reads him excerpts from the newspaper once a week. He also receives the Braille-large type version of the New York Times approximately two weeks after its publication.

He used NewsTalk twice as part of the initial phase of this study.

- **Edward** is an interface designer in his 30s. He listens to radio news 90 minutes a week, and reads online news 15 minutes a week. He uses a computer at work and at home. He used NewsTalk twice as part of the initial phase of this study.

Findings

Before participating in this study, participants read one page of instructions on how to use NewsTalk (see Appendix 3). In this study, NewsTalk was available for use via telephone 24 hours a day.

First impressions of new users

Overall, early experiences with the interface were positive. Here are some comments that were typical of users' first impressions of the interface:

Beth: "It's simple. I don't have to remember much to use it."

Richard: "It's very easy to use. I felt I was productive right away."

Charles: "I have found myself wanting to use it to get the news. For example, last night my girlfriend and I were at a restaurant having an argument about something that was in the news that day. And I thought to myself, 'I should call [NewsTalk], find the information, and we can get on with dinner!'"

However, the majority of new users had trouble understanding the synthesizer's speech:

Diane: "The [synthesized] voice bothers me. It sounds strange, odd."

Edward: "I have trouble understanding the [host's] voice. It's easier for me to understand the [news reader's voice], perhaps because there is more context—it's reading complete sentences."

But, consistent with the results of [Schwab 1985], users were able to understand the synthesizer much better after approximately three conversations:

Nancy: "I'm getting used to [the synthesizer]. At first, you have to listen very hard. I wasn't comfortable the first few times I called, but since then it's fine."

Occasionally, even an experienced user had trouble understanding a specific synthesizer utterance. In these cases, the conversation would momentarily or permanently break down. The command "Repeat that" was added to the grammar and is available after any interface utterance. When this command is received, the interface instructs the synthesizer to repeat its most recent sentence, and to speak the utterance at a slower rate.

Reactions to the spoken conversation metaphor

One of the main goals of the interface design was to enable new users to feel competent using it. Familiar metaphors, including spoken conversation (as discussed in the previous chapter), were chosen in order to aid this goal.

One of the questions about NewsTalk was how people would react to this metaphor. In general, users found it natural and comfortable:

Richard: "I like the conversational milieu that it puts me in. Because it's conversational, I'm listening for instructions. I trust it immediately. It's very positive. It called me by my first name right away. I find that very reassuring."

Beth: "I like talking to it because it keeps you involved. If you just push buttons it's rather rote: I don't think you pay as much attention. If you have to verbalize then you have to think about it, and I think you retain more of the news."

Edward: "It's much better than pushing buttons. It's more natural. And you don't have to reach over and press any buttons."

Despite the generally favorable reaction to the spoken conversation metaphor, it did cause problems. One of the most common was that users sometimes forgot commands that they didn't use often:

Nancy: "I wish it would remind me what the names of the sections are. I listen to the International section a lot, and the Business section, so I remember those. But sometimes I forget some of the other section names."

Also, one user found it tedious to respond verbally (i.e., by saying "yes" or "no") to each headline:

Charles: "I would like to have a keypad interface for things like "yes" and "no". But I like saying everything else."

Information organization

Another concern was how users would react to the way the articles were organized:

Beth: "I like that it's broken down into sections. I can skip entire sections. I can find what I want."

Richard: "At first, I had some trouble building up an image of how the articles are organized. But I just had to use it a few times. Now it seems quite natural."

Nancy: "One of the things I like is that I can focus on just the sections I want, rather than a newspaper where you get everything every day."

Fixing vague prompts

In the early versions of the interface, it began a conversation by saying "Welcome to NewsTalk. Who's there?". The interface expected users to say their first and last name only. However, the fairly open-ended question "Who's there?" led to a wide variety of responses from users, including "Joe", "It's Joe", "This is Joe", "This is Joe Smith", "Hi, I'm Joe." Since these responses were outside of the recognizer's grammar, many recognition substitution errors occurred. For a system that is tracking each user's actions and making decisions based on those actions, this is an unacceptable result.

To solve this problem, the opening prompt was changed from the friendly "Who's there?" to the more specific "What is your first and last name?" After making this change, users were much more likely to say only their first and last name and, as a result, the recognizer was much more likely to return the correct name.

Also, when the recognizer reports that a user spoke one of the Sign Off utterances, the interface always explicitly confirms that the user wants to hang up. The original confirmation, "That's it?" (see Chapter 3), led to confusion and was replaced by "Oh, you're hanging up now?". However, one new user heard this prompt as a statement, "OK, you're hanging up." As a result, the confirmation was changed again to the more clear question "Are you hanging up now?".

Adapting confirmations based on user experience

When an early version of the interface suggested a headline, a user could ask to hear the corresponding article by saying "yes", or reject it explicitly by saying "no." However, in several cases when a user said "no", the recognizer reported "yes." Because the users were unfamiliar with the interface, and because they assumed that their response had been understood, they did not realize that the interface had made a mistake and was now reading an article.

Two changes were made to fix this problem. First, the grammar was expanded to include “read it” (synonymous with “yes”) and “skip it” (synonymous with “no”). (The recognizer reported “Read it” and “skip it” accurately more often than “yes” and “no”.) Second, the interface was changed to implicitly confirm a user’s acceptance of an article by saying “OK, here’s the article” before reading each article.

This confirmation proved useful for a short time, but users reported that it quickly became tedious. In the final design, the implicit confirmation is used only before the first article read in each conversation; no confirmation is given before subsequent articles. Also, as discussed in Chapter 3, one synthetic voice is used to read articles and another voice is used for all other interface utterances. This change in voice helps a user realize when the interface has started reading an article. Similarly, the return of the Host voice after a long article alerts a user that the article is over and that the Host is offering the next headline.

Using NewsTalk while involved in another task

As discussed, Richard used NewsTalk while commuting in his car to and from work. He spends approximately 30 minutes driving to work, and that time is split between freeway driving and fighting heavy city traffic. One concern of the investigator and of Richard was that he would become distracted or immersed in NewsTalk, and his driving performance might suffer. In his experiences, this was not the case—using NewsTalk did not prove to be a distraction:

Richard: “I’m completely surprised that I don’t zone out when I get the news [from NewsTalk]. [NewsTalk] is like having a conversation with another person. It’s a very natural thing to do—to talk and listen.”

One of the main advantages of a speech interface is that you can use it while doing other tasks. As discussed, Richard used the interface while driving. The other two long-term users used NewsTalk at home and in their office; they both expressed a desire for a speaker phone so that they could use NewsTalk while doing other tasks around their home or office:

Beth: “A speaker phone would come in handy, because then I could do other things: eat, clean my room, get dressed.”

Nancy: “I wish I had a speaker phone, so I could do other things: clean up my desk, organize papers, write notes. I also think it would be useful if I was driving.”

Similarly, some of the short-term users wanted to use NewsTalk while attending to other tasks:

Ted: "I'd like to use it while I eat breakfast. If I had a speaker phone, I could just talk and my hands would be free for preparing the food and eating. I wouldn't be tied to the phone."

Nat: "I'd like to try using it every day so I can read mail and listen to [NewsTalk] at the same time. I like it because you don't have to focus on it fully."

Charles: "One of the nice things is I could use it while doing something else, like washing the dishes."

Edward: "I wouldn't use it over a telephone, but I would with speaker phone. Also, I'd use it while driving. Now I listen to radio news while I'm making coffee, eating breakfast, getting dressed. I'd like to use NewsTalk during that time. I don't want to make getting news an act in and of itself."

Getting information with less effort

Late in this study, the investigator designed a new feature with the goal of enabling a user to receive the same useful information but with less effort. The purpose of the feature was to reduce the actions required of a user while still providing news of interest. Reducing the number of actions also reduced the opportunity for speech recognition errors.

Specifically, when NewsTalk offered a headline a user could respond with the utterance "Just do it". This caused the interface to read each article in sequence without input from the user. When the user wanted to switch back to conversing with NewsTalk, she could press 1 on the touchtone keypad. Pressing any other key caused NewsTalk to skip the current article and to move on to the next.

In general, the three users who tried this feature liked it, although two of the three concluded that they were retaining less information:

Beth: "What's nice about [the new feature] is that there isn't a gap between the headline and the article—I can get more information about an article in the same amount of time, so I find out faster what the article is about and I can decide to listen to it or not in less time."

“Whether I’d use [the new feature] depends on the conversation. If I’m not very awake, then I probably wouldn’t, because talking to it keeps me involved and more alert.”

Nancy: “I used [the new feature] in pretty much the same way as before. The only difference is that, because I no longer have to do anything to hear an article I don’t listen to each article as closely, and I think I don’t take in as much of the information.”

Improving the overview

The current interface suggests the best sections for each user, but a better overview is clearly needed. One user put it this way in an e-mail message:

Nancy: “Could it give a quick list of everything first and then ask for a choice? Since time is likely to be limited, and under the current structure something next could always offer more interest but you have no knowledge of what is coming up later, the process of actually choosing a story can become fairly random -- (e.g., Say I have heard 5 headlines and rejected them all and now feel that this one sounds vaguely interesting, I don’t know how many are left in the list, so I’ll say yes.)”

Several solutions were proposed for this problem. The most promising is applying a text understanding system that groups articles according to subjects that occur most in the news that day. For example, articles in the National section might be subdivided by Congress, Clinton, and Oklahoma City (during the period of the terrorist bombing there and its aftermath). The interface could make these subjects known to a user, perhaps when a user goes to the corresponding section, and a user could jump to articles on a particular subject by speaking that subject’s name.

