

# 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 accuracy:** Improving SPHINX-II by better acoustic and language models and more efficient search algorithms. Obtaining task-independent acoustic models and senone mapping tables using large speech corpora. Developing task-adaptation techniques for both acoustic and language models.
- **Fluent human/machine interfaces:** Developing tools that provide fluent communication between users and computers by voice and understanding the role of voice in the computer interface.
- **Robust language modeling:** Developing techniques for combining multiple language training sources to create improved language models. Applying new knowledge sources to strengthen existing language models.
- **Understanding spoken language:** Developing flexible recognition and parsing strategies to cope with spoken language. Developing and evaluating methods of integrating speech recognition and natural language understanding. Developing automatic training procedures for grammars.
- **Acoustical and environmental robustness:** Extending environment-adaptation procedures to telephone channels and other more difficult environments.

## RECENT RESULTS

- A task-independent very-large vocabulary version of SPHINX-II has been developed from 37,000 WSJ utterances. WSJ acoustic models have been adapted to ATIS by use of pruning and interpolation. Recognition accuracy has been further improved by the use of phone-dependent codebooks and multiple speaker clusters.
- A one-pass trigram language model has been implemented that reduces search time by using a lexical tree.
- An adaptive long-distance language model was created that reduces perplexity by 32%-39% and word error rate by 14%-16% over the conventional trigram.

- User interaction for the ATIS system has been improved by incorporating clarification and mixed-initiative dialogs, speech output, and form-based displays.
- The Recursive Transition Network grammars used in the PHOENIX system have been combined with bigram language models in the A\* portion of the recognition search of SPHINX-II. This reduces sentence understanding errors.
- Error rates for telephone speech and other secondary environments have been reduced by 40 percent by implementing reduced-bandwidth analysis, phone-dependent cepstral normalization, and codebook-adaptation procedures.
- A phonetically-transcribed 100,000-word dictionary has been completed and released to the public.
- SPHINX-II has been adapted for use in CMU's Project LISTEN, a prototype reading coach funded by NSF.

## PLANS FOR THE COMING YEAR

- Develop a 100,000-word implementation of SPHINX-II that runs in real time and uses less than 50 megabytes.
- Add multiple dynamic language models and other forms of incremental learning to SPHINX-II.
- Investigate the use of fully-continuous HMMs to take full advantage of the large amount of available training data.
- Investigate technologies to reduce recognition errors caused by OOV words.
- Develop speech coding algorithms for real time implementation and use compressed speech signal in the speech recognition system.
- Develop techniques for combining multiple language models from different sources, and develop new knowledge sources to strengthen existing language models.
- Implement a speech-only interface for a workstation environment, supporting both command and dictation capabilities.
- Continue development and evaluation of systems in which both stochastic and grammar-based language model information are used to improve the speech recognition search.
- Improve the representation of context information and its relation to the parser in our Spoken Language System
- Continue to develop methods for automatic acquisition of Natural Language information used by an SLS system.
- Further extend and improve environmental robustness.
- Port spoken-language technologies to a new secretarial-assistant task.