Evaluating the Use of Prosodic Information in Speech Recognition and Understanding

Mari Ostendorf   Patti Price

Boston University   SRI International
Boston, MA 02215   Menlo Park, CA 94025

PROJECT GOALS

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.

RECENT RESULTS

- Investigated duration lengthening at different levels of prosodic breaks over different types of units (e.g., final syllable, interstress interval), finding that the primary region of lengthening is the phrase-final, word-final syllable rhyme.

- Implemented a binary tree quantizer in our HMM prosodic phrase boundary detection system, which enabled the use of multiple cues determined from the above analysis. A new algorithm for adapting model parameters has also been implemented. The resulting algorithm has speaker-independent break recognition performance above the level of our previous speaker-dependent system. A paper describing this work (Wightman and Ostendorf) appears in the *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*.

- Submitted for publication an article describing a perceptual, phonological and phonetic analysis of the relationship between prosodic structure and syntactic structure. A shortened version of this paper (Price et al.) appears in this volume. (This work was also funded by a related grant on prosody, NSF grant number IRI-8805680.)

- Developed an initial version of an analysis-by-synthesis approach for scoring a syntactic parse in terms of the observed prosodic constituents. This algorithm leverages recent results on the role of prosody in disambiguation, mentioned above. We assessed this method by using the algorithm to decide between pairs of ambiguous sentences, finding that the algorithm performance is close to human perception. A paper describing this work (Wightman et al.) appears in this proceedings.

PLANS FOR THE COMING YEAR

- Further improve the break index algorithms by adding an intonation feature to the quantizer. We expect this to improve performance considerably because the main source of errors is confusion between break indices 3 (no boundary tone) and 4 (marked with boundary tone).

- Evaluate the break index detection algorithms on paragraphs of speech (as opposed to sentences) and on spontaneous speech as opposed to read speech.

- Implement the parse scoring algorithm using automatically obtained parses and evaluate on additional data.

- Utilize the parse scoring algorithm in speech understanding.