Hyperspeech

Barry Arons

Speech Research Group
MIT Media Laboratory
20 Ames Street, Cambridge, MA 02139
+1-617-253-2245, barons@media-lab.mit.edu

ABSTRACT
Hyperspeech is a speech-only hypermedia application that explores issues of speech user interfaces, navigation, and system architecture in a purely audio environment without a visual display. The system uses speech recognition input and synthetic speech feedback to aid in navigating through a database of digitally recorded speech segments.

KEYWORDS
Speech user interfaces, speech applications, hypermedia, speech as data, speech recognition, speech synthesis, conversational interfaces.

OVERVIEW
The hyperspeech system demonstrated in the video is an interface that presents "speech as data," allowing a user to wander through a database of recorded speech without any visual cues [1,3,4]. Navigating in the audio domain is more difficult than in the graphical domain, since the ear cannot browse a set of recordings the way the eye can scan a screen of text and images.

Audio interviews were automatically gathered by telephone, and manually segmented by topic into approximately 80 nodes. Over 700 typed hypertext-style links [2] were created to connect logically related comments and ideas. After listening to a particular node, links can be followed for: exploring comments from a particular speaker, hearing supporting and opposing views, or browsing the database at several levels of detail.

The interface allows the user to actively drive through the database rather than being passively "chauffeured" around by menus and prompts. This ability is based on a common set of navigational commands (i.e., link types that can be followed) that are independent of location in the database.

Every effort has been made to streamline the conversational interaction since time is such a valuable commodity in voice systems. For example, transitions from one interaction modality to another (e.g., from recognizing a word to playing a recorded segment) are designed for low system response time. Secondly, all actions by the system are easily interruptible by the user, and provide immediate feedback that an interrupt was received. Finally, the speech segments are played back at a faster rate than originally recorded, without a change of pitch. However, if a user asks to repeat a segment, it is replayed at normal speed for maximum intelligibility.

The system currently runs on a Sun SPARCstation; recorded segments are played directly on the workstation. A speaker-dependent isolated word recognizer and text-to-speech synthesizer are used for input and feedback. Note that the entire system (i.e., recognition, synthesis, and time-compression of speech) could be entirely software-based, and run in real-time on a current generation workstation.

The user interface evolved significantly during an iterative design process. Navigation in the initial design was difficult because of too much feedback, changing menus, and lack of stable audio landmarks. The current system, as shown in the video, demonstrates that interacting with a computer by voice can be very powerful, particularly when the same modality is used for both input and output. While the video presents the interface to a particular hypermedia application, the project's goal is to explore more general forms of interaction with speech data.

THINGS TO NOTE IN THE VIDEO
There are several points to listen for in the video:
- The quick transitions and system response time.
- The change in speed when a speech segment is repeated.
- The use of speech as both the primary data of the system, as well as for I/O.
- The difference between navigating and retrieving information in the speech domain and a traditional hypertext system that uses a display and mouse.

REFERENCES
1. Arons, B. Hyperspeech: Navigating in speech-only hypermedia. In *Hypertext '91 Proceedings*, pp. 133-146. ACM, 1991.

2. Conklin, J. Hypertext: An introduction and survey. *IEEE Computer*, 20(9):17-41, Sept. 1987.

3. Muller, M.J. and Daniel, J.E. Toward a definition of voice documents. In *Proceedings of COIS '90*, pp. 174-183. ACM, 1990.

4. Stifelman, L.J., et al. VoiceNotes: A speech interface for a hand-held voice notetaker. In *Proceedings of CHI '93*. ACM, 1993 (this volume).