Future of Speech and Audio in the Interface
A CHI'94 Workshop

Barry Arons and Elizabeth Mynatt

Speech and audio are rapidly becoming integrated into our daily computing environments. While the traditional "terminal beep" still exists, it is being supplemented with a variety of other speech and non-speech sounds, as well as speech input to control applications and user interfaces. This workshop explored current and future applications, research areas, and interaction techniques that use audio in the user interface. The emphasis of the meeting was on a user interface, or "CHI perspective", of using speech and sound in interactive contexts where the audio channel can be exploited for the user's benefit.

The workshop was in many ways a follow-on to the half-day "Sound-Related Computation" workshop held at ACM Multimedia 93 (organized by Carla Scaletti, the founder of the ACM SIGSound e-mail forum). The CHI 94 workshop brought together members and ideas from the CHI community to further define the emerging area of sound in user interfaces and applications. There was much interest in the workshop and we were forced to limit the number of attendees, and permit only one participant on multi-authored papers.

Participants

Mike Albers	Center for Human-Machine Systems Research, Georgia Institute of Technology
Barry Arons	Speech Interaction Research
Peter Astheimer	Fraunhofer-Institute for Computer Graphics (IGD)
David Burgess	Graphics, Visualization and Usability Center, Georgia Institute of Technology
Jonathan Cohen	Interval Research Corporation
William N. Creager	William Creager Consulting
Stephanie S. Everett	Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory
Bill Gaver	Rank Xerox EuroPARC
Nick Haddock	Hewlett-Packard Laboratories
Debby Hindus	Interval Research Corporation
Demetrios Karis	GTE Laboratories Incorporated
Craig McQueen	Department of Computer Science, University of Toronto
Elizabeth Mynatt	Graphics, Visualization and Usability Center, Georgia Institute of Technology
David E. Owen	AirTight Garage
Albert Papp	Lawrence Livermore National Laboratory
Alan Schell	Ciba Corning Diagnostics Corp.
Lisa J. Stifelman	Speech Research Group, MIT Media Lab
David Thiel	Microsoft Corp.
Govert de Vries	Philips Corporate Design
Steve Whittaker	Lotus Development Corporation
Catherine G. Wolf	IBM Thomas J. Watson Research Center
Nicole Yankelovich	Sun Microsystems Laboratories, Inc.

Format

Based on the position papers submitted by workshop attendees, four primary research areas emerged, and the workshop was organized around these themes:

This report appeared in the SIGCHI Bulletin (October 1994, Vol 26, No. 4, pages 44–48).

• Non-speech audio interfaces
• Speech interfaces
• Combined speech and non-speech audio interfaces
• Spatial displays, virtual and collaborative environments, etc.

Two participants were selected from each of these areas to briefly present their work to stimulate conversation. To maximize interactive discussions and information interchange these presentations were strictly limited to ten minutes (only 80 minutes of prepared presentations during the one and a half days of the workshop). To encourage active participation, attendees were split into two groups for most of the discussions. An attempt was made to balance the groups based on backgrounds, interests and experience, and to distribute the presenters. Halfway through the first day, the groups were shuffled to allow the participants to spend time with as many new people as possible. On the second day of the workshop, the discussion groups were self organizing.

1. Non-Speech Audio Interfaces

Bill Gaver spoke on "The Time is Now Because the Periphery is Central" in which he proposed that focusing on issues of ambient audio and peripheral awareness are critical to future user interfaces and applications. Gaver summarized much previous work (e.g., Gaver 89, Gaver 93) to show how providing non-speech cues was important for supporting direct interaction in interactive interfaces. He pointed to "peripheral computing" where a primary goal of the interface is to provide peripheral awareness of people and events. He suggested that audio designers focus on applications such as virtual augmented reality, "bungee-jumping on the web", and remote communication as domains which leverage the strengths of audio interfaces.