The editors’ list and the user’s list

Another issue was whether users would understand the purpose of the editors’ list and their list. Would they choose one or the other depending on the type of news they were looking for? From their comments, it appeared that they did understand the purposes of these lists. Also, it appeared that users understood that they needed to use the system for some amount of time before they could expect the system to select articles that matched their interests.

Beth: "It's not a broadening experience if I only listen to things I'm interested in. It doesn't broaden my perspective. I like having other stories to choose from other than the ones that matches my interest exactly."

Nancy: "At the beginning I didn't listen to my list because I didn't think it would [have articles that matched my interests]. But after a while I decided to give it a chance and I thought it did quite well."

Other long-term users echoed Nancy's comments that the Filter Agent found articles that matched her interests:

Richard: "It finds information I need to know that I would never have found on my own."

Beth: "It's hard to know how well it's finding articles for me without knowing what it isn't giving me. But I think by and large it's doing a good job of giving me articles that concern topics that I am interested in."

Final impressions from users

At the completion of the study, participants were asked their overall impressions of the system, the things that they liked about it and the things they found frustrating. These quotes are typical of their responses:

Beth: "I find it a really useful tool since I can't see a newspaper. It meets my needs."

"Except for radio, I have a hard time getting timely news. But I can't control what news they cover on the radio, and I can't control how long the articles are. With [NewsTalk] I'm in control and I make the choices. That's great."

"I wish it had editorials and advertisements. And it should describe political cartoons—Newsweek [on cassette tape] does that. Also, I'd like to tell it to spell certain names, the ones that [the synthesizer] has trouble pronouncing."

Nancy: "I like it because there are useful stories that I don't get in my other news sources. However, I'd like to be able to be more specific in my search, for example, "What's the economic news from London today?"

"I like that it has related articles. The first one is usually main-stream, the second one is more unusual."

Richard: "Of course, there's a learning curve: it took me a few times to figure out how I should talk to it and what I can say. But I am much better at using it now. When it doesn't work, it's not particularly frustrating. By and large the whole experience has been quite positive. It's a useful tool."

Diane: "There's definitely a market for something like this that provides convenience, accessibility, and immediacy, especially for information hounds, or people with physical disabilities."

"An enhancement would be if you could select your sources. For example, an investor who was tracking mutual funds might benefit the most from listening to the Morningstar daily report."

Charles: "The concept of NewsTalk is very appealing. I can see using it in a car if I really wanted to know about something specific. But otherwise, I'd rather just listen to music or listen to whatever was on the radio news. I don't have a computer at home, so I could see using at home on the weekends too."

Ted: "I'd like to move within an article: back up a sentence, or move forward a paragraph. And I would like to access ads. But it's the easiest way I've seen to get news without using your eyes. I wish someone would market it."

Edward: "It drives me nuts that I can't get what I want from the radio. If I want the weather, I have to sit around and wait until the weather comes on. The part [of NewsTalk] that I really like is the control."

"Now, if there is something interesting on the radio in the morning then I have to wait to turn on the coffee grinder, because I can't hear the radio over the grinder. I can't stop the radio, but I'd like to. It's having control over the information that entices me [about NewsTalk]."

All of the long-term participants asked to continue using NewsTalk after the study was completed.

5. Personalized Information Agents

Introduction

“The idea of an agent originated with John McCarthy in the mid-1950’s, and the term was coined by Oliver Selfridge a few years later, when they were both at the Massachusetts Institute of Technology. They had in view a system that, when given a goal, could carry out the details of the appropriate computer operations and could ask for and receive advice, offered in human terms, when it was stuck. An agent would be a “soft robot” living and doing its business within the computer’s world.

—Alan Kay [Kay, 1984]

[Maes 1993] defines an interface agent as a computer program that learns the preferences of users and automates repetitive or predictable computer-related tasks. [Laurel 1995] describes interface agents as “metaphors with character.” Agents can perform many useful tasks including teaching [Chin 1991], scheduling meetings [Kozierok 1993], and filtering information [Sheth 1993].

[Schneiderman 1995] has described agents as systems that exhibit most or all of the following characteristics:

- anthropomorphic representation
- adaptive behavior
- accepts vague goal specifications
- gives you just what you need
- works while you don’t
- works where you aren’t

Information filtering

One of the most common roles of an interface agent is that of an information filter. Both information retrieval (IR) and information filtering (IF) systems are designed to enable users to find useful information; traditional systems use the following steps [Ellis 1990] :

- 1) A text query is submitted to the system. (The query is submitted by a user or another program.)
- 2) The system calculates the similarity, i.e., relevance, of the query to each document.
- 3) The system returns a list of relevant documents.

However, [Oddy 1977a; Oddy 1977b] and [Belkin 1992] point out two areas where IF systems diverge from IR systems:

- Typically, a user will interact with an IF system on a number of occasions and, often, with long-term goals. As a result, IF system designers can take advantage of these frequent interactions and long-term information needs by maintaining a profile of each user, and using this profile to personalize the information for each user.
- In contrast to a typical IR user, a user of an IF system is less likely to be highly motivated, less likely to have a well-defined information need, and is less likely to express an information need in the terms expected by the system. As a result, it is more important for IF systems to be clear and easy to use rather than being overloaded with myriad options.

The next section reviews information filtering systems, with a focus on IF systems that provide information using non-traditional methods.

Related work

THOMAS

THOMAS [Oddy 1977a; Oddy 1977b] was an information filter with which a user participated in a screen-based “dialog” that was intended to resemble a human-to-human conversation. THOMAS’s design was based on the following observations:

- Users of an information retrieval system often do not have a clear query.
- Users find it difficult to express their query in a form that yields useful results from the system.³⁹

³⁹[Norman 1986b] calls this the gulf of execution.

- Users' interactions with an information retrieval system are often heuristic and interactive.

As a result, THOMAS was designed with the assumption that a user's information need was not a static entity, but something which would likely change during the course of a retrieval session.

Specifically, THOMAS presented a series of screens which guided the user through a dialogue and, ideally, to items of interest. The first screen enabled a user to enter document titles, subject terms, or authors. THOMAS then presented a list of relevant documents, which the user could open. In addition, a user had the option to: 1) indicate whether a returned document was of interest or not, 2) select or reject terms displayed in a document's representation, and 3) enter additional document titles, subject terms, or authors.

THOMAS relied on user feedback and a dialogue history to create and refine a model of the user's requirements, and used the model to choose which documents to offer to the user. If THOMAS came to a point in the dialogue where it had no relevant documents to offer, it would either prompt the user to reconsider documents already rejected, or ask the user to once again enter document titles, subject terms, or authors.

Oddy determined that, when using a small collection, THOMAS performed as well as traditional interactive query-based systems. THOMAS was an early and significant step in suggesting how designers could reconsider how users search for information, and what users' thought processes were while searching.

Guides

Guides [Oren 1990] is a multimedia database that contains information about American history during the period 1800-1850. A primary goal of the system is to "reduce the cognitive load on users that is created by 'navigating' while trying to learn."

Anthropomorphized agents, or "guides", are used to assist a user in determining what information she should or could proceed to next.

In addition, guides are used to create a "narrative metaphor" and to communicate a specific point of view. For example, if a user wants to learn about life in Northern California during the 1849 Gold Rush, she could click one of the guide icons. The guide appears in a video on the computer screen and relates the event from his/her perspective. (The types of guides include a scout, a slave, an inventor, and an Indian.)

Each guide icon shows the corresponding guide's degree of interest in the current information by posing in one of four "stock" gestures. Each gesture is consistent with each guides "personality" and other character traits [Laurel 1991].

A user can also create a "custom guide" by naming a new guide, then specifying areas of interest from scrolling lists of topics. The custom guide is then available to direct the user to areas that matched the user's specifications.

Newt

Newt⁴⁰ [Sheth 1993] is an on-screen interface to an electronic news filter. In Newt, each user has one or more agent filters. Each agent contains a list of news groups it should use to search for articles, and terms it searches for within the articles. Users can edit these lists.

Each agent proposes a list of articles (ranked by relevance) to a user. A user may give positive or negative feedback to the agent by clicking its plus (+) button or a minus (-) button. This feedback adjusts the agent's profile so that it will be more or less likely to find similar articles in the future. Newt uses a genetic algorithm [Holland 1975] in order to evolve the agents so that they locate more relevant articles over time.

Sheth conducted a two week user study and noted that, despite encouraging results, users had a number of difficulties with the system, including:

- After reading an article, users often would forget to provide feedback (either positive or negative). Because the learning process relied on this explicit feedback, if a user failed to give feedback then the agent did not learn.
- To use the system effectively, users needed to create an accurate "model" of the agent, i.e., the criteria it used to select articles and the reasons it selected specific articles. In a survey of users after the user study, Sheth found that "people had mixed reactions when asked if they could develop good agent models."

⁴⁰Newt is an abbreviation of News Tailor.

- User feedback did not affect the system until the next time the system collected articles and presented the filtered list to a user.⁴¹ Sheth noted that this “lack of immediate response prevented efficient communication” between a user and the system.

First!

First! [Individual 1995] is a personalized news filtering service created by Individual, Inc., of Burlington, MA.

To begin using First!, a new customer writes her interests in a text file. Each day, this text file is used by the information retrieval system SMART [Salton 1983] to determine which news articles may be of greatest interest to the customer. An editor double-checks SMART's choices, and the filtered articles are delivered to the customer via e-mail, fax, Lotus Notes, or the World Wide Web.

Periodically, Individual, Inc., sends each customer a list of headlines she received in the preceding period, and the customer indicates the relevance of each headline. In addition, the customer is asked to describe articles they don't want to receive, and desired articles she thinks are missing from her filtered articles. The editors use the results of the survey to fine-tune the customer's profile.

