

Speech Research at Carnegie Mellon

Raj Reddy, Principal Investigator

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Carnegie Mellon University is engaged in a broad program of research whose goal is to provide the knowledge, currently missing, that will allow us to effectively integrate speech into the computer interface. We have identified a number of areas that we believe are of crucial importance to the acceptance of speech as a standard modality of human-computer communication. Our separate research areas are united by the common theme of eliminating the fundamental limitations of current speech recognition technology. We are currently working in the following areas:

- **Improved Recognition Techniques** — We are beginning the development of a 5000-word, speaker independent, connected speech recognition system. To maintain high accuracy on these larger tasks, we are pursuing a number of different strategies. We are investigating better subword models, improved learning algorithms, and rapid configuration for new vocabularies and tasks. We are also collecting a large database of task-independent speech and investigating more efficient ways of handling larger amounts of training data.
- **Fluent Human/Machine Interfaces** — We are investigating the utility of speech in several day-to-day interactive tasks. We have developed an effective interface for complex problem-solving applications using spoken language, and implemented a real-time spoken-language system for the Office Manager (OM) system. We have also developed strategies that will allow users to dynamically modify tasks by adding new words and grammatical constructs. Our plans include extending the OM system and deploying it within the group to study spoken-language interaction for meaningful daily tasks like mail, management, appointment scheduling, and personal database manipulation.
- **Acoustical and Environmental Robustness** — We have recently developed algorithms that deal with several classes of variability in the speech signal, including noise-subtraction algorithms based on traditional approaches that substantially eliminate stationary noise interference. More recently, we have developed techniques that allow us to approach comparable recognition performance from a variety of microphones, including table-top microphones, without the need to specifically train for each microphone. This latter result has brought us close to microphone independence.
- **Understanding Spontaneous Spoken Language** — Moving beyond small languages and rigid syntax, situations in which the user cannot (or will not) learn a restricted command language, requires the use of sophisticated parsing techniques that can deal with ill-formed speech. Ill-formedness can be mispronunciation, agrammaticality, the presence of restarts, as well as more mundane phenomena such as interjections (um's and ah's) and pauses. We have developed a frame-based parsing approach that has been remarkably successful in parsing ill-formed input and in dealing with a variety of common natural-language phenomena such as anaphora and ellipsis. We have successfully integrated this parser into ATIS, the Air Travel Information System, and OM, the Office Manager system.
- **Dialog Modeling** — Real-world tasks have numerous constraints that allow us to predict successfully the course of an interaction. We have been extending the work begun with the MINDS system to additional domains and have successfully demonstrated that similar dialog-level constraints can be applied to recognition. We are currently focusing on the ATIS domain to develop dialog models which have the potential to reduce perplexity by an order of magnitude.