Craig McQueen presented "Sound as a Continuous Background Mechanism for Guiding Behavior" a sonification technique (see Kramer 94) to provide feedback when using pen input if the visual channel is overloaded or not available. Several mappings from location to sound were tried (e.g., mapping x-y position to pitch-loudness or to pitch-timbre). McQueen shared the results from testing his technique (called "AudioStrokes") with the manipulation of pie menus. He discussed possible applications of his work, including teaching handwriting skills of handheld computers.

We discussed a variety of issues relating to the use of non-speech audio in human-computer interfaces.

One frustration was that although we now know a variety of techniques (sonification, earcons, auditory icons) to create auditory interfaces, it is still difficult to design good sounds. A common complaint is that sounds in computer interfaces are annoying (see Berglund 94) and distracting. We discussed strategies for making auditory cues less urgent, such as playing slowly changing sounds at low volumes, as well as general techniques for controlling the aesthetics of auditory interfaces.

A more central concern was how to effectively convey information with non-speech auditory cues. Sounds can be interpreted at several levels. For example footsteps or door knocks can be used as auditory icons in a variety of ways. Yet there can be a deeper more expressive level to sounds as well; different footsteps or knocks can subtly indicate size, importance, and a host of other feelings. Current user interfaces have not yet addressed this deeper expressive level in their use of sounds. Perhaps this is partly attributable to the fact that people that create this kind of sounds for movies are called Foley "artists" not "scientists" or "engineers". Current audio interfaces are just beginning to get basic ideas conveyed through sound, and have not yet addressed the harder issues of expressiveness or emotion.

The culture of television and the movies have partially defined the sounds that we know and recognize. One common piece of advice was to involve sound designers or musicians in the design process (see Kaye 92), just as one would consult with graphical designers when creating a new screen-based interface. Audio designers from various domains (film, video games, music) all

have design knowledge which may be applicable to auditory interface design. Understanding the contributions and limitations of these design domains is important, albeit difficult, due to the different vocabularies and overall design goals in each domain.

2. Speech Interfaces

Steve Whittaker spoke about rethinking the how and why of audio applications and the use of "speech as data". Conversational speech is pervasive and critical in the workplace, but we don't currently capture it in our computers. Whittaker discuss the importance of recording, accessing and manipulating recorded speech data even without the power of full speech recognition. Filochat (Whittaker 94) provides user centered indexing for random access to conversations by co-indexing digital notes and a speech recording.

William Creager presented "Simulated Conversations: Speech as an Educational Tool" on the use of speech and audio in multimedia educational contexts. He noted that the CD authoring process was relatively expensive, and that the content must therefore be dead or unchanging—that is why there are so many educational CDs on dinosaurs. He showed portions of an educational CD that contained a variety of spoken interviews. He discussed the attractiveness of speech interfaces for creating a "mental dialogue" between the student and the education material being presented. Creager argued that both the pace and story-telling flow of the speech are more engaging than the stereotypical visual presentations in educational software.

Is it possible to do "audio highlighting" of natural or synthetic speech? It may be feasible to use background sounds merged with the speech or some form of audio enhancement (e.g., MIDI "vocal harmonizers" or "aural enhancers", see Cohen 91), although changing the pitch and timing of speech may conflict with prosodic cues already used for emphasis and other natural meanings. This conflict is analogous to using bold or italics in text for emphasis, then overloading the these features to indicate links in a hypertext system.

Participants voiced concern over issues of privacy and the social implications of recording conversations such as in Filochat. Some people thought that only those people that participated in a conversation should be able to access the recorded speech information. There was also concern that using a Filochat-like system may change the way people take notes at meetings, thus detracting from the inherent naturalness of the note taking task.

It was felt that we need better tools and techniques to navigate, index, and find structure in recordings. Along with pen (Whittaker 94) and graphical (Hindus 93) access to recordings, there is a strong desire to be able to quickly browse or skim the recorded material (see Arons 93a, Arons 94).