The next several sections describe the collection and filtering process designed and implemented for this thesis project, with a special emphasis on how the system adapts over time and, as a result, selects more useful articles for each user.

A personalized information agent

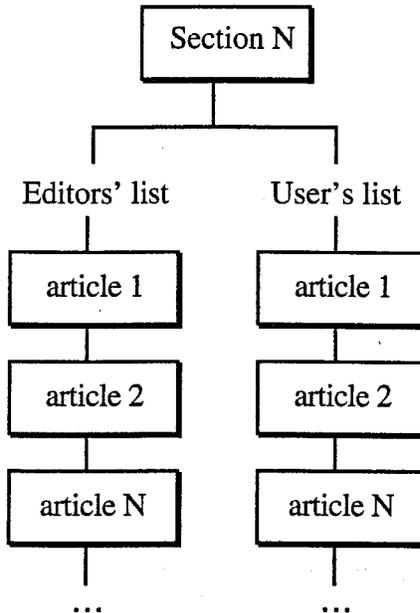
Collecting articles

The News Collection program gathers approximately 500 text articles daily and approximately 15 audio broadcasts hourly from a number of sources. The News Assignment program assigns each item to one or more news sections: Top Stories, National, International, Local, Business, Technology, Living, or Sports. The assignment is made based on one or more properties of the article, including its source (e.g., all articles from Edupage are assigned to the Technology section) or labels assigned by a

⁴¹In Newt, articles were collected and filtered once a day.

news editor (e.g., all ClariNet articles in world.europe.ireland are assigned to the International section).

Each news section contains two lists—the editors' list and the user's list:



- The editors' lists contain the top stories in a specific section as determined by news editors. These are the articles that one would typically see on the front page of a newspaper, or hear at the beginning of a television or radio newscast. The editors' lists are the same articles in the same order for all users. Editors' lists typically have ten to twenty articles.
- The user's list contains the most relevant articles in a specific section; relevance is determined by each user's Filter Agent. The Filter Agent selects articles for this list by comparing the incoming articles with articles the user has shown a particular interest in in the past. The Filter Agent lists the articles from most to least relevant. User lists may have zero to twenty articles.

When a user begins using NewsTalk, she has articles in the editors' lists only; her lists are empty because her Filter Agent knows nothing about her news interests. The following sections describe the steps the Filter Agent goes through in order to select articles for and order articles in each list for each user.

Filtering articles

The system (see Figure 1) records in the User Profile the text of each article a user has listened to (in NewsTalk) or explicitly shown an interest in (in NewsBlast⁴²). In the User Profile, the articles added on a particular day are considered to be a set. The Filter Agent uses the User Profile articles in order to query the incoming articles (using the information retrieval system SMART) in order to determine which incoming articles are likely to be of interest to each user.

The Filter Agent biases the results of the SMART queries so that more recent sets in the User Profile are considered to be more relevant than older sets. As a result, a user is more likely to hear articles related to the news she heard yesterday than what she heard last month. The rest of this section explains the details of this process.

To begin the filtering process for a user, the Filter Agent sends an article in her User Profile as a query to SMART.⁴³ SMART returns a list of incoming articles and their relevance to the profile article. The Filter Agent biases the relevance values returned by SMART toward incoming articles related to more recent profile articles.

Specifically, the articles heard in each dialogue are considered as a set in the User Profile. The relevance values of the incoming articles related to the most recent set are multiplied by 1, the bias of that set. The bias of each succeeding set is the bias of the previous set multiplied by .95. The result is

⁴²The feedback loop from NewsPage has not been implemented.

⁴³SMART [Salton 1983] is an information retrieval system developed by Gerard Salton and his associates, first at Harvard University and now at Cornell University. It was chosen for this project because of its accuracy [Salton 1989], computational efficiency, and performance [Buckley 1985].

SMART accepts queries of an arbitrary length, i.e., anywhere from one word to an entire document. It retrieves documents using automatic indexing and the vector space model. *Automatic indexing* (including a negative dictionary, a synonym dictionary, a phrase dictionary, word stemming, statistical term associations, syntactic analysis, and hierarchical term expansion) produces a list of a document's descriptors, or terms. The *vector space model* determines each terms' weight, i.e., its descriptive power, using two factors: term frequency and inverse document frequency. Terms which appear more often in a document are assigned higher weights (term frequency), and terms which appear in fewer documents (i.e., the more specific terms) are assigned higher weights (inverse document frequency).

an exponential decay of the bias of each set from most recent to oldest. For example, the bias of each set after the first 20 dialogues is depicted as (Figure 4):

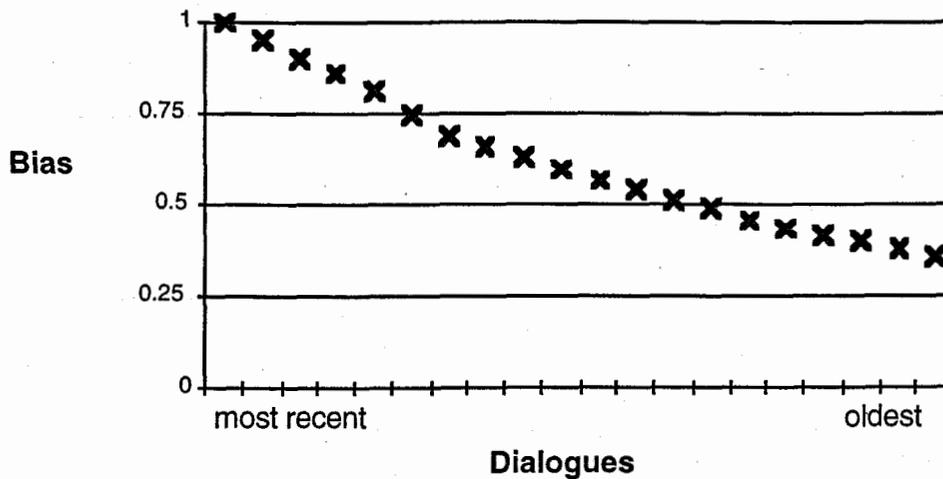


Figure 4: The bias of each set in the User Profile exponentially decays from the most recent to the oldest set.

Each profile article is sent to SMART in turn, and the Filter Agent accumulates the biased relevance of each new article. At the completion of this process, the new articles with the highest accumulated relevance are added to the user's list in the appropriate sections in their order of relevance.

This entire process is repeated for each User Profile so that each user has access to new articles filtered according to her profile.

Finding related articles

As a separate process, each incoming article is sent to SMART as a query, and the system records the incoming article most similar to it.

NewsTalk uses this information to make suggestions to a user based on the interest she shows in specific articles in the current conversation.

Specifically, after a user has shown a particular interest in an article, NewsTalk will try to continue the theme of that article by suggesting the article most similar to it.⁴⁴ In this way, NewsTalk is actively changing the

⁴⁴Refer to the section "Suggesting information" in Chapter 3 to see specifically how the interface suggests related articles.

order of the editors' and user lists according to what it has learned about a user in the current conversation.

Using a software tool to find related articles is not a fool-proof methods and mistakes do occur. In the current scheme, two articles that have a relevance rating between .4 and .5 (as determined by SMART) are judged to be related. Two articles with a relevance greater than .5 are judged to be essentially the same articles, and the system will remove one. Two articles with a relevance .4 are judged to be unrelated.

Supporting serendipity

An important goal of any personalized information agent is that it not become too focused on a small set of topics. In NewsTalk, one of the purposes of the editors' lists is to prevent this from happening. Articles are assigned to one of the editors' lists because of a human editor's judgment that the article is important, regardless of a user's interest in the article.

In addition, the editors' lists have other purposes. The lead articles in any news service are intended to inform all recipients of important information. As a result, these articles create a common dialogue and focus a community on important issues and events of the day.

Also, when a user first uses NewsTalk, the editors' list articles are the only ones available. This is because her User Profile is empty and, as a result, the personalized information agent hasn't selected any articles specifically for her. As a result of listening to articles in the editors' lists, articles are added to her User Profile, and these articles are used to filter future incoming articles for her lists.

One user study participant used NewsTalk for two months. At the conclusion of the study, she was asked why she sometimes chooses to hear articles in the editors' list rather than articles in her list. Her response reflects one of the main purposes of the editors' list:

"It's not a broadening experience if I only listen to things I'm interested in. It doesn't broaden my perspective. I like having other stories to choose from other than the ones that match my interest exactly."

Scaling up

The process just described is computationally intensive and would need to be altered to support a significantly larger user population. Specifically, it

takes approximately 1 second to process each article in a User Profile, or one hour to process 3600 articles.

The process could be altered in a number of ways to support more users. One option is to assign different tasks in the process to different computers, rather than running all tasks on one computer. A second option is to save in the User Profile only the salient features in each article, rather than the entire article. A third option is to limit the number of articles in each user's profile. These options are not mutually exclusive.

Quantitative evaluation of results

To determine whether the Filter Agent is indeed selecting articles of particular interest to specific users, a set of experiments were conducted.

On several occasions, each of the three long-term users were presented with a sample of articles retrieved by their Filter Agent (15 samples out of approximately 100 articles retrieved). They were also presented with a sample of articles not retrieved (i.e., discarded) by their Filter Agent (15 samples out of approximately 400 articles discarded). The thirty articles from both groups were sorted in a random order and presented to a user in sequence. The user was asked to indicate her relevance judgment, or interest level, in each article. The results are shown in Figures 5-7.

