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. Component technologites are being developed in the context of spoken language systems in two domains: the DARPA-standard ATIS travel planning task, and CMU's office management task. Research in spoken language is currently focussed in the following areas:

- **Improved speech recognition technologies.** Research is directed toward increasing the useful vocabulary of the speech recognizer, using better subword models and vocabulary-independent recognition techniques, providing for rapid configuration for new tasks.

- **Fluent human/machine interfaces.** The goal of reserach in the spoken language interface is the development of an understanding of how people interact by voice with computer systems. Specific development systems such as the Office Manager are used to study this interaction.

- **Understanding spontaneous spoken language.** Actual spoken language is ill-formed with respect to grammar, syntax, and semantics. We are analyzing many types of spontaneous speech phenomena and developing appropriate syntactic and semantic representations of language that enable spontaneous speech to be understood in a robust fashion.

- **Dialog modeling.** This goal of this research is to identify invariant properties of spontaneous spoken dialog at both the utterance and dialog level, and to apply constraints based on dialog, semantic, and pragmatic knowledge to enhance speech recognition. These knowledge sources can also be used to learn new vocabulary items incrementally.

- **Acoustical and environmental robustness.** The goal of this work is to make speech recognition robust with respect to variability in acoustical ambience and choice of microphone, so that recognition accuracy using desk-top or bezel-mounted microphones in office environments will become comparable to performance using close-talking microphones.

RECENT RESULTS

- Incorporation of semi-continuous HMMs and speaker adaptation has produced speaker-adaptive recognition performance that is comparable to speaker-dependent performance reported previously by other sites. Speaker-adaptation algorithms using neural networks have also been developed with encouraging preliminary results.

- A vocabulary-independent speech recognition system has been developed. Improvements including the use of second order cepstra, between-word triphones and decision-tree clustering have produced a level of vocabulary-independent performance that is better than the corresponding vocabulary-dependent performance.

- A dynamic recognition-knowledge base has been incorporated into the Office Manager system, as well as models of noise phenomena. The natural language and situational knowledge capabilities of the system have also been extended.

- The ATIS system has been augmented by incorporating the use of padded bigrams and models for non-lexical events, providing for increased coverage at reduced perplexity. These changes have produced major improvements in accuracy using both speech and transcripts of ATIS dialogs as input.

- Six principles of dialog that characterize spontaneous speech at the pragmatic and semantic levels were identified. Algorithms were developed to invoke these principles at the utterance levels to constrain the search space for speech input and transcripts of ATIS dialogs.

- Pre-processing algorithms that normalize cepstral coefficients to compensate for additive noise and spectral tilt have been made more efficient.

PLANS FOR THE COMING YEAR

- We will continue to investigate neural-network-based speaker normalization and its application to speaker-independent speech recognition.

- The vocabulary-independent system will be improved by refinements in decision-tree clustering, pruning strategies, and selection of contextual questions. Non-intrusive task and environmental normalization will be introduced.

- We will continue refining the Office Manager system and begin using it as a testbed for the development of error repair strategies and intelligent interaction management.

- The constraints imposed by dialog models will be extended to allow more dialog and pragmatic knowledge to be used by the ATIS system in the understanding process. The ATIS system will be improved by the addition of out-of-vocabulary models and an improved rejection capability and user interface.

- Dialog-level knowledge will be applied to incremental word learning.

- Passive environmental adaptation will be incorporated into the speech recognizer in the Portable Speech Library. We will measure the extent to which processing using multiple microphones and physiologically-motivated front ends complements the robustness provided by acoustical pre-processing.