Automatic Phone Recognizer

Sponsor Information:

Dr. Okim Kang
Department of English (TESL/Applied Linguistics)
Northern Arizona University
Okim.Kang@nau.edu

Project Description

In linguistics, prosody is the rhythm, stress, and intonation of speech. Prosody may reflect various features of the speaker or the utterance: the emotional state of the speaker; the form of the utterance (statement, question, or command); the presence of irony or sarcasm; emphasis, contrast, and focus; or other elements of language that may not be encoded by grammar or by choice of vocabulary. The Applied Linguistics Speech Lab (ALSL) is developing a number of programs to extract prosodic features from audio speech files. These features are then provided as input to machine learning classifiers whose output is being utilized to measure the speaker's English proficiency and comprehensibility. The KALDI ASR is an open source automatic speech recognizer (ASR) that is used by the ALSL to recognize speech phones. Phones are the fundamental sounds that make up all human speech and are independent of any particular language or dialect. Using an automatic speech recognizer is somewhat of an over-kill for recognizing phones. Several Windows based Matlab phone detectors are available on-line. The goal of this project would be to research the many phone detectors, implement one, and improve it until it can recognize phones as well as or better than the KALDI ASR.

Knowledge, skills and expertise required for this project

• MATLAB

Equipment Requirements

• MATLAB (ALSL will purchase licenses for the students)

Deliverables:

• Matlab code
• Regular progress meetings
• User manual
• Documentation sufficient that another programmer could add additional functionality to the code with minimal effort