Recall and precision are two of the most common measures of a filter system's effectiveness [Salton 1983]. Recall is the percentage of items that match a user's interest that were retrieved. Precision is the percentage of articles retrieved that match a user's interests. If RETREL is defined as the number of items retrieved and relevant, RETNREL is the number retrieved but not relevant, and NRETREL is the number not retrieved but relevant, then the formulas for recall and precision are:

$$\text{recall} = \frac{\text{RETREL}}{\text{RETREL} + \text{NRETREL}}$$

$$\text{precision} = \frac{\text{RETREL}}{\text{RETREL} + \text{RETNREL}}$$

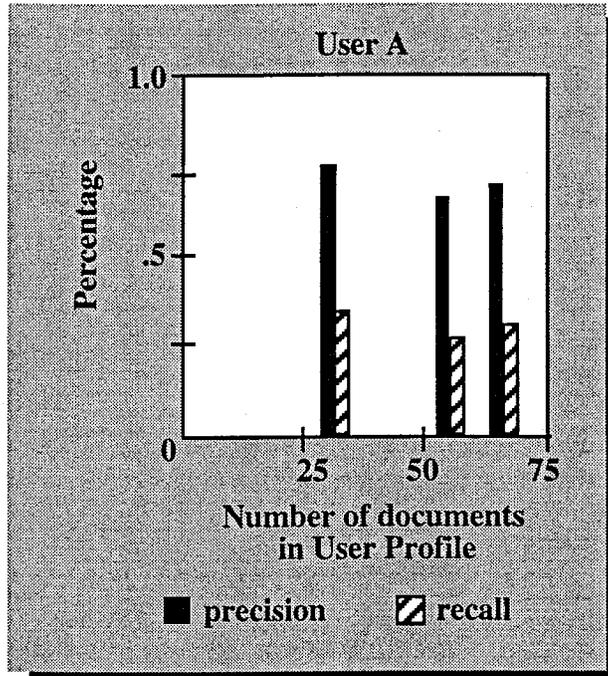


Figure 5: Filter Agent results for Beth.⁴⁵

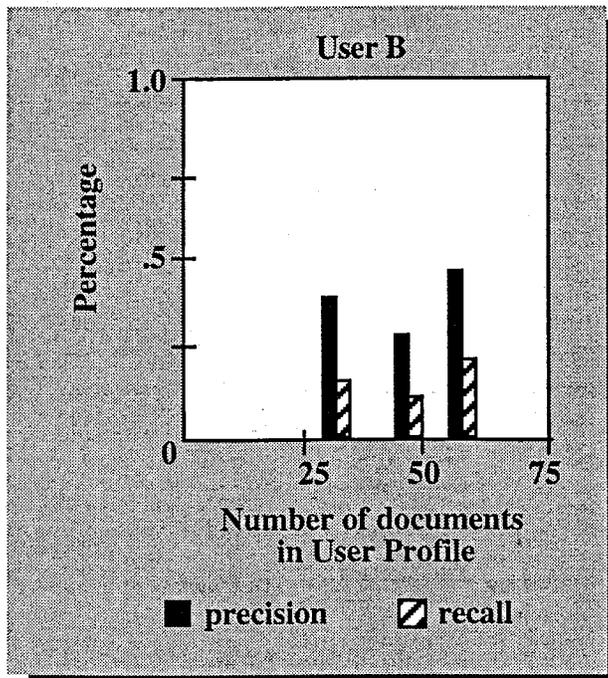


Figure 6: Filter Agent results for Nancy.

⁴⁵See Chapter 4 for a description of each user.

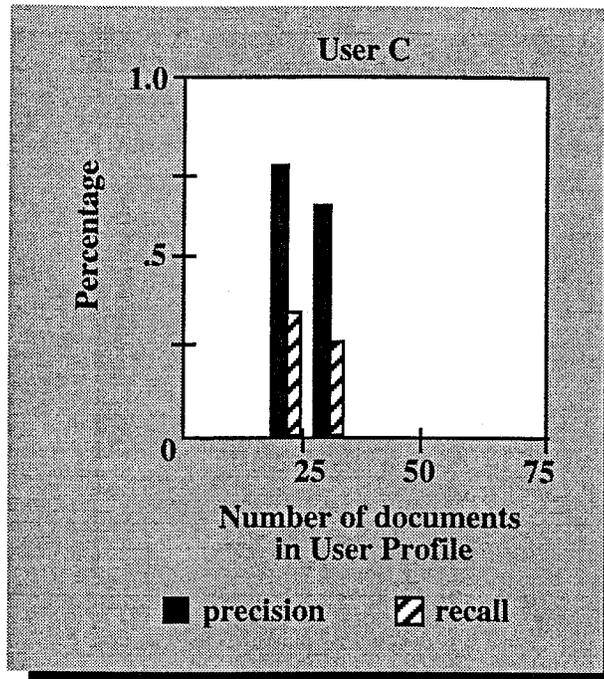


Figure 7: Filter Agent results for Richard.

6. NewsBlast (World Wide Web user interface)

Introduction

The personalized information agent described in the previous chapter makes its results available to, and could accept feedback from, a number of different interfaces. NewsTalk, a speech interface to the agent, has already been discussed. This chapter focuses on NewsBlast, a personalized on-screen newspaper available via the World Wide Web (WWW).⁴⁶

Related work

A number of online newspapers exist, including FishWrap and Crayon. FishWrap enables you to choose the news *topics* you are interested in; Crayon enables you to choose the news *sources* you want.

FishWrap [Chesnais 1993] is an experimental online newspaper available to students, faculty, and staff at the Massachusetts Institute of Technology. The original impetus for the project was to give students access to local news from their hometown. However, the project has accomplished and surpassed this early goal. Currently, FishWrap provides access to news articles, photographs, audio clips, comics, advertisements, and advice columns.

A user can request news from approximately 50 pre-defined topics, including national news, movie reviews, architecture, and cycling. FishWrap includes articles from several sources, including the Associated Press and the Boston Globe.

⁴⁶The World Wide Web [Berners-Lee 1992] is a distributed hypermedia system developed by Tim Berners-Lee and his associates at the European Laboratory for Particle Physics (known as CERN) in Switzerland. It provides computer users with a simplified, consistent method to publish and access a variety of media (text, audio, and video).

A user accesses information on the WWW by using a "web browser" (or application) such as Netscape or Mosaic. A user can click on hyperlinks, i.e., "hot" text or graphics, in order to jump to related information. In the WWW, information is contained in "pages", or windows.

Crayon⁴⁷ [NB 1995], created at Bucknell University, is an index to approximately 100 existing online news sources, including daily newspapers (e.g., San Francisco Chronicle and The Irish Times), magazines (e.g., Sports Illustrated), and electronic news services (e.g., Edupage). A user can request articles from one or more of these sources, and articles from one of several broad categories (including Nation, World, Weather, Entertainment, Comics, and World Wide Web). Crayon has over 10,000 subscribers.

User interface design description

NewsBlast includes many of the same features available in NewsTalk, including personalized news separated into sections and lists, related articles, and user feedback. However, NewsBlast is available on-screen via the World Wide Web. As a result, a user must choose related articles and give feedback explicitly, rather than implicitly as in NewsTalk.

Also, NewsBlast includes a WWW interface to the SMART information retrieval system. Using this interface, a user can type in a natural language query (e.g., "What is Bill Clinton up to today?") and receive articles relevant to that query.

Presenting information

When users traverse information in a graphical user interface, they sometimes become disoriented, i.e., they lose track of where they are, and they are unable to determine where they can go and how they can get back to information they have already seen [Conklin 1987]. NewsBlast attempts to maintain a user's orientation by creating only two pages of information.

The first page is a list of headlines, grouped by section. Just as in NewsTalk, each section contains an editors' list and a user list. Headlines are assigned to the appropriate lists within each section.

⁴⁷Crayon is an acronym for CReAte Your Own Newspaper, and is available on the World Wide Web at <http://sun.bucknell.edu/~boulter/crayon/>.

