RESEARCH IN LARGE VOCABULARY CONTINUOUS SPEECH RECOGNITION*

Janet Baker, Larry Gillick, and Robert Roth

Dragon Systems, Inc.
320 Nevada St.
Newton, MA 02160

PROJECT GOALS

The primary long term goal of speech research at Dragon Systems is to develop algorithms that are capable of achieving very high performance large vocabulary continuous speech recognition. At the same time, in the long run we are also concerned to keep the demands of those algorithms for computational power and memory as modest as possible, so that the results of our research can be incorporated into products that will run on moderately priced personal computers.

RECENT RESULTS

Much of the past year's effort has been devoted to work on speaker independent training, linear discriminant analysis, and acoustic modeling, using the Wall Street Journal corpus as our development vehicle, with the goal of attaining very high accuracy large vocabulary SI CSR. In the past, Dragon had focused primarily on speaker dependent and speaker adaptive recognition, so that speaker independent research was a new departure for us in this past year. Similarly, in the past Dragon had confined itself to recognition algorithms that were highly parsimonious in both computation and memory usage, but we have now, temporarily, dropped those constraints in the interest of seeing just what is attainable with greater computational resources.

Our research in the area of speaker independent training has focused on the use of tied mixture models with multiparameter streams in representing the effect of context on the acoustic realization of a phoneme. These models are built using Bayesian smoothing in the context of the EM algorithm. The prior distribution for a tied mixture model for a phoneme in a specific context represents our opinion about that model based on other more generic models for that phoneme that have typically been built from much more data.

An important theme of our research has been the exploration of the tradeoff between the greater ability to model the dependence among acoustic parameters when streams are high dimensional versus the greater acoustic resolution possible in streams with fewer parameters. Another important thread has been the investigation of a variety of strategies for building the basis components for the mixtures.

The use of IMELDA, a particular form of linear discriminant analysis, has played a crucial role in our development in the last year as a way of eliminating some of the interspeaker variability represented in the parameters and, perhaps as a consequence, some of the dependence among our parameters, thus reducing the need for high dimensional streams. IMELDA also was used for the purpose of eliminating some of the variability due to the microphone in the DARPA "stress test". A further beneficial effect of this method lay in the reduction of the number of signal processing parameters (from 32 to 16) and thus the size of the acoustic models.

The need to retrain our system many times during the course of our research led us to re-engineer the training so that models for different phonemes could simultaneously be built on different computers.

PLANS FOR THE COMING YEAR

Dragon plans to continue to focus on improving the overall quality of the acoustic models for large vocabulary speaker independent continuous speech recognition. We intend to enhance the set of acoustic parameters which are an input to the IMELDA transform, which is likely to continue to play a key role in our system. We will again cluster the output distributions of the states into PELs, or phonetic elements, as our earlier systems did, both for the purpose of reducing the size of the models and for the purpose of smoothing noisy estimates based on insufficient training data. Another direction for exploration will be the use of mixture models without tied basis components.

We also intend to investigate cross-word modeling more intensively, through the use of the position of the word boundary as part of the context of a triphone model, and possibly through the use of probabilistic phonological rules.

*This work was sponsored by the Defense Advanced Research Projects Agency under contract number J-FBI-92-060.