3. Combined Speech and Non-Speech Audio Interfaces

Lisa Stifelman presented "Conversation Interfaces Integrating Speech and Sound". She reviewed VoiceNotes, a speech interface for categorizing and accessing small segments of digitized speech that provides both speech and non-speech audio feedback (Stifelman 93). She also described "Conversational VoiceNotes", a system that uses continuous speech recognition input to enable new ways of navigating, organizing, and accessing notes. The system uses a context free grammar to specify speech and non-speech audio feedback to the user, and incorporates techniques to allow the user to repair speech recognition errors (such as through editing expressions, e.g., "No I said..."). In subsequent discussions Stifelman recommended a variety of work in the area of correcting spoken utterances, including self repair and the use of editing expressions (Levelt 83, Levelt 91, Cutler 83, Shriberg 92).

Nicole Yankelovich presented "SpeechActs and the Design of Speech Interfaces", a framework for creating speech applications with a set of tools for building speech user interfaces (Yankelovich 94). Audio tapes of the system were played, including some amusing examples of speech recognition errors, and problems reading e-mail with a text-to-speech synthesizer.

When is it best, or most appropriate, to use a speech-based system? Speech is best when detailed textual information needs to be transmitted. Speech, however, is often difficult to use in public or social environments (see Schmandt 94). When is it best, or most appropriate, to use a non-speech audio system? Non-speech audio is best when something needs to be transmitted with speed or urgency, or for the continuous monitoring of background information (Cohen 92a). Also note that some people prefer speech feedback and some people prefer non-speech feedback.

Some participants felt that users need to be able to speak to computers in the same way that they speak to people, keeping in mind that we speak differently to different people. A design principle to keep in mind is that a conversational system should not use any output words that cannot subsequently be spoken by the user for input.

There are often conflicting requirements in designing speech systems in an attempt to reduce speech recognition errors. In some instances it is best to get the user to speak as little as possible to reduce errors, yet some systems work best with long utterances.

The ability to interactively correct, or repair, speech recognition errors is important. For example, if a user speaks “Schedule a meeting with Barry on Tuesday”, but the speech recognition system gets “Thursday”, it is better for the user to be able to say something like “No, I said Tuesday” than to be required to repeat the whole utterance (see Schmandt 86, Schmandt 94). It is often best for a system to assume that the recognition was correct, and provide enough feedback so that the user can fix anything that is broken.

Designers should consider integrating training into applications, to both train the speech recognition system as well as the user on what is permissible to say, and to allow the user time to get comfortable with the system. Consider changing or shortening system feedback over time; as a user gets accustomed to a system, it may be desirable to provide different levels of speech or auditory feedback (Stifelman 93).

4. Spatial Displays, Virtual and Collaborative Environments, etc.

David Burgess presented a variety of “Open Issues in Spatial Audio”. Topics included: cost considerations (it should be possible to compute 3-D audio on a $10 digital signal processing chip, see Burgess 92), intellectual property issues (there are only a few good “Head Related Transfer Function” models, HRTFs, see Wenzel 92, and they are guarded closely), modeling of the environment is needed (Astheimer 93, Pompetzki 94) to get auditory images to appear outside the head (Durlach 92), the burden of wearing headphones, and the overall effectiveness of spatial audio systems is not well understood.

Stephanie Everett spoke on using speech recognition and natural language understanding in human-computer interfaces with an emphasis on virtual environments (Wauchope 94). A natural language interface was added to an existing graphical interface to a naval command and control system. This multimodal interface allows both database queries and imperatives to be specified through speech or text. The system has a focus stack (Grosz 86) and supports anaphora (pronouns) and ellipsis (sentence fragments). Note that in such a system speech and graphics can compensate for one another (Cohen 92b).

Are HRTFs really needed, and is it possible to synthesize a spatial audio environment with loudspeakers? It may be possible to go pretty far without requiring a full spatial audio system, for example, simple left-to-right panning may be sufficient in some applications. For video conferencing, the addition of strong visual cues help in localizing sounds (Cohen 91, Cohen 93, Sellen 92a, Sellen 92b). Besides spatial location, consider using other auditory streaming principles (Bregman 90) to encode audio that is presented to a user.