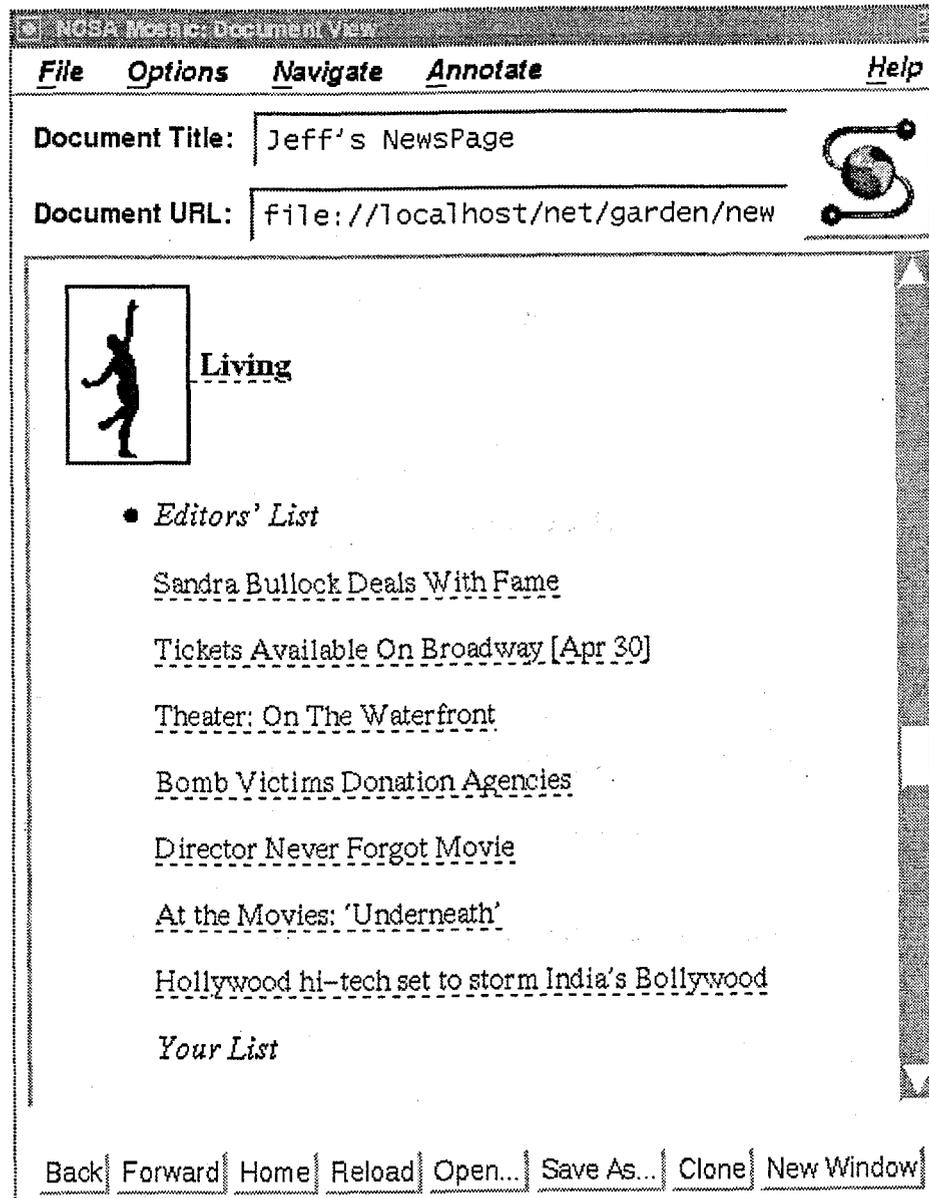


Figure 8: NewsBlast's front page contains a list of headlines divided into sections (in this case, Living) and lists (the editors' list and your list).

When a user clicks a headline, the text of the corresponding article appears in a new page (figure 9). All of the articles are on one page in order to: enable a user to browse easily from one article to another; to reduce the number of windows open on a user's screen; and to take advantage of WWW browsers' built-in text searching capability.

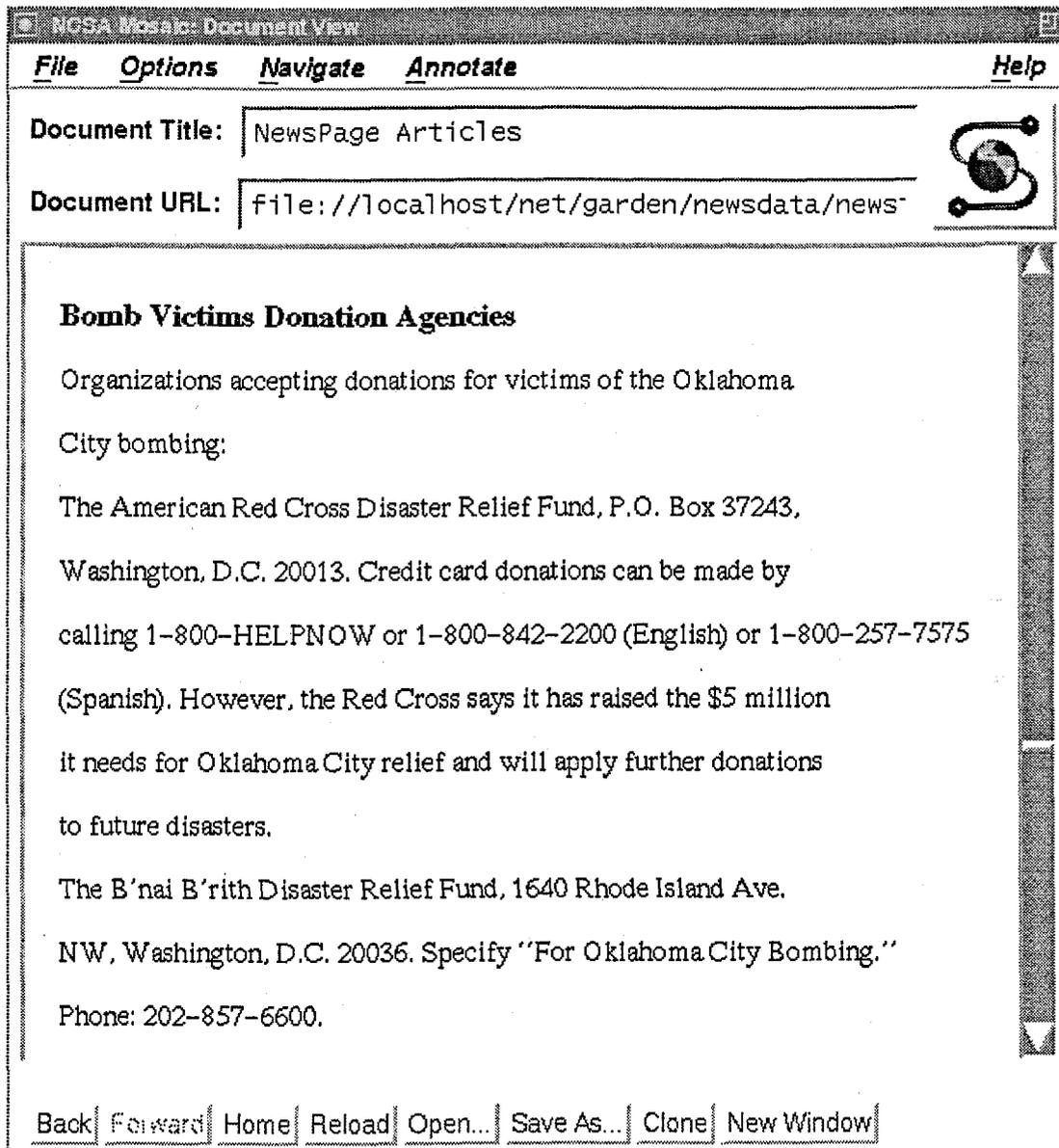


Figure 9: A headline and article in NewsBlast.

In contrast to NewsTalk's implicit assumptions about what articles a user is interested in, in NewsBlast a user would have to take explicit action and click a "Thumbs Up" icon associated with an article in order to add the article to her User Model.⁴⁸

⁴⁸This explicit feedback from NewsBlast to the User Model has not been implemented.

In NewsBlast, the headline of any related articles appear at the end of an article. This enables a user to jump to that related article and to continue the theme of the first article.

Digitized audio

As part of the collection process described in Chapter 5, text and audio from ABC News Radio are included in NewsTalk and NewsBlast. In NewsBlast, a user can click the ABC icon which accompanies an article in order to hear the accompanying audio clip. Typically, the clip is of a newsmaker or a reporter, and it may last anywhere from 5 seconds to 2 minutes.

The screenshot shows a web browser window titled "News Calls". The address bar is empty. The browser's navigation toolbar includes buttons for Back, Forward, Home, Reload, Images, Open, Print, Find, and Stop. A large "N" logo is visible on the right side of the toolbar. The main content area displays two news items, each with a small thumbnail image of a person's face.

Thursday, March 23, 1995, 9am, #1

After Wednesday's shouting and name-calling on the house floor over welfare reform, we asked William Kristol, a leading Republican consultant and one of the architects of the GOP's Contract with America, why the welfare bill caused congressmen to abandon their customary civility.

VERBATIM: *"This is an emotional issue because it's an important issue. The changes the Republicans are proposing are fundamental changes in policies that have been built up over decades. They're real emotions connected to these policies, there's really a lot at stake here. And I think when there's a lot at stake you get an emotional debate."*

Thursday, March 23, 1995, 9am, #2

Kato Kaelin this morning/Thursday returns to the witness stand for more questions from the prosecution. O.J Simpson's defense lawyers have their work cut out for them when they begin Kaelin's cross-examination. The task is expected to go to Robert Shapiro. What kind of a witness does Shapiro have when he faces Kaelin? Here's defense lawyer Gigi Gordon, speaking from Los Angeles on ABC's Good Morning America:

VERBATIM: *"Mr. Kaelin is sort of a guy who is walking through a minefield with both sides shooting at him and it's gonna be a very tough row to hoe. He wants to keep the witness under control and get certain things out of him without having him go off like a time bomb!"*

The browser's status bar at the bottom shows the page number "1/1" and a progress indicator.

Figure 10: The NewsBlast presentation of the ABC text and audio. A user may click the ABC icon in order to hear the audio clip which accompanies the article.

7. Conclusions and Future Directions

Future directions

Overcoming the "bootstrapping" problem

At least twenty users interacted with the system at least twice, and in all cases people had some difficulty getting started with the system: learning the vocabulary, the syntax, the conversation limitations, and the abilities of the interface.⁴⁹ This problem could be eased in a number of ways.

People often learn better how to do a task by doing the task rather than reading about it. Most likely, new users would have an easier time learning to use NewsTalk if they first interacted with a speech tutorial, which introduced the abilities, limitations, vocabulary, and syntax of the system.

More specific queries

Current NewsTalk users are able to ask only very general queries, i.e., queries at the section-level. This design is useful when a user is browsing and has no specific interest in mind, but this restriction causes frustration when a user wants articles regarding a specific topic.

It may be possible to enable users to ask for more specific topics by replacing the limited vocabulary speech recognizer NewsTalk uses now with a large vocabulary speech recognizer such as BYBLOS⁵⁰ from BBN. Such a recognizer would drastically change a user's experience because, instead of being limited to broad search criteria (e.g., "Go to Business") you could ask for very specific topics ("What's new with Apple Computer?").

Such a fundamental change may affect the usability of other parts of the interface, and would need to be tested thoroughly to determine how the interface should change.

⁴⁹[Owen 1986] refers to this as the "bootstrapping" problem.

