Name(s)	Project Number
David C. Liu	**S1610**

Project Title

Continual Adaptation of Acoustic Models for Domain-Specific Speech Recognition

Abstract

Objectives/Goals

Advances in automatic speech understanding bring a new paradigm of natural interaction with computers. The Web-Accessible Multi-Modal Interface (WAMI) system developed by MIT provides a speech recognition service to a range of lightweight applications for Web browsers and cell phones.

However, WAMI currently has two problems. First, to improve performance, it requires continual human intervention through expert tuning--an impractical endeavor for a large shared speech recognition system serving many applications. Second, WAMI is limited by its global set of models, suboptimal for its variety of unrelated applications.

Methods/Materials

In this research I developed a method to automatically adapt acoustic models and improve performance. The system automatically produces a training set from the utterances recognized with high confidence in the application context. I implemented this adaptive system and tested its performance using a data set of 106,663 utterances collected over one month from a voice-controlled game.

To solve the second problem, I also extended the WAMI system to create separate models for each application.

Results

The utterance error rate decreased 13.8% by training with an adaptation set of 32,500 automatically selected utterances, and the trend suggests that accuracy will continue to improve with more usage.

The system can now adapt to domain-specific features such as specific vocabularies, user demographics, and recording conditions. It also allows recognition domains to be defined based on any criteria, including gender, age group, or geographic location.

Conclusions/Discussion

This research has enabled the WAMI system to automatically learn from its users and reduce its error rate. The extended WAMI can create customized models to optimize performance for each application and user group. These improvements to WAMI bring it one step closer towards being an "organic," automatically-learning system.

Summary Statement

I extended MIT's speech recognition system to make it learn on-the-fly as more people use it. The system serves many Web and mobile applications simultaneously. My work brings it closer to being an "organic" and self-learning system.

Help Received

I worked under the general guidance of Ian McGraw at the Spoken Language Systems group of MIT CSAIL.