Using speech in combination with immersive virtual environments that use goggles or a boom-mounted display makes sense since it is easy to use microphones and headphones in conjunction with the visual display hardware.

Group Discussion: Future Research in Speech-Related Interfaces

We discussed the idea of general purpose speech recognizers versus special purpose recognizers (see also Roe 93 on the future of speech recognition). One area of work is in hybrid speech recognizers that combine the features of a large vocabulary, speaker independent, continuous speech recognition system with the ability to add new vocabulary items on-the-fly as is often found in speaker dependant, isolated word system. This could, for example, be accomplished by using multiple speech recognizers, or modifying the underlying recognition software.

Tools and toolkits are needed to assist developers. For example, in continuous speech recognition systems, there should be standard tools for implementing commonly used subgrammars such as plug in modules for recognizing dates, times, and numbers. Recognizers should work cooperatively with applications rather than as a black box that spits out a sentence. An application may have semantic knowledge that could aid in the recognition process, but this may be difficult to embed into a grammar.

Common software interfaces are needed for both recognition and text-to-speech systems. Speech synthesis needs improvements in naturalness, prosody, and emotion. Text-to-speech systems need software "hooks" so that it is easy to add new rules for preprocessing (e.g., for speaking acronyms or e-mail addresses).

Group Discussion: Future Research in Non-Speech Audio Interfaces

One topic of discussion that surfaced multiple times during the workshop was how to design integrated auditory interfaces. Although we are now versed in a number of techniques for conveying information with non-speech auditory cues, our knowledge of how to incorporate these techniques with other concepts common in user interfaces is limited. One obvious goal is to deftly combine speech and non-speech audio to create powerful audio-only interfaces. But we first must understand when to use which type of output and, potentially more difficult, how to effectively overlay the techniques.

Another related area of interest was how to compare and integrate auditory and graphical interfaces. One goal expressed by workshop participants was to create a hear-and-feel standard much like the look-and-feel standards for graphical interfaces. This line of research is being addressed in multiple ways. First, research in transforming graphical interfaces into auditory interfaces (Mynatt 94a) is examining how the constructs in graphical interfaces can be conveyed through non-speech auditory cues. Second, the lack of design guidelines that are common for the creation of graphical interfaces has plagued interface designers who want to effectively build on previous research in auditory interfaces. Mynatt (Mynatt 94b) is exploring a design methodology for the quantitative and qualitative evaluation of auditory icons. This methodology aims to aid the interface designer in selecting auditory cues to convey actions and objects in interactive interfaces.

Only the simplest auditory interface contains just one sound. Any reasonably complex interface will include multiple sounds, potentially presented concurrently. Yet, most research has focused on working with sounds on a one-at-a-time basis. Working with multiple sounds requires designing auditory spaces. This design must be addressed at multiple conceptual levels. At the lowest level, designers wishing to present concurrent sounds must address problems with masking. Although masking has been a common term in psychoacoustic literature, it has not be sufficiently evaluated for multiple complex sounds. The aesthetics of working with multiple sounds requires the designer to avoid cacophony in the resulting interface. At the highest level, the expressive power of sounds must be taken into account. The auditory space can lend itself to multiple interpretations as users conceptually chain sounds together in novel and unanticipated ways (Cohen 94).

A common complaint was the limitations of current technology to support the creation of interesting non-speech auditory interfaces. Although audio hardware is now becoming

commonplace on desktop computers, the infrastructure to design and evaluate auditory interfaces is still lacking. Two specific areas of research were discussed. First, the difficulty of working with sampled sounds has led Gaver (Gaver 94) to develop algorithms for synthesizing auditory icons. The potential design flexibility these algorithms may offer is clear. Second, the requirement for a full featured audio server was ubiquitous. These servers need to gracefully work with multiple sounds at many levels such as mixing, sample rate conversion, priority scheduling and associating sounds to various events.

