Speech Representation and Speech Understanding

William S. Meisel
Speech Systems Incorporated
18356 Oxnard Street
Tarzana, California 91356
(818)881-0885

OBJECTIVE

The core objective of the research is to encode speech into segments which retain necessary information for accurate continuous speech recognition, but are more efficient to deal with than the usual encoding of short frames of speech. We use a multi-stage decision-tree encoder with linear combinations of features at the decision nodes; the result is segments which cover multiple frames and which are coded with the terminal node number of the final tree.

Additional contract objectives include showing that application knowledge can be efficiently applied to these codes to produce accurate transcriptions. The Knowledge Systems Laboratory at Stanford University is to help test the result in an application-development environment. The contract is to produce results which can be employed in a commercial system at the end of the contract.

SUMMARY OF ACCOMPLISHMENTS

We showed that segmenting and coding speech using SSI's phonetic encoding significantly improves both speed and accuracy for a system using Markov modelling.

We reduced utterance error rate by 25.7% with a further 40% increase in speed by reducing the number of ways words were spelled in the dictionary and by re-defining the phonetic classes.

The word error in decoding of phonetic codes into words was further decreased by a typical 20% using a penalty that reduced the erroneous insertion of small words.

Speed of recognition was further increased by a factor of two by using a more efficient structure in the decoding software.

Software was modified to provide access to transcriptions other than the best guess (e.g., the second through tenth best guesses) to aid the user in making corrections.

Software was also modified to give application developers access to semantic knowledge inherent in the structure of the language model used in the recognition process; for example, various names that a radiologist called a tumor seen in an xray (mass, density, tumor, etc.) would all be labelled "tumor."

PLANS

Larger segments and other variations will be tested in a Markov model environment.

Contract improvements in SSI's Phonetic Engine will be integrated, tested as a whole, and performance reported on RM data.