

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

- Developed an algorithm for recognizing intonational cues (pitch accents or prominences and boundary tones) using a modified version of our previously developed break detection algorithm. Improved break detection performance by including probability of boundary tone as an additional feature.
- Extended previous work in analysis/synthesis parse scoring by introducing a new probabilistic scoring technique that uses a decision tree to predict prosodic phrase breaks. Disambiguation results show that this automatically trainable synthesis technique yields performance comparable to the rule-based synthesis algorithms previously investigated, even though the syntactic structures represented in the training corpus were quite different from those in the testing corpus.
- Investigated, in conjunction with SRI's SLS project, acoustic attributes of hypothesized repair locations, finding that relative durations of two repeated words and existence and duration of an intervening pause can be reliable cues to repairs.
- Discovered that pause fillers in spontaneous speech were a frequent source of recognition errors, analyzed acoustic cues to pause fillers, and found a sig-

nificant difference between the pitch of a pause filler and that of its local context, suggesting a simple detection algorithm.

- Played a major role in organizing and leading a workshop aimed at developing a common core prosodic transcription standard; the impending availability of large corpora of data including syntactic annotations makes the need for agreement on prosodic standards especially critical.

PLANS FOR THE COMING YEAR

- Evaluate the break index and prominence recognition algorithms on paragraphs of speech (as opposed to sentences) and on spontaneous speech as opposed to read speech. Investigate new acoustic features for improving recognition results.
- Extend parse scoring algorithm to include prominence information making use of tree-based prominence prediction algorithms.
- Utilize the parse scoring algorithm in speech understanding.
- Continue study of acoustic cues to repairs and other spontaneous speech effects; specifically, analyze the proportion of such events in different SLS configurations, the effect of such events on types of errors in different SLS configurations, and the role of intonation and duration patterns in their detection.
- Coordinate project with research on prosody at IPO, Eindhoven; this laboratory is the research center for the largest group in the world focussed in prosody. Dr. Price will spend three months at this institution working on the current project half-time.