⁵⁰BYBLOS is also speaker-independent and allows continuous speech. Its syntax and vocabulary can be optimized for a specific domain. For example, BBN has created a version that is optimized for the syntax and vocabulary found in the Wall Street Journal.

More navigation options within a text article or digitized audio file

In the current interface, a user has only two navigation options while listening to an article: she can jump to the end of the article, or she can continue listening to the article. The interface would improve significantly if a user could move forward or back (for example, by one sentence or one paragraph), or if she could jump to the beginning or the end of the article.

The current system relies on significant text processing, but it does not do any processing of the digitized audio. The interface may benefit from a number of techniques that attempt to detect salient points in the audio, using such cues as energy and pauses [Kato 1992; Arons 1994], and speaker indexing [Roy 1995].

Additional features

A number of features could be added to the interface to make it more useful.

For example, a user could ask the interface to send the text and/or audio of an article to her e-mail or that of a colleague, save it to her disk, or send the text of an article to a specific printer or fax machine.

One participant in the long-term user study wanted to customize the interface's grammar so that he could add phrases that were more natural to him. This capability might make it easier for a user to remember what utterances are legal.

Users might find it useful to mark a specific article, and then later in the conversation return to that marked article. [Arons 1994] proposes a bookmark metaphor for this purpose.

Currently there are two parts of an article: the headline and the body. An intermediate level of summarization would be useful as well, so that a user could hear a headline and ask for a summary of the article (rather than, or in addition to, the entire article).

If the interface offers an article, a user may want to ask why the article was chosen (for example, "because it is about Apple Computer"). Further, a user may want to skip to the section of the article relevant to that topic. (You might think of this as audio highlighting.)

You could ask the system to call you if a major article happened on a specific topic—or it may know you so well you don't have to specify this.

You could subscribe to agents that collect information that is biased to a specific point of view, e.g., a conservative or liberal political slant.

In the long-term, it may be possible to query such a system in order to ask questions about an article (as people did in the person-to-person user study). For example, "What did President Clinton have to say about that?"

Conclusions

Many people ask if speech systems are good enough to use. As with any tool, the answer depends on the user's abilities, tasks, and environment (including the acoustic and social situation). Who is using the system, for what purpose, and what else are they attempting accomplish simultaneously? Speech interfaces have the advantage that they can be used while a person's hands and eyes are busy with another task and they take advantage of the physical abilities of people who are blind or visually impaired.

Of course, much work remains in the area of speech interfaces; synthesizers need to be more understandable and pleasant to listen to; speech recognizers need to be more robust, especially in acoustic situations that are less than ideal; interfaces need to be more natural and adhere more closely to the conversational conventions that people are accustomed to from talking one another. This thesis is another step on the path to making speech interfaces a common and useful part of more people's lives.

NewsTalk has been used by several short-term users and three long-term users. Users reacted positively to NewsTalk's spoken conversation metaphor, and expressed a desire to use the interface in many more situations. Users reported, and quantitative results confirm, that NewsTalk and the Filter Agent adapted to the interest of each user, given only implicit feedback from users. Also, users were able to interpret NewsTalk's feedback to detect and correct recognition errors. Finally, as one measure of the success of the interface and of the system as a whole, all of the long-term users asked to continue using NewsTalk beyond the completion of the studies.

References

- Apple 1987 Apple Computer, Inc. *Apple Human Interface Guidelines: The Apple Desktop Interface*. Reading, MA: Addison-Wesley Publishing Company, Inc. 1987.
- Arons 1991a B. Arons. "Hyperspeech: Navigating in Speech-only Hypermedia." In *Proceedings of Hypertext '91*, pp. 133-146. ACM, 1991.
- Arons 1991b B. Arons. "Authoring and Transcription Tools for Speech-Based Hypermedia Systems." In *Proceedings of 1991 Conference*, American Voice I/O Society, Sep. 1991, pp. 15-20.
- Arons 1994 B. Arons. "Interactively Skimming Recorded Speech." Ph.D. thesis, MIT Program in Media Arts and Sciences, Feb. 1994.
- Blazie 1991 Blazie Engineering. "Products for Blind and Visually Impaired People." 1991.
- Belkin 1992 N.J. Belkin and W.B. Croft. "Information Filtering and Information Retrieval: Two Sides of the Same Coin?", In *Communications of the ACM*, Vol. 35, No. 12, pp. 29-38, 1992.
- Berners-Lee 1992 T. Berners-Lee, R. Caillau, J. Groff, and B. Pollerman. *World-Wide Web: The Information Universe, Electronic Networking: Research, Applications, and Policy*. Meckler Publications, 2(1), pp. 52-58, Spring 1992.
- Brennan 1990 S.E. Brennan. "Conversation as Direct Manipulation: An Iconoclastic View." In *The Art of Human-Computer Interface Design*, edited by Brenda Laurel. Addison-Wesley Publishing Company, Inc., Reading, MA. 1990.

- Buckley 1985 C. Buckley. "Implementation of the SMART Information Retrieval System." Technical Report 85-686, Cornell University. 1985.
- Chesnais 1993 P. Chesnais. "The Freshman FishWrap." MIT Media Laboratory. December 13, 1993.
- Chin 1991 D. Chin. "Intelligent interfaces as agents." In *Intelligent User Interfaces*, edited by J. Sullivan and S. Tyler. ACM Press, NY. 1991.
- Clark 1992 H.H. Clark. *Arenas of Language Use*. University of Chicago Press, Chicago, IL. 1992.
- Conklin 1987 J. Conklin. Hypertext: An introduction and survey. *IEEE Comput.* 20, 9, pp. 17-41.
- Deerwester 1990 S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. "Indexing by Latent Semantic Analysis", *Journal of the American Society for Information Science*, 41(6):391-407, 1990.
- Ellis 1990 D. Ellis. *New Horizons in Information Retrieval*. Library Association Publishing Ltd., London. 1990.
- Engelbeck 1989 G. Engelbeck, and T.L. Roberts. "The effects of several voice-menu characteristics on menu-selection performance." *US West Advanced Technologies*. 1989.
- Erickson 1990 T. Erickson. "Working with Interface Metaphors." In *The Art of Human-Computer Interface Design*, edited by Brenda Laurel. Addison-Wesley Publishing Company, Inc., Reading, MA. 1990.
- Ericsson 1984 K.A. Ericsson and H.A. Simon. *Protocol Analysis, The* MIT Press, 1984.
- Gould 1983 J.D. Gould, J. Conti, T. Hovanyecz. "Composing Letters with a Simulated Listening Typewriter." In *Communications of the ACM*, 26, 4, pp. 295-308.

- Grice 1975 H.P. Grice. "Logic and Conversation." (From the William James lectures, Harvard University, 1967.) In *Syntax and Semantics 3: Speech Acts*. P. Cole and J. Morgan, eds. New York: Academic Press, 1975.
- Grosz 1986 B.J. Grosz and C. L. Sidner. "Attention, Intentions, and the Structure of Discourse." In *Computational Linguistics*, 12(3), pp. 175-203. MIT Press, Cambridge, MA. 1986.
- Hayes 1983 P.J. Hayes and D.R. Reddy. "Steps toward graceful interaction in spoken and written man-machine communication." In *International Journal of Man-machine Studies*, 19, 231-284. 1983.
- Holland 1975 J.H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. University of Michigan Press, Ann Arbor, 1975.
- Holtzblatt 1993 K. Holtzblatt and S. Jones. *Contextual Inquiry: A Participatory Technique for System Design*. In *Participatory Design: Principles and Practice*, Aki Namioka and Doug Schuler (Eds.). Lawrence Earlbaum Pub., Hillsdale, NJ. 1993.
- Horner 1993 C. Horner. "NewsTime." SM thesis, Department of Media Arts and Sciences, Massachusetts Institute of Technology, May 1993.
- Individual 1995 Individual, Inc. 8 New England Executive Park West, Burlington, MA, 01803. 1995.
- Irish Times 1994 The Irish Times. p. 2, August 28, 1994
- Kamm 1994 C. Kamm. "User Interfaces for Voice Applications." In *Voice Communications Between Humans and Machines*. National Academy Press, Washington, DC. 1994.
- Kato 1992 Y. Kato and K. Hosoya. "Fast Message Searching Method for Voice Mail Service and Voice Bulletin Board Service."

- In *Proceedings of 1992 Conference*, American Voice I/O Society, pp. 215-222. 1992.
- Kay 1984 "Computer software", in *Scientific American*. Vol. 251, no. 3, pp. 52-59. September 1984.
- Kozierok 1993 R. Kozierok. A Learning Approach to Knowledge Acquisition for Intelligent Interface Agents. SM thesis, Department of Media Arts and Sciences, Massachusetts Institute of Technology, 1993.
- Lakoff 1980 G. Lakoff and M. Johnson. *Metaphors We Live By*. The University of Chicago Press, Chicago. 1980.
- Laurel 1991 B. Laurel. *Computers as Theater*. Addison-Wesley. 1991.
- Laurel 1995 B. Laurel. "Secret Agent Manifestations." In *I.D.: The International Design Magazine*, March-April 1995, p. 60.
- Levelt 1982 W.J.M. Levelt and S. Kelter. "Surface Form and Memory in Question Answering." In *Cognitive Psychology*, Vol. 14, pp. 78-106. 1982.
- Lewis 1986 C. Lewis and D.A. Norman. "Designing for Error." In *User Centered System Design*, edited by D.A. Norman and S.W. Draper. Lawrence Erlbaum Associates, Hillsdale, NJ. 1986.
- Ly 1993 E. Ly. "Chatter: a conversational telephone agent." SM thesis, Department of Media Arts and Sciences, Massachusetts Institute of Technology, May 1993.
- Maes 1993 P. Maes and R. Kozierok. "Learning Interface Agents." In *Proceedings of AAAI*, 1993.
- Maes 1994 P. Maes. "Agents that reduce work overload and information overload." In *Communications of the ACM*, 37 (7), pp. 31-40, 146. 1994.

