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 of SPHINX-II by use of better phonetic models and better search techniques, providing for rapid configuration for new tasks.
- **Fluent human/machine interfaces:** Developing an understanding of how people interact by voice with computer systems, in the context of Office Management and other domains.
- **Understanding spontaneous spoken language:** Developing flexible parsing strategies to cope with phenomena peculiar to the lexical and grammatical structure of spoken language. Development of automatic training procedures for these grammars.
- **Dialog modeling:** Applying constraints based on dialog, semantic, and pragmatic knowledge to identify and correct inaccurate portions of recognized utterances.
- **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.

RECENT RESULTS

- The SPHINX-II system incorporated sex-dependent semi-continuous hidden Markov models, a speaker-normalized front end using a codeword-dependent neural network, and shared-distribution phonetic models.
- Vocabulary-independent recognition was improved by introducing vocabulary-adapted decision trees and vocabulary-bias training, and by incorporating the CDCN and ISDCN acoustical pre-processing algorithms.
- SPHINX-II has been extended to the Wall Street Journal CSR task by incorporating a practical form of between-word co-articulation modeling in the context of a more efficient beam search.
- The Carnegie Mellon Spoken Language Shell was reimplemented and additional applications for the Office Management domain were developed, including a telephone dialer and voice editor.
- Grammatical coverage in the ATIS domain was extended. An initial set of tools was developed to create the grammar in a semi-automatic fashion from a labelled corpus.
- The MINDS-II system was developed which identifies and reprocesses mis-recognized portions of a spoken utterance using semantics, pragmatics, inferred speaker intentions, and dialog structure in the context of a newly-developed finite-state recognizer.
- Acoustical pre-processing algorithms for environmental robustness were extended, made more efficient, and demonstrated in the ATIS domain. Pre-processing was combined microphone arrays and with auditory models in pilot experiments.

PLANS FOR THE COMING YEAR

- We will extend shared-distribution models to produce senonic baseforms, addressing the problem of new word learning and pronunciation optimization, and the the decision-tree-based senone will be made more general. The CDNN-based approach will be extended for both speaker and environment normalization. The use of long-distance semantic correlations in language models to improve the prediction capability will be explored.
- We will incorporate confidence measures, audio feedback, and the latest recognition technologies into the Office Manager system. We will investigate the behavior of multi-modal systems that incorporate speech recognition.
- We will develop architectures and automatic learning algorithms for SLS systems with greater integration of recognition, parsing, and dialog and pragmatics. Work will be initiated on the identification of misunderstood portions of a complete utterance, and the use of partial understanding and clarification dialogs.
- We will continue to develop parallel strategies for robust speech recognition, and we will demonstrate these methods in more adverse acoustical environments.