Given that the field of auditory interface design is still quite small compared to related fields such as graphical interface design, visualization and virtual reality, a pressing concern was the need to identify the tasks that would benefit the most from non-speech auditory interfaces. As discussed earlier, Gaver argued for focusing on tasks which require users to monitor their peripheral surroundings. On-going research is also addressing such tasks as monitoring complex systems, supporting computer access for the visually impaired (Mynatt 94c), reading maps (Blattner 94), and debugging parallel programs (Jackson 94). The overall goal of this work is to mimic the ways that we constantly use sound in our natural environments to tell us about the things that we cannot see.

Informal Activities

During breaks several participants informally showed some of their work. David Thiel showed a video tape of a conversational agent “Peedy the Parrot” that recognized continuous speech and played back recorded segments of speech. This system was shown live during one of the CHI 94 Demonstration sessions (Ball 94). Barry Arons showed video tapes of the Conversational Desktop (Schmandt 86, Schmandt 87) and Hyperspeech (Arons 91, Arons 93b), and gave a live demonstration of SpeechSkimmer (Arons 93a, Arons 94).

Future Plans

We have compiled an on-line list of resources (journals, conferences, mailing lists, etc.) for people working in, or joining, this field. We plan to make this information available on many sites on the World Wide Web, including the home page for research in auditory interfaces located in the Graphics, Visualization and Usability Center at Georgia Tech (http://www.gatech.edu/gvu/gvutop.html).

We discussed having a follow-on workshop at a future CHI, or perhaps a focused multi-day mini-conference. We are also investigating future publications based on the workshop or follow-on activities.

Conclusions

The workshop was an interesting and intense two days. It is difficult to capture the flavor and content of the workshop in just a few pages. This document has attempted to present only major points made during the workshop, and hence probably does not it justice. We have attempted to cite all the literature referenced during the workshop.

We succeeded in bringing together a variety of people that are researching, or using speech and audio in their work. For some of the participants, some of the discussions were reviews of old and familiar material, while for others it was a whole new world of thinking about speaking and listening rather than looking.

While we attempted to focus the discussions on user interface related issues many times we found ourselves asking questions like “how did you do that”, “where can I buy one of those”, or “doesn’t somebody have software that will do this for me”. As Catherine Wolf noted in a post-workshop conversation: We are currently stymied by technical problems, we (the workshop participants and the entire community) think that we can address deep and important user interface issues if we can get over the current hardware and software problems. In many ways the

speech and audio user interface community is where the graphics community was ten or twenty years ago. The tools and techniques for using speech and audio in the user interface are just beginning to become available, and the insights, experience, and lore of audio interfaces are not widely known.

Acknowledgments

Thanks to all who participated in the workshop for their free exchange of ideas and discussions on these important topics. We hope that we have done justice in distilling many hours of interactive conversation into a few pages of text.