- Malone 1983 T.W. Malone. "How do people organize their desks: Implications for designing office automation systems." In *ACM Transactions on Office Information Systems*, 1, pp. 99-112. 1983.
- Martin 1989 G.L. Martin. "The utility of speech input in user-computer interfaces." In *International Journal of Man-Machine Studies*, 30:355-375, 1989.
- Marx 1995 M. Marx. Personal communication. April 5, 1995.
- Mountford 1990 S.J. Mountford. "Tools and Techniques for Creative Design." In *The Art of Human-Computer Interface Design*, edited by Brenda Laurel. Addison-Wesley Publishing Company, Inc., Reading, MA. 1990.
- Muller 1990 M.J. Muller, and J.E. Daniel. "Toward a definition of voice documents." In *Proceedings of COIS '90*, pp. 174-182. ACM, 1990.
- Mynatt 1994 E. Mynatt and G. Weber. "Nonvisual Presentation of Graphical User Interfaces: Contrasting Two Approaches." In *CHI '94 Conference Proceedings*, ACM, New York. 1994.
- NB 1995 Nation's Business. "Create Your Own Newspaper with Crayon." March 29, 1995.
- Norman 1986a User Centered System Design, edited by D.A. Norman and S. W. Draper. Lawrence Erlbaum Associates, Hillsdale, NJ. 1986.
- Norman 1986b D.A. Norman. "Cognitive Engineering." In User Centered System Design, edited by D.A. Norman and S. W. Draper. Lawrence Erlbaum Associates, Hillsdale, NJ. 1986.
- Norman 1988 D. Norman. *The Psychology of Everyday Things*. Basic Books, Inc., Publishers, New York. 1988.

- Norman 1992 D. Norman. Turn Signals Are The Facial Expressions of Automobiles. Addison-Wesley. 1992
- Nye 1982 J.M. Nye, "Human factors analysis of speech recognition systems." In *Speech Technology*, I, p. 50-57. 1982.
- Oddy 1977a R.N. Oddy, "Information retrieval through man-machine dialog", In *Journal of Documentation*, 33, 1977, pp. 1-14.
- Oddy 1977b R.N. Oddy, "Retrieving references by dialog rather than by query formulation", In *Journal of informatics*, 1, 1977, pp. 37-53.
- Oren 1990 T. Oren, G. Salomon, K. Kreitman, and A. Don. "Guides: Characterizing the Interface." In *The Art of Human-Computer Interface Design*, edited by Brenda Laurel. Addison-Wesley Publishing Company, Inc., Reading, MA. 1990.
- Owen 1986 D. Owen. "Naive Theories of Computation." In *User Centered System Design*, edited by D.A. Norman and S.W. Draper. Lawrence Erlbaum Associates, Hillsdale, NJ. 1986.
- Reichman 1986 Rachel Reichman. "Communication Paradigms for a Window System." In *User Centered System Design*, edited by D.A. Norman and S.W. Draper. Lawrence Erlbaum Associates, Hillsdale, NJ. 1986.
- Richards 1984 M.A. Richards and K.M. Underwood. "How Should People and Computers Speak to One Another?" In *Proceedings of Interact '84*. pp. 33-36. International Federation for Information Processing, London. 1984.
- Roy 1995 D. Roy. "NewsComm: A Hand-held Device for Interactive Access to Structured Audio." SM thesis, Department of Media Arts and Sciences, Massachusetts Institute of Technology. Expected May 1995.

- Rudnicky 1991 A. I. Rudnicky, and A. G. Hauptmann. "Models for evaluating interaction protocols in speech recognition." In *Proceedings of CHI '91*, pp. 285-291.
- Rudnicky 1994 A. I. Rudnicky, and A. G. Hauptmann and K. Lee. "Survey of Current Speech Technology." In *Communications of the ACM*, 37 (3), pp. 52-57. 1994.
- Sacks 1974 H. Sacks, E.A. Schegloff, and G. Jefferson. "A simplest systematics for the organization of turn-taking for conversation. In *Language*, 50 (4) : pp. 696-735, 1974. Reprinted in *Studies in the Organization of Conversational Interaction*, edited by J. Schenken, Academic Press, 1978.
- Salton 1981 G. Salton. "The SMART environment for retrieval system evaluation—advantages and problem areas". In *Information retrieval experiment*, K. Sparck Jones, editor. London, Butterworth, 1981.
- Salton 1983 G. Salton and M. McGill. Introduction to Modern Information Retrieval. McGraw-Hill, New York. 1983.
- Salton 1989 G. Salton. Automatic Text Processing. Addison-Wesley, Reading, MA. 1989.
- Schalk 1992 T. Schalk. "Guest Editor's Introduction." In *Journal of the American Voice I/O Society*, volume 11, February 1992.
- Scharf 1994 I. Scharf. "A Language for Interactive Speech Dialog Specification." SM thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, May 1994.
- Schegloff 1973 E.A. Schegloff and H. Sacks. "Opening up closings." In *Semiotica*, Vol. 8, pp. 289-327. 1973.
- Schmandt 1985 C. Schmandt, B. Arons, and C. Simmons. "Voice Interaction in an Integrated Office and Telecommunications Environment." In *Proceedings of the 1985 American Voice*

I/O Society Conference, pp. 51-57. San Jose, CA.
September 1985.

- Schmandt 1986 C. Schmandt and B. Arons. "A Robust Parser and Dialog Generator for a Conversational Office System." In *Proceedings of the 1985 American Voice I/O Society Conference*, pp. 355-365. San Jose, CA. September 1986.
- Schmandt 1994 C. Schmandt. *Voice Communication with Computers*. Van Nostrand Reinhold: New York. 1994.
- Schneiderman 1987 B. Schneiderman. *Designing the User Interface: strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley Publishing Company, Inc. 1987.
- Schneiderman 1995 B. Schneiderman. "Perspectives: Looking for the Bright Side of User Interface Agents." In *Interactions: New visions of Human-Computer Interaction*. January 1995.
- Sheth 1993 B. Sheth and P. Maes. "Evolving agents for personalized information filtering." In *Proceedings of the Ninth Conference on Artificial Intelligence for Applications*. IEEE Computer Society Press, 1993.
- Sheth 1994 B. Sheth. "A Learning Approach to Personalized Information Filtering." SM thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Jan. 1994.
- Sparks 1994 R. Sparks, L. Meiskey, and H. Brunner. An Object-oriented Approach to Dialogue Management in Spoken Language Systems. In *Proceedings of CHI '94*, pp. 211-217. ACM, New York. 1994.
- Stalnaker 1978 R.C. Stalnaker. "Assertion." In *Syntax and Semantics: Pragmatics 9*, edited by P. Cole. Academic Press, New York. 1978.

- Stanfill 1986 C. Stanfill and D. Waltz. "Toward Memory-based Reasoning." In *Communications of the ACM*, December 1986.
- Stifelman 1992 L.J. Stifelman. "VoiceNotes: An Application for a Voice-Controlled Hand-Held Computer." SM thesis, Department of Media Arts and Sciences, Massachusetts Institute of Technology, May 1992.
- Stifelman 1993 L.J. Stifelman, B. Arons, C. Schmandt, E.A. Hulteen. "VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker." In *Proceedings of INTERCHI '93*, ACM, New York. 1993.
- Wheatley 1992 B. Wheatley, J. Tadlock, and C. Hemphill. Automatic efficiency improvements in telecommunications application grammars. First IEEE Workshop on Interactive Voice Technology for Telecommunications Applications. 1992.
- Wildfire 1994 Wildfire Communications, Inc. Lexington, MA. 1994.
- WGBH 1993 Media Access Research and Development Office, WGBH Educational Foundation. The Print Access Project: an investigation into making daily newspapers accessible to print-disabled people. Boston, MA. 1993.
- Yankelovich 1994a N. Yankelovich and E. Baatz. "SpeechActs: A Framework for Building Speech Applications." In *Proceedings of the American Voice Input/Output Society*, San Jose. 1994.
- Yankelovich 1994b N. Yankelovich. "Talking vs. Taking: Speech Access to Remote Computers." In *Proceedings of CHI '94*, ACM, New York. 1994.
- Yankelovich 1995a N. Yankelovich, G. Levow, and M. Marx. "Designing SpeechActs: Issues in Speech User Interfaces." In *Proceedings of CHI '95*. ACM, New York. 1995.
- Yankelovich 1995b N. Yankelovich. Personal communication. May 1995.

Appendix 1: NewsTalk Grammar

Below is the context-free grammar used in NewsTalk. Terminals are in quotes. The user names have been crossed out to protect their privacy. See [Wheatley 1992] for information on the format of this grammar.

```
start(Root).
Root ---> nameG.
Root ---> sectionG.
Root ---> goToSectionG.
Root ---> goToLongSectionG.
Root ---> listG.
Root ---> goToListG.
Root ---> noG.
Root ---> yesG.
Root ---> goToSleepG.
Root ---> changeSpeedG.
Root ---> repeatG.
Root ---> goBackG.
Root ---> signOffG.

export(nameG).
nameG ---> XXX XXX.
nameG ---> Jeff Herman.
nameG ---> XXX XXX.

export(sectionG).
sectionG ---> A B C.
sectionG ---> the B B C.
sectionG ---> N P R.
sectionG ---> Business.
sectionG ---> International.
sectionG ---> Living.
sectionG ---> Local.
sectionG ---> National.
sectionG ---> Sports.
```

sectionG ---> Technology.
sectionG ---> Top Stories.

export(goToSectionG).
goToSectionG ---> go to sectionG.

export(goToLongSectionG).
goToLongSectionG ---> go to the sectionG section.

export(listG).
listG ---> my list.
listG ---> the editors' list.

export(goToListG).
goToListG ---> go to listG.

export(noG).
noG ---> skip it.
noG ---> no.

export(yesG).
yesG ---> read it.
yesG ---> yes.

export(goToSleepG).
goToSleepG ---> stop talking.
goToSleepG ---> stop listening.

export(changeSpeedG).
changeSpeedG ---> talk very fast.
changeSpeedG ---> talk fast.
changeSpeedG ---> talk at a normal speed.

export(repeatG).
repeatG ---> repeat that.

export(goBackG).
goBackG ---> go back.

export(signOffG).

signOffG ---> that's it for now.

signOffG ---> good bye.

signOffG ---> hang up.

