Student: Minshu Zhan
Term: 2012 Summer
Faculty Supervisor: Stefanie Shattuck-Hufnagel
Date: June 21, 2012

Automatic Generation of Phonetic Landmarks

The overall goal of the project is to model the process by which human listeners recognize and understand spoken language. We assume that this process involves the recognition of sound segments and their features using acoustic cues. One particular kind of acoustic cue is the pattern of robust spectral changes, called landmarks. Under the assumption that these robustly implemented cues play a particularly important role in human's sound perception, the project requires developing programs that processes this aspect of spoken utterances automatically.

My job in this project includes the following main tasks: (a) predict landmarks from word sequences using existing rules; (b) detects the mutations, deletions, and insertions in hand landmark labels compared with the predicted landmarks; (c) extract contextual information, e.g. syllable stress and position, associated with each landmark; (d) based on the contexts, generate and evaluate possible decision trees that model the differences between predicted and actual landmarks.

During spring term UROP, I have completed three of the python scripts required for the first task. **LexiconExtract.py** extracts words from the conversation textgrid files and outputs a mini lexicon for the given speech samples**; LMdicParsor.py** parses the phoneme-to-LM matrix; **LMPredictor.py** outputs for each input file a new textgrid file with four additional tiers - LM, glottalization, velopharyngeal, and the phonemes.

During the summer of 2012, I will: (i) modify our current landmark comparison code to ensure its performance and correctness when applied to various input files; (ii) implement the data representation of landmarks' context information, including stresses and positions of relevant syllables, words, and phrases, as well as word frequencies; (iv) add context extraction functions into the landmark prediction script; (iv) apply the data to 'C4.5', an existing decision tree module. Upon completion of these tasks, I will start implementing similar programs for other types of non-landmark acoustic cues related to the distinctive features of voicing and place.

My motivation for joining this project is to acquire a better understanding of human's

acoustic perception process. I have always been interested in the field of automatic sound perception and wish to be more familiar to the basic techniques and models in sound processing through this project. The project is a very good practice in programming and artificial intelligence for me.