References

Arons 91	B. Arons. Hyperspeech: Navigating in Speech-Only Hypermedia. In Proceedings of Hypertext (San Antonio, TX, Dec. 15-18), ACM, New York, 1991, pp. 133-146.
Arons 93a	B. Arons. SpeechSkimmer: Interactively Skimming Recorded Speech. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST), ACM SIGGRAPH and ACM SIGCHI, ACM Press, Nov. 1993, pp. 187-196.
Arons 93b	B. Arons. Hyperspeech (videotape). ACM SIGGRAPH Video Review 88 (1993). InterCHI '93 Technical Video Program.
Arons 94	B. Arons. Interactively Skimming Recorded Speech. Ph.D. dissertation, MIT, Feb. 1994.
Astheimer 93	P. Astheimer. What You See is What You Hear: Acoustics Applied in Virtual Worlds. In Proceedings of IEEE Symposium on Research Frontiers in Virtual Reality (San Jose, CA, Oct. 25-26), IEEE Computer Society Press, 1993, pp. 100-107.
Ball 94	J. E. Ball, D. T. Ling, D. Pugh, T. Skelly, A. Stankosky, and D. Theil. ReActor: A System for Real-Time, Reactive Animations. In CHI '94 Conference Companion (Boston, MA, Apr. 24-28), SIGGCHI, ACM, New York, 1994, pp. 39-40.
Berglund 94	B. Berglund, K. Harder, and A. Preis. Annoyance Perception of Sound and Information Extraction. Journal of the Acoustic Society of America (Mar. 1994), 1501-1509.
Blattner 94	M. M. Blattner, A. L. Papp and E. P. Glinert. Sonic Enhancement of Two-Dimensional Graphic Displays. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading, MA: Addison-Wesley Publishing Company, Inc., vol. XVII, Santa Fe Institute Studies in the Sciences of Complexity, 1994, 447-470.
Bregman 90	A. S. Bregman. Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, MA: MIT Press, 1990.
Burgess 92	D. A. Burgess. Techniques for Low Cost Spatial Audio. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST), ACM SIGGRAPH and ACM SIGCHI, ACM Press, Nov. 1992, pp. 53-59.
Cohen 91	M. Cohen and L. F. Ludwig. Multidimensional Window Management. International Journal of Man/Machine Studies 34 (1991), 319-336.
Cohen 92a	J. Cohen. Kirk Here: Using Genre Sounds to Monitor Background Activity. In INTERCHI 93 Adjunct Proceedings (Monterey, CA, May 3-7), ACM, New York, 1992, pp. 63-64.
Cohen 92b	P. R. Cohen. The Role of Natural Language in a Multimodal Interface. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST), ACM SIGGRAPH and ACM SIGCHI, ACM Press, Nov. 1992, pp. 143-149.
Cohen 93	M. Cohen. Integrating Graphic and Audio Windows. Presence 1, 4 (Fall 1993), 468-481.
Cohen 94	J. Cohen. Monitoring Background Activities. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading, MA: Addison-Wesley

Publishing Company, Inc., vol. XVII, Santa Fe Institute Studies in the Sciences of Complexity, 1994, 499-531.

Cutler 83 A. Cutler. Speaker's conceptions of the function of prosody. In A. Cutler and D.R. Ladd, editors, Prosody: Models and Measurements, chapter 7, pages 79-91. Springer-Verlag, 1983.

Durlach 92 N. I. Durlach, A. Rigopulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel. On the Externalization of Auditory Images. Presence 1, 2 (1992), 251-257.

Gaver 89 W. W. Gaver. The SonicFinder: An Interface that uses Auditory Icons. Human-Computer Interaction 4, 1 (1989), 67-94.

Gaver 93 W. W. Gaver. Synthesizing Auditory Icons. In Proceedings of INTERCHI (Amsterdam, The Netherlands, Apr. 24-29), SIGGCHI, ACM, New York, 1993, pp. 228-235.

Gaver 94 W. W. Gaver. Using and Creating Auditory Icons. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading, MA: Addison-Wesley Publishing Company, Inc., vol. XVII, Santa Fe Institute Studies in the Sciences of Complexity, 1994, 417-446.

Grosz 86 B. J. Grosz and C. L. Sidner. Attention, Intentions, and the Structure of Discourse. Computational Linguistics 12, 3 (1986), 175-204.

Hindus 93 D. Hindus, C. Schmandt, and C. Horner. Capturing, Structuring, and Representing Ubiquitous Audio. ACM Transactions on Information Systems 11, 4 (Oct. 1993), 376-400.

Jackson 94 J. Jackson and J. Francioni. Synchronization of Visual and Aural Parallel Program Performance Data. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading, MA: Addison-Wesley Publishing Company, Inc., vol. XVII, Santa Fe Institute Studies in the Sciences of Complexity, 1994, 291-306.

Kaye 92 D. Kaye and J. LeBrecht. Sound and Music for the Theater: The Art and Technique of Design. Back Stage Books, 1992.

Kramer 94 G. Kramer. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading, MA: Addison-Wesley Publishing Company, Inc., vol. XVII, Santa Fe Institute Studies in the Sciences of Complexity, 1994.