Appendix 2: User Study 1

The goals of this initial study were to determine how two people would communicate over the telephone in order to: request articles on specific topics; summarize articles; reject or accept suggested news articles; learn about and act on each other's news interests.

Methodology

Each participant was randomly paired with one other participant (there were four pairs in all). None of the participants knew anything about their partner. In each pair of participants, one was the "manager", or news requester: her goal was to retrieve the news of interest to her. The other participant was the "assistant", or news provider: her goal was to provide the news of interest to the manager.

Each pair of participants had two telephone conversations. Each conversation began when the investigator set up a conference call between himself and the two participants. Each conversation lasted approximately 15 minutes.

Findings

Specification of interests

Typically, conversations began by the manager listing the types of articles she was interested in hearing. She often gave examples to clarify what she meant by specific topics:

Manager: I'm interested in stories about law enforcement. By that I mean things local police departments are doing, new procedures for dealing with domestic violence. Just sort of trends that have to do with that.

Manager: I also would like significant national political stories. For example, if Bill Clinton is going to come to Massachusetts and campaign for Teddy Kennedy, that would be a major political story.

Asking for specific topics more than general sections

Managers asked for specific topics (e.g., the O.J. Simpson case, World Cup soccer) much more often than news by section (e.g., international news, sports).

Assistants summarized stories, frequently leading to follow-up questions from the manager

Assistants almost always communicated a story by summarizing it and speaking key phrases, rather than another strategy such as starting at the beginning of the story and reading until the manager said she had heard enough.

Since the summary didn't always cover all of the topics the manager was interested, she frequently asked the assistant follow-up questions to fill in any missing pieces.

Assistants filled in the search time

While the assistants were searching for an answer to a follow-up question, they would often say whatever text they were scanning at that moment, perhaps for two reasons: (1) to give the manager additional information as they searched for the answer to the specific question, and (2) to let the manager know they were still searching.

Occasionally this practice led to the initial question not being answered, perhaps because conversants were flooded with other information and forgot the original question.

Assistants explained their article choices

When an assistants selected an article, she would often begin by saying why she chose it. For example:

Assistant: Oh, here's a story about TCI possibly merging with Bell Atlantic—you said you were interested in business mergers.

Assistants specified the source of a story

Assistants frequently specified what source they were using, either at the beginning of a conversation or when they switched to a new source:

Manager: OK, I'd like to know more about World Cup soccer and the results of that final game between Brazil and Italy.

Assistant: I'm going to get that for you out of the Boston Globe.

Results of debriefing

After each participant had finished the study, I asked them two open-ended questions about the experience:

- I asked the managers what they found useful and what they found frustrating with the service provided.
- I asked the assistants what they found easy and what they found difficult in providing the service.

Because the questions were so open-ended, it's interesting how uniform the responses were.

Managers' hesitancy to direct assistants

Despite being told at the beginning of the experiment that they were in charge of the conversations, three managers said that they didn't direct the assistant in all cases they wanted to:

- "Sometimes [my assistant] would tell me things I didn't want to know, but I wanted to be polite and let her go on."
- "It was hard to say I didn't want to hear something because it's a person and you don't want to shut them off."
- "He wanted to give me all of the details without letting me make a decision about it first. Because it was a human being, I didn't want to interrupt."

Serendipity

After they had answered all other question, I explained to the managers that I was creating a computerized, personalized news service. Two of the four managers were concerned that such service would present only articles that related to topics they had specified, rather than allowing them to run across topics they hadn't thought to specify.

Instructions to the managers

At the beginning of the first user study, the investigator read these instructions only to the participants assigned to the role of manager.

Your scenario

You've just hired a new assistant. Your assistant's job is to provide you with the daily news most useful to you.

Your task is to talk to your assistant via telephone in order for you to receive the news of most interest to you. You're in charge, so don't hesitate to direct your assistant to news of interest to you, or to reject suggestions that aren't helpful.

Starting the conversation

I'll call you at the agreed upon times and set up a conference call between you and your partner.

Ending the conversation

The conversation ends when any of the following occurs:

- 1) as manager, you decide you have received enough news;
- 2) 15 minutes has passed;
- 3) either you or your partner has to attend to something else and can't continue the conversation.

Restrictions

All of your conversations with your partner should be restricted to news.

Comments, questions, or concerns

Please feel free to call me at any time at my office (617-253-2245) if you have any comments, questions, or concerns about this study.

Instructions to the assistants

At the beginning of the first user study, the investigator read these instructions only to the participants assigned to the role of assistant.

Your scenario

You've just been hired as an assistant. Your job is to provide your new manager with the news most useful to her.

Your task is to talk to your manager via telephone in order to provide her with the news she is most interested in. The manager is in charge, so it's likely she will direct you to news of interest to her, and she will accept or reject suggestions you make.

Starting the conversation

I'll call you at the agreed upon times and set up a conference call between you and your partner.

Ending the conversation

The conversation ends when any of the following occurs:

- 1) the manager decides that she has received enough news;
- 2) 15 minutes has passed;
- 3) either you or your partner has to attend to something else and can't continue the conversation.

Restrictions

All of your conversations with your partner should be restricted to news.

Comments, questions, or concerns

Please feel free to call me at any time at my office (617-253-2245) if you have any comments, questions, or concerns about this study.

Appendix 3: Materials for User Study 2

Participants in User Study 3 read these instructions before using NewsTalk.

I'm testing NewsTalk, not you. If you have trouble using NewsTalk that's evidence that the design is flawed in some way, and the purpose of these studies is to uncover those flaws.

You can stop any of the conversations at any time, and you can choose not to participate in any more experiments at any time.

You should talk to NewsTalk using the same volume and pacing you typically use when talking to someone over the telephone. Note that you don't have to pause between words, and you can't talk to it while it is talking to you.

NewsTalk has divided the news into 8 sections: Top stories, National, International, Local, Living, Business, Technology, and Sports. Within each of these sections, NewsTalk has divided the articles into two lists: the editors' list and your list. The articles in the editors' list are the top stories for today according to news editors. The articles in your list are the top stories selected by NewsTalk specifically for you.

After you identify yourself, you can go to any section at any time, and you can go to any list within the current section at any time.

After you go to a section, NewsTalk will suggest a headline. If you want it to begin reading the article, you can say "yes" or "read it." If you want it to suggest the next headline, you can say "no" or "skip it."

To interrupt NewsTalk when it is reading an article, press any of the touch tones on your telephone.

If you want NewsTalk to repeat its most recent sentence, say "repeat that."

If you want NewsTalk to stop talking, say "stop talking." NewsTalk will stop talking until you press any of the touch tone keys.

If you want NewsTalk to go back to its most recent statement, say "go back." (For example, this is useful if it does not recognize your name correctly.)

If you want NewsTalk to talk faster or slower, you can say "talk very fast", "talk fast", or "talk slowly."

When you are done and you want to hang up, you can say “good-bye” or “hang up.”

Appendix 4: Biographies of the committee

Chris Schmandt

Chris Schmandt received the BS and MS degrees from MIT, where he has been building speech systems since 1979. He is director of the Speech Research Group at the Media Laboratory, a position he has held since the creation of the Lab. Before that he worked on speech applications research at the Architecture Machine Group, including "Put That There" and "Phone Slave" projects, as well as projects in digital video typography and gestural input for stereoscopic video displays.

His current research focuses on user interfaces and applications of speech processing technology, voice as a data type on workstations and hand-held computers, and computer-mediated telephony. Key to this work is gaining a better understanding of how people use speech for communication in a conversational context, and how to apply this knowledge to more effective voice interaction with computers.

Pattie Maes

Pattie Maes obtained a Ph.D. in Computer Science from the University of Brussels in 1987. She worked as a Visiting Professor and Research Scientist at the MIT AI Laboratory from 1989 to 1991. Since then she has been a faculty member at the MIT Media Laboratory, where she performs research on autonomous intelligent agents, in particular focusing on Human-Computer Interaction (so-called "Interface Agents") as well as applications in Interactive Training and Entertainment.

Candy Sidner

Candy Sidner has been a member of the research staff at Lotus Development Corp. since 1993. Her research interests are in artificial intelligence, especially in natural language processing, collaboration and negotiation, and planning and acting, and in human interfaces, especially using models of conversation and users. Her current research concerns the enhancement of user interfaces with a model of collaboration and discourse processing.

Candy was a member of the research staff at Cambridge Research Lab, Digital Equipment Corp. and a Division Scientist at Bolt Beranek and Newman. She has served on numerous program committees and journal boards for societies in artificial intelligence, and natural language processing. She is past president of the Association of Computational Linguistics (1989), and is a Fellow and past Councilor of the American

Association of Artificial Intelligence. She received her Ph.D. from MIT in Computer Science in 1979.