

Spoken Language Recognition and Understanding

Victor Zue and Lynette Hirschman

Spoken Language Systems Group
Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

1. PROJECT GOALS

The goal of this research is to demonstrate spoken language systems in support of interactive problem solving. The system accepts continuous speech input and handles multiple speakers without explicit speaker enrollment. The MIT spoken language system combines SUMMIT, a segment-based speech recognition system, and TINA, a probabilistic natural language system, to achieve speech understanding. The system engages in interactive dialogue with the user, providing output in the form of tabular displays, as well as spoken and written output. The system has been demonstrated on several applications, including travel planning and direction assistance.

2. RECENT RESULTS

- Reduced spontaneous speech recognition word error rate by more than a factor of two since the February 1991 evaluation through the use of low perplexity language models and context-dependent phonetic models.
- Reduced natural language weighted error by almost a factor of 2 on class A sentences through the use of a robust parsing mechanism, which integrates parsed phrases into a single semantic representation, using a slight extension of the existing discourse processing.
- Demonstrated a near real-time interactive spoken language system, running on a Sun SPARCstation or an IBM RX6000.
- Developed and experimented with alternative metrics for the evaluation of interactive spoken language systems, including use of task completion and time-to-completion for air travel planning tasks, as well as a log file based evaluation procedure.
- Collected nearly 20,000 sentences for the Wall Street Journal pilot corpus in support of research and development in large-vocabulary speech recognition systems.

- Chaired the MADCOW multi-site ATIS data collection effort, contributed over 5000 sentences of spontaneous ATIS data, and participated in the common evaluation for speech, spoken language and text input.

3. PLANS FOR THE COMING YEAR

- Improve SUMMIT recognition performance by developing phonetic and language models for silence and filled pauses, and mechanisms to detect and “repair” false starts and repeats.
- Continue experimentation on low perplexity language models (N-gram, LR parser, probabilistic parsing) to improve speech recognition performance.
- Model discourse and dialogue, including the use of error and clarification messages, to improve both recognition performance and the interactive nature of spoken language systems.
- Develop evaluation metrics for interactive spoken language systems based on experiments using the real-time ATIS spoken language system.
- Experiment with alternative user interaction strategies using near real-time data collection system.