

Speech Research at Carnegie Mellon

Principal Investigator: Raj Reddy

Carnegie Mellon has been active in speech recognition research for over 20 years. Research in the 70's have led to important systems such as Hearsay, Harpy, and Dragon. Since 1980, the Carnegie Mellon speech group has been exploring techniques for speaker-independent recognition. Various avenues were explored, including a knowledge engineering approach to identify robust speech features that are independent of speaker and environment, and a statistical learning system that utilized human knowledge and detailed speech modeling.

The latter system, Sphinx, was demonstrated in 1987. It is the first recognition system that could accurately recognize continuous speech from a large vocabulary by any speaker, without training. This was accomplished through the availability of ample training, the use of a powerful learning algorithm, and the design of detailed speech models. Currently, Sphinx has a word accuracy of about 96% using a word-pair grammar with perplexity 60. In 1988, Sphinx was ported to the Beam search acceleration machine, which made real-time recognition a reality. In 1988, the Minds System was demonstrated. Minds used Sphinx and Beam, and demonstrated the utility of dialog and semantic knowledge in improving recognition accuracy.

The overall goal of speech research at Carnegie Mellon is to develop new technologies that address the major problems currently inhibiting automatic speech recognition in realistic environments. Areas for future work include:

- **Improved Recognition Techniques** — We plan to develop a 5000- to 20,000-word vocabulary, speaker independent, connected speech, spoken English recognition system with grammar perplexity of approximately 200. To obtain a high accuracy on such a difficult task, we must improve Sphinx substantially. We plan to investigate better subword modeling, improved learning algorithms, and rapid configuration for new vocabularies and tasks.
- **Fluent Human/Machine Interfaces** — We will investigate, measure, and improve the utility of speech in several day-to-day interactive tasks that involve collaborative, human/computer problem solving.
- **Acoustical and Environmental Robustness** — We will improve the robustness of Sphinx against changes in microphone, noise characteristics, and environmental acoustics. Areas of research include employing multiple microphones, noise-subtracting, and homomorphic filtering.
- **Understanding Spontaneous Spoken Language** — We will work on the problematic phenomena associated with spontaneous speech, such as repetitions and restarts, speech-like noise (“um”s and “ah”s), and interruptions. The “missing science” includes techniques for understanding fragmentary input and for parsing in the presence of errors.
- **Task Semantics and Pragmatics** — Our recent research on the Minds system has shown that effective use of knowledge from dialogue, task-level semantics, and user models can reduce perplexity over ten fold. In the future, we intend to demonstrate the utility of such techniques in a wide range of domains.