Two New Techniques for Natural Spoken User Interfaces

Garrett Weinberg, Bhiksha Raj
Mitsubishi Electric Research Labs
201 Broadway
Cambrdige, MA 02139, USA
Tel: 1-617-621-7547
E-mail: *weinberg@merl.com*
bhiksha@merl.com

Kaustubh Kalgaonkar
Georgia Institute of Technology
Atlanta, GA 30332
Tel: 1-404-894-2900
E-mail: kaustubh@ece.gatech.edu

ABSTRACT
Speech can be an excellent modality for simple information retrieval tasks, such as finding points of interest in an automotive navigation system or selecting music from a personal player. However, to achieve natural yet robust retrieval from spoken inputs in noisy, real-world environments, new techniques are needed. We will demonstrate two such techniques, one software-based and one hardware-based. In the former we use the whole search space of the speech recognition system to perform retrieval; in the latter, we use a new device, an ultrasonic Doppler microphone, to detect the start and end of utterances, thus enabling hands free operation. Together these two techniques can significantly improve the usability of speech interaction for our target applications.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors

Keywords: Speech, retrieval, Doppler

INTRODUCTION

Speech is a natural input modality for human-computer interactions that involve retrieval tasks such as finding points of interest in an automotive navigation system or selecting music from a personal player. To be most natural, a spoken UI must allow the user to speak naturally and to interact with the system in a hands-free manner. These are difficult requirements: i) automatic speech recognition (ASR) engines perform relatively poorly on natural speech, particularly in noisy real-world environments, and ii) it is hard to detect when a talker is actually addressing the system in hands-free operation.

In this paper we demonstrate two techniques designed to address these problems. In the first, we use the space of hypotheses considered by the recognizer to perform retrieval robustly from free-form spoken input. In the second, we use an ultrasonic Doppler device to detect when a talker is addressing the system.

UIST'06, October 15–18, 2006, Montreux, Switzerland.

SpokenQuery

SpokenQuery is a patented information-retrieval technology that lets users locate items they need by saying descriptive search-words instead of typing them. SQ extends the usability of ASR engines by post-processing n-best results from the recognition of free-form speech, then retrieving items from an application-specific database. SQ integrates with state-of-the-art commercial ASR engines, relieving developers of speech grammar- and vocabulary-management issues, letting them focus instead on the application's interface and its integration with database items.

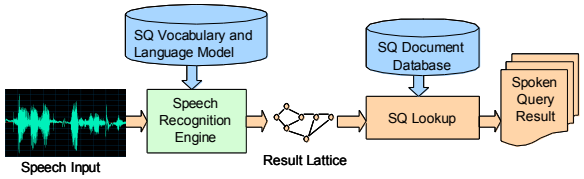

Figure 1: The speech recognizer's result lattice is used to create a query term vector that is compared against the document feature vectors.

SpeakPod. SpeakPod leverages SpokenQuery to provide an easy and intuitive way to enjoy music from an iPod in an automobile. Instead of pressing numerous buttons or navigating complex menus, users simply say any combination of song title, artist name, and album name, and the system presents a sorted list of the best matches.

Ultrasonic Doppler Microphone

The ultrasonic Doppler microphone consists of an ultrasonic transmitter/receiver pair coupled to an audio microphone. The ultrasonic transmitter emits a high-frequency tone in the direction of the talker's face. The frequency of this tone is modulated by speech-related movements of the talker's face through the Doppler effect. The Doppler modulated signals are captured through the ultrasonic sensor. Speech is indicated only when known speech-related spectral patterns are observed in the Doppler signal, meaning that the talker is speaking in the direction of the microphone.