

Analysis and Symbolic Processing of Unrestricted Speech

M. Margaret Withgott, Ronald M. Kaplan, Principal Investigators

XEROX PALO ALTO RESEARCH CENTER
3333 Coyote Hill Road
Palo Alto, CA 94304

This is a basic research project whose thrust is both theoretical and practical in nature. The core technology consists of techniques using machine learning and statistical theory as well as fundamental linguistic and phonetic theory. The investigations aim at furthering the understanding of requirements for future speech recognition systems, and in developing strategies for extracting significant information from noisy and/or large quantities of language data.

Both the theoretical and practical sides of the research have demonstrated advances. Phonetic regularities have been discovered; phonetic processing architectures and parameter tracking methods, and the CDT algorithm have been developed, all of which take into account contextual factors associated with phonetic variation. Along with progress in variation, a distance metric has been developed for co-channel speech-interference. Finally, progress has been furthered in understanding information extraction in unrestricted language data, and part-of-speech annotation has been demonstrated. The most recent accomplishments include:

- Developed the Clustered Decision Tree algorithm (an n-ary classification induction method) which makes use of machine learning and statistical techniques to organize data into structures representing the contextual factors associated with phonetic variation. [see Chen *et al.*, DARPA 89 Feb. & Oct. Proceedings]
- Using the CDT methodology, developed a program to create probabilistic pronunciation models in the SRI RULE format.
- Discovered that while identically transcribed phones from different phoneme sources may not differ spectrally, they can differ temporally; and an account of the phenomenon was developed [Peet & Withgott, reported at 116th meeting of the Acous. Soc. Amer.]
- Developed an LPC-based distance metric for recognition in the presence of competing speech in which target-interference separation and target recognition are performed simultaneously by matching subsets of LPC predictor roots; and achieved error reduction of 70% at low to moderate signal-to-noise ratios as compared with conventional whole-spectrum matching in speaker-dependent isolated-word recognition experiments. [see Kopec & Bush, ICASSP 89]
- Applied Markov random fields to (1) extract speech formants without appeal to a fixed number of expected resonant frequencies, and (2) “restore” formants, employing continuity constraints to allow the missing and noisy formants to be filled in.
- Developed automatic annotation for text using ordinary parts of speech and simple long-distance dependencies, without reliance on handmarked training data or upon uniformly higher-order Markov models; and achieved 95% correct annotation in a test text unrelated in form or content to the training document. [see Kupiec, DARPA 89 Feb. & Oct. Proceedings]