

CONSISTENCY MODELING

*Hy Murveit, Principal Investigator
Vassilios Digalakis, Peter Monaco, Mitch Weintraub*

SRI International
Speech Technology and Research Laboratory
Menlo Park, CA 94025

PROJECT GOALS

SRI's consistency modeling project began in August 1992. The goal of the project is to develop *consistency modeling* technology. That is, we aim to reduce the number of improper independence assumptions used in traditional speech recognition algorithms so that the resulting speech recognition hypotheses are more self-consistent and, therefore, more accurate. Consistency is achieved by conditioning HMM output distributions on state and observations histories, $P(x/s,H)$. The goal of the project is finding the proper form of the probability distribution P , the proper history vector, H , and the proper feature vector, x , and developing the infrastructure (e.g. efficient estimation and search techniques) so that consistency modeling can be effectively used.

RECENT RESULTS

Highlights of our accomplishments to date include a large reduction in our speech recognition error rate due to the development on Genonic HMM technology, and the development of a real-time version of our system. A summary of our accomplishments include:

- SRI developed and refined Genonic HMM technology. This is a form of continuous density HMM that, combined with other advances, allowed us to reduce our error rate by over a factor of two over the past year. Currently our best performance is 9.3% word error as measured on ARPA's 20K Nov. 1992 evaluation set and 13.6% on ARPA's 20K Nov. 1993 test set, both using ARPA's standard grammars.
- SRI developed an information-theoretic framework for estimating the effect of the history H in the conditional HMM output distribution $P(x/s,H)$ when H is constrained to be a previous frames x_{t-i} .
- SRI implemented a version of continuous local-consistency modeling. SRI verified that the information-theoretic framework above indeed predicts recognition accuracy improvements. We

measured performance of continuous local consistency for several different frame lags.

- SRI developed progressive search: a framework for using hierarchies of recognition algorithms in order to achieve fast yet accurate speech recognition.
- SRI developed tree-based search schemes for implementing large-vocabulary speech recognition systems. This resulted in a real-time 20,000 recognition system with about 30% word error.
- SRI developed Gaussian shortlist technology and other techniques for avoiding Gaussian distribution evaluation. This resulted in a net decrease of Gaussian evaluations by a factor of 30 with no recognition accuracy degradation. By combining this with the above search technologies, we expect implement a real-time near full-accuracy speech recognition in the near future.
- SRI transferred, improved, and evaluated feature mapping technology developed on NSF funding. This allows our system to be virtually microphone independent for large classes of microphones. For instance, using models developed for a Sennheiser close-talking microphone, accuracy degraded only 10% (5.9% to 6.4% error) when tested on data recorded with an Audio Technica desk-mounted microphone.

PLANS FOR THE COMING YEAR

Our plans for the coming year include:

- Advancing the state of our Genonic HMM technology, including incorporating segmental features to improve the consistency of hypotheses.
- Implementing real-time near full accuracy HMM systems by combining tree-search, multi-pass and shortlist technology.
- Further improving and evaluating feature-mapping microphone-independent HMM technology.