

Introduction to the Special Issue on Speech as Data

Although humans developed speech to fulfill a need for communication long before written languages appeared, it is only in the last decade that computer-mediated speech communication has become practical. Widespread availability of networked, multimedia computers allowed the perhaps surprising interest in low-quality “internet telephony.” Low-cost mass storage now allows even home computers to manage audio and video.

As a result, we see increasing use of “speech as data” in our computer systems. That is, audio bits that serve their purpose without any transcription or automatic word recognition. This poses research questions: what is the utility of speech as data? how do we manipulate it? how do its real-time uses differ from retrieval from memory? The two articles in this special issue on “speech as data” address these questions.

The first article, “SpeechSkimmer: A System for Interactively Skimming Recorded Speech” by **Barry Arons**, describes how speech, with its temporal constraints, can be managed effectively by ordinary end-users. The article covers techniques for structuring, filtering, and presenting recorded speech in a way that allows users to navigate and interactively find information in audio recordings. The specific system, SpeechSkimmer, is a user interface for skimming speech at various levels of detail. It incorporates special processing techniques to allow users to hear recorded sounds much more quickly than normal playing speed, but without the distortion we ordinarily associate with speech played too fast.

The second article, “Hanging on the ‘Wire: A Field Study of an Audio-Only Media Space” by **Mark Ackerman**, **Debby Hindus**, **Scott Mainwaring**, and **Brian Starr**, investigates real-time speech integrated into the workplace. These researchers set up an audio-only communication system, called Thunderwire, which allows users who are in different physical places to communicate with each other verbally, as if they were in the same physical space. The study described in the article addresses many of the social issues that arise when speech is introduced into a work environment, including the development of social norms and the concern for privacy.

Taken together, the two articles cover the spectrum of speech as data, from when the speech is produced to after some of it has been recorded for later reuse. The Thunderwire study provides some valuable insights about the role and, to some extent, the limitations of audio communication in the workplace. Once such audio communication is captured, it can become an overwhelming task to retrieve useful information from the archive. In this regard, the SpeechSkimmer article is particularly important in that it address new user interface paradigms for dealing with stored audio. In the past, serious work on these interface techniques has been sparse, mainly because speech was an impractical medium. Now it is encouraging to see techniques emerging for handling speech as a first-class citizen in the computing environment.

CHRIS SCHMANDT
MIT Media Lab

NICOLE YANKELOVICH
Sun Microsystems Laboratories