Evaluating the Use of Prosodic Information in Speech Recognition and Understanding

Mari Ostendorf   Patti Price

Boston University   SRI International

Objective:

The goal of this project is to investigate the use of different levels of prosodic information in speech recognition and understanding. The work will involve determining which aspects of prosody are useful in speech recognition and understanding, developing reliable algorithms for analysis of prosodic information, and incorporating these into a Spoken Language System (SLS) being developed at SRI. Our approach is multi-disciplinary: combining linguistic theory, speech knowledge and statistical modeling techniques. This research is sponsored jointly by DARPA and NSF, and is coordinated with another NSF project entitled "Prosody Analysis/Synthesis Using Probabilistic Models and Linguistic Theory."

Summary of Accomplishments:

- Specified conventions for labeling prosodic phenomena and hand-labelled several minutes of speech with this convention.

- Developed a method to automatically phonetically label and align speech data given the orthographic transcription using the SRI HMM word recognition system.

- Conducted perceptual experiments with phonetically ambiguous sentences to examine the role of prosody in parsing.

- Developed a formalism for providing prosodic information – phrase boundaries (or word connectivity) and lexical stress – to a parser.

- Developed and evaluated algorithms for automatically extracting word connectivity values based on duration and pause cues, and predicting lexical stress from duration cues.

Plans:

- Analyze recognition and parsing errors to understand what aspects of prosody will be most important to a spoken language system and what components of the system could use this information.

- Develop a phrase boundary detection algorithm using breaths, pauses, boundary tones and duration information and evaluate accuracy with respect to hand-labelled data.

- Develop a formalism for incorporating phrase boundary information into a parser and evaluate the usefulness of hand-labeled and automatically detected boundary information in the SRI spoken language system.