Spoken Language Recognition and Understanding

Victor Zue

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 develop and demonstrate spoken language technology in support of interactive problem solving. 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 accepts continuous speech input and handles multiple speakers without explicit speaker enrollment. It engages in interactive dialogue with the user, providing output in the form of tabular and graphical displays, as well as spoken and written responses. We have demonstrated the system on several applications, including air travel planning and urban navigation/exploration; it has also been ported to several languages, including Japanese and French.

2. RECENT RESULTS

- **Improved Recognition and Understanding:** Reduced word error rate by over 50% from last year (while using a larger vocabulary with higher perplexity) through the use of improved acoustic-phonetic alignment and pronunciation modelling; reduced spoken language understanding error rate by over 25% from last year (while using a larger application back-end) by making use of stable corpus of annotated data.

- **On-Line Travel Planning:** Developed PEGASUS, an interactive spoken language interface for on-line travel planning connected to American Airlines' EAASY SABRE system.

- **Multi-lingual SLS:** Extended the bilingual VOY-AGER system to other languages including Italian, French, and German. The system uses a single semantic frame to capture the meaning irrespective of the language, and the langauge generation component has also been unified and enhanced. In addition, a segment-based language identification approach has been formulated and implemented. The resulting system, when evaluated on the OGI Multi-Language Telephone Speech Corpus, achieved an identification rate of 55.8%.

- **Phonological Parsing for Letter/Sound Generation:** Developed and implemented a framework for bi-directional letter/sound generation, using a version of our probabilistic natural language system, TINA. The system can parse nearly 95% of unseen words, and achieved word accuracies of 71.8% and 55.8% for letter-to-sound and sound-to-letter generation on the parsable words.

- **Tranformation-Based, Error-Driven Learning:** Refined and extended this technique for part of speech tagging, and achieved accuracies of 97.2% with 267 simple non-stochastic rules.

- **HLT Community Service:** Collected and distributed more than 1400 ATIS-3 sentences from 58 subjects. Distributed our POS tagger to over 150 sites. Vice-Chair of 1994 HLT workshop.

3. FUTURE PLANS

- **Technology Development:** Continue to improve speech recognition and language understanding technologies for large vocabulary spoken langauge systems. Areas of research include acoustic modelling, lexical access, adaptation techniques, SR/NL integration strategies, dialogue modelling, grammar induction, multilingual porting (e.g., Spanish and Mandarin), and discovery/learning of unknown words.

- **System Development:** Explore research issues within the context of developing a system that enables users to access and manipulate various (real) sources of information using spoken input in order to solve specific tasks. Initial focus will be in the travel domain, which include urban navigation, air travel planning, and weather information.