Evaluating the Use of Prosodic Information in Speech Recognition and Understanding

M. Ostendorf P. Price S. Shattuck Hufnagel
Principal Investigators

Boston University SRI International MIT RLE
Boston, MA 02215 Menlo Park, CA 94025 Cambridge, MA 02138

PROJECT GOALS

The goal of this project is to investigate the use of different levels of prosodic information in speech recognition and understanding. The two main thrusts in the current work involve the use of prosodic information in parsing and detection/correction of disfluencies, but we have also investigated duration modeling for continuous speech recognition. The research involves determining a representation of prosodic information suitable for use in speech understanding systems, developing reliable algorithms for detection of prosodic cues in speech, investigating architectures for integrating prosodic cues in speech understanding systems, and assessing potential performance improvements by evaluating prosody algorithms in actual spoken language systems (SLS). This research is sponsored jointly by ARPA and NSF, NSF grant no. IRI-8905249.

RECENT RESULTS

Recent results on this project are summarized below, with names of the students primarily responsible for the work indicated in parentheses.

- Extended the prosodic prominence and break acoustic models by implementing an iterative pruning algorithm (N. Veilleux), integrating text and acoustic models (D. Macannuco), and developing a new energy feature based on results of recent linguistic studies.
- Continued work in prosody-parse scoring, running tests on a larger set of sentences and optimizing on word error rate and achieving 10% reduction in word error by combining the prosody-parse score with acoustic and language scores from the MIT ATIS system. Experiments on the SRI ATIS system are in progress. (N. Veilleux)
- Further explored parametric duration modeling in CSR using maximum likelihood clustering and speaking rate adaptation. Observed a 10% reduction in word error on the RM task, but no improvement in initial experiments on the WSJ task. (C. Fong)
- In analyses of the ATIS corpus, found that: hesitations are associated with intonation patterns similar to those in filled pauses (in addition to long pauses and lengthened segments) and occur at locations with relatively higher perplexity in the language model (N. Veilleux and A. Schlosser), that filled pauses occur almost exclusively in low-probability word sequences and have longer schwa duration than in the determiner "a", and that there are differences in the f0 patterns of fluent vs. disfluent single word repetitions (E. Shriberg).
- Developed methods for automatic detection of fragments from acoustic cues and patterns in the N-Best recognizer output using decision trees. (M. Hendrix)
- Developed a taxonomy for disfluencies and analyzed distributional properties of 5000 hand-labeled disfluencies from the ATIS corpus, the Switchboard corpus, and a third comparison corpus of human-human air travel planning speech. Findings include general models for predicting overall disfluency rates, relative rates of disfluency types, and relationships between disfluency type and type-independent features (e.g. presence of a word fragment or editing phrase). (E. Shriberg)
- Analyzed patterns of occurrences of word-initial glottalization, finding a high coincidence rate with phrase onset and prominence marking, which suggests new cues for prosodic pattern detection. (L. Dilley) [This work is also funded by an NIH grant.]

PLANS FOR THE COMING YEAR

- Further refine the break index and prominence recognition algorithms to improve accuracy on the ATIS corpus, and investigate the use of detected prominence in the SRI ATIS system as a knowledge source for rejecting or correcting template matcher output.
- Improve the parse scoring algorithm performance by exploring new syntactic features, and assess performance on SRI vs. MIT ATIS systems.
- Refine the fragment detection algorithm and extend to detection of other disfluencies by integrating acoustic and pattern matching text cues, and evaluate usefulness in the SRI ATIS system.