SPOKEN-LANGUAGE RESEARCH AT CARNEGIE MELLON

Raj Reddy, Principal Investigator

School of Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

PROJECT GOALS

The goal of speech research at Carnegie Mellon continues to be the development of spoken language systems that effectively integrate speech processing into the human-computer interface in a way that facilitates the use of computers in the performance of practical tasks. Research in spoken language is currently focussed in the following areas:

- **Improved speech recognition technologies:** Extending the useful vocabulary of SPHINX-II by use of better phonetic and linguistic models and better search techniques, providing for rapid configuration for new tasks.
- **Fluent human/machine interfaces:** Developing tools that allow users to easily communicate with computers by voice and understanding the role of voice in the computer interface.
- **Understanding spontaneous spoken language:** Developing flexible recognition and parsing strategies to cope with phenomena peculiar to the lexical and grammatical structure of spontaneous spoken language. Investigate methods of integrating speech recognition and natural language understanding. Development of automatic training procedures for these grammars.
- **Acoustical and environmental robustness:** Developing procedures to enable good recognition in office environments with desktop microphones and a useful level of recognition in more severe environments.
- **Rapid integration of speech technology:** Developing an approach that will enable application developers and end users to incorporate speech recognition into their applications quickly and easily, as well as the dynamic modification of grammars and vocabularies.

RECENT RESULTS

- SPHINX-II has been extended with a multi-pass search algorithm that incorporates two passes of beam search and a final A-star pass that can apply long-distance language models as well as produce alternative hypotheses.
- Joint training of acoustic models and language models is currently being explored in the context of the Unified Stochastic Engine (USE).
- A framework for long-distance language modeling was developed, in collaboration with IBM researchers. A pilot system using this model yielded significant reduction in perplexity over the trigram model.
- Developed improved recognition, grammar coverage and context handling that reduced SLS errors for the ATIS Benchmark by 67%. We also improved the robustness and user feedback in our live ATIS demo.
- Developed and evaluated two methods for more tightly integrating speech recognition and natural language understanding, producing error reductions of 20% compared to the loosely-coupled system.
- Added automatic detection capability for out-of-vocabulary words and phrases. New words are now entered instantly into the phone dialer application given only their spelling.
- Acoustical pre-processing algorithms for environmental robustness were extended to the CSR domain and made more efficient.

PLANS FOR THE COMING YEAR

- Use our existing language modeling framework to model long-distance dependence on words and word combinations. These new models will be allow the recognizer to take advantage of improved linguistic knowledge at the earliest possible stage.
- Implement confidence measures for large-vocabulary SLS systems, for new-word detection and greater accuracy.
- Continue to explore issues associated with very large vocabulary (100,000-word) recognition systems.
- Continue to develop methods for automatically acquisition of Natural Language information used by an SLS system.
- Improve user interaction in the ATIS system, including clarification and mixed initiative dialogs, speech output and form-based displays.
- Begin to develop a new SLS application, such as a telephone-based form filling application.
- Provide grammar switching and instantaneous new word addition for the general SPHINX-II decoder.
- Develop and test a 100,000-word pronunciation lexicon that will be available in the public domain.
- Continue to improve our cepstrum-based environmental compensation procedures.
- Demonstrate more robust microphone-array techniques.
- Extend our work on environmental robustness to long-distance telephone lines.
- Continue to enhance our spoken language interfaces, by introducing speech response capabilities and facilities for user customizing. Continue to investigate the appropriate use of speech in multi-modal interfaces.