

Spoken Language Recognition and Understanding

Victor W. Zue and Lynette Hirschman

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

OBJECTIVE

The goal of this research is to develop a spoken language system that will demonstrate the usefulness of voice input for interactive problem solving. The system will accept continuous speech, and will handle multiple speakers without explicit speaker enrollment. Combining *SUMMIT*, a segment-based speech recognition system, and *TINA*, a probabilistic natural language system, to achieve speech understanding, the system will be demonstrated in an application domain relevant to the DoD.

SUMMARY OF ACCOMPLISHMENTS:

- Developed a procedure for determining context-dependent models for lexical labels in the *SUMMIT* speech recognition system. Reduced word error rate by almost a factor of two on the 1,000 word Resource Management task.
- Performed experiments investigating the integration of syntax and semantics with acoustic evidence to improve system performance. Achieved a 33% improvement on performance score in the *VOYAGER* domain.
- Performed signal representation comparisons on the task of speaker-independent vowel classification. Demonstrated the robustness of the auditory model over other signal representations, particularly in the presence of noise.
- Developed a data collection procedure within the *ATIS* domain, and collected nearly 3,700 spontaneously generated sentences from over 100 speakers. Performed comparative analyses on the data collection at *TI* and *MIT*.
- Developed a preliminary version of the *MIT ATIS* system and participated in the common evaluation with both speech and text input.

PLANS:

- Improve *SUMMIT* recognition performance by incorporating more complex context-dependent models and experimenting with alternative classification algorithms.
- Provide tighter coupling of *TINA* and *SUMMIT* in order to exploit speech and natural language symbiosis. In particular, investigate how parse probability can be affected by discourse context.
- Investigate the modelling of discourse and dialogue, including the use of error and clarification messages, to improve both recognition performance and the interactive nature of spoken language systems.
- Collect additional speech and text data during actual problem solving for system development and evaluation in the *ATIS* domain.