

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. There are two thrusts in the current work: use of prosodic information in parsing and detection/correction of disfluencies. 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 speech understanding system, and evaluating the potential performance improvements possible through the use of prosodic information in a spoken language system (SLS). This research is sponsored jointly by DARPA and NSF, NSF grant no. IRI-8905249, and in part by a DARPA SLS grant to SRI.

RECENT RESULTS

- Evaluated the break index and prominence recognition algorithms on a larger corpus, with paragraphs (as opposed to sentences) of radio announcer speech.
- Extended the prosody-parse scoring algorithm to use a more integrated probabilistic scoring criterion and to include prominence information, making use of tree-based recognition and prediction models.
- Collaborated with a multi-site group for development of a core, standard prosody transcription method: TOBI, (TOnes and Break Indices), and labeled over 800 utterances from the ATIS corpus with prosodic break and prominence information. Analyses of consistency between labelers shows good agreement for the break and prominence labels on ATIS.
- Ported prosody-parse scoring algorithms to ATIS, which required: developing new features for the acoustic and prosody/syntax models and representing new classes of breaks to represent hesitation; currently evaluating the algorithm for reranking the N-best sentence hypotheses in the MIT and SRI

SLS systems. (This work was made possible by researchers at MIT and SRI who provided the parses and recognition outputs needed for training and evaluating the prosody models.)

- Developed a new approach to duration modeling in speech recognition, involving context-conditioned parametric duration distributions and increased weighting on duration.
- Developed tools for analysis of large number of repairs and other disfluencies; analyzed the prosody of filled pauses in ATIS data and extended the work on disfluencies to data in the Switchboard corpus of conversational speech.
- Developed methods for automatic detection and correction of repairs in ATIS corpus, based on integrating information from text pattern-matching, syntactic and semantic parsing.

PLANS FOR THE COMING YEAR

- Evaluate the break index and prominence recognition algorithms on spontaneous speech, specifically the ATIS corpus, and further refine algorithms to improve performance in this domain.
- Improve the parse scoring algorithm performance in the ATIS domain by exploring new syntactic features, and assess performance on SRI vs. MIT SLS systems.
- Investigate alternative approaches to integrating prosody in speech understanding.
- Continue study of acoustic and grammatical cues to repairs and other spontaneous speech effects.
- Based on the results of the acoustic analyses, develop automatic detection algorithms for flagging repairs that are missed by the syntactic pattern matching algorithms and develop algorithms for classifying detected repairs to aid in determining the amount of traceback in the repair.