

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. In particular, the current focus of the work is the use of prosodic phrase boundary information in parsing. The research involves determining a representation of prosodic information suitable for use in a speech understanding system, developing reliable algorithms for detection of the prosodic cues in speech, investigating architectures for integrating prosodic cues in a parser, and evaluating the potential improvements of prosody in the context of the SRI Spoken Language System. This research is sponsored jointly by DARPA and NSF.

Summary of Accomplishments:

- Developed algorithms for the automatic detection of breaths and silences; the initial breath detection algorithm missed only 7% of all breaths with one insertion in 6 minutes of speech.
- Implemented and evaluated several methods for automatically detecting prosodic break indices based on relative duration; the current algorithm has a correlation coefficient of 0.85 with the labels produced by expert listeners and can handle changes in speech rate within a sentence.
- Integrated the prosodic information with the SRI natural language component by modifying the grammar to use the indices to rule out prosodically inconsistent syntactic hypotheses; this integration reduces the number of syntactic parses by about 25% and yields a unique parse for many otherwise ambiguous sentences.

Plans:

- Evaluate the break index detection algorithms in a speaker-independent context, and on new corpora, such as paragraphs of speech (as opposed to sentences) and spontaneous speech.
- Improve the break index algorithms by combining different acoustic cues, including as duration, intonation and pauses/breaths.
- Develop a scoring algorithm for ranking parses according to the correlation of prosodic and syntactic information.