Levelt 83 W.J.M. Levelt and A. Cutler. Prosodic marking in speech repair. Journal of Semantics, 2(2):205-217, 1983.

Levelt 91 W.J.M. Levelt. Self-monitoring and self-repair. Speaking: >From Intention to Articulation, chapter 12, pages 458-499, 1991.

Mynatt 94a E. Mynatt. Auditory Presentation of Graphical User Interfaces. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading, MA: Addison-Wesley Publishing Company, Inc., vol. XVII, Santa Fe Institute Studies in the Sciences of Complexity, 1994, 533-555.

Mynatt 94b E. Mynatt. Designing with Auditory Icons: How Well Do We Identify Auditory Cues? In CHI '94 Conference Companion (Boston, MA, Apr. 24-28), SIGGCHI, ACM, New York, 1994, pp. 269-270.

Mynatt 94c E. Mynatt and G. Weber. Providing Access to Graphical User Interfaces: Contrasting Two Approaches. In Proceedings of CHI '94 (Boston, MA, Apr. 24-28), SIGGCHI, ACM, New York, 1994, pp. 166-172.

Pompetzki 94 W. Pompetzki and J. Blauert. A Study on the Perceptual Authenticity of Binaural Room Simulation. In Proceedings of Wallace Clement Sabine Centennial Symposium (Cambridge, MA, Jun. 5-7), Acoustical Society of America, 1994, pp. 81-84.

Roe 93 D. B. Roe and J. G. Wilpon. Whither Speech Recognition: The Next 25 Years. IEEE Communications Magazine 31, 11 (Nov. 1993), 54-62.

Schmandt 86 C. Schmandt and B. Arons. A Robust Parser and Dialog Generator for a Conversational Office System. In Proceedings of 1986 Conference, American Voice I/O Society, 1986, pp. 355-365.

Schmandt 87 C. Schmandt and B. Arons. Conversational Desktop (videotape). ACM SIGGRAPH Video Review 27 (1987).

Schmandt 94	C. Schmandt. Voice Communication with Computers: Conversational Systems. New York: Van Nostrand Reinhold, 1994.
Sellen 92a	A. Sellen, B. Buxton, and J. Arnott. Using Spatial Cues to Improve Video Conferencing. In Proceedings of CHI (Monterey, CA, May 3-7), ACM, New York, 1992, pp. 651-652.
Sellen 92b	A. Sellen. Speech Patterns in Video-Mediated Conversations. In Proceedings of CHI (Monterey, CA, May 3-7), ACM, New York, 1992, pp. 49-59.
Shriberg 92	E. Shriberg, J. Bear and J. Dowding. Automatic detection and correction of repairs in human-computer dialog. In Proceedings of the DARPA Workshop on Spoken Language Systems, 1992.
Stifelman 93	L. J. Stifelman, B. Arons, C. Schmandt, and E. A. Hulteen. VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker. In Proceedings of INTERCHI (Amsterdam, The Netherlands, Apr. 24-29), ACM, New York, 1993, pp. 179-186.
Wauchope 94	K. Wauchope. Eucalyptus: Integrating Natural Language Input with a Graphical User Interface. Naval Research Laboratory, technical report no. NRL/FR/5510-94-9711, Feb. 1994.
Wenzel 92	E. M. Wenzel. Localization in Virtual Acoustic Displays. Presence 1, 1 (1992), 80-107.
Whittaker 94	S. Whittaker, P. Hyland, and M. Wiley. Filochat: Handwritten Notes Provide Access to Recorded Conversations. In Proceedings of CHI (Boston, MA, Apr. 24-28), SIGCHI, ACM, New York, 1994, pp. 271-277.
Yankelovich 94	N. Yankelovich. Talking vs. Taking: Speech Access to Remote Computers. In CHI '94 Conference Companion (Boston, MA, Apr. 24-28), SIGGCHI, ACM, New York, 1994, pp. 275-276.