Robust Continuous Speech Recognition Technology Program Summary *

Clifford J. Weinstein and Douglas B. Paul, Principal Investigators

Lincoln Laboratory, MIT
Lexington, Ma. 02173-9108

PROGRAM GOALS

The major objective of this program is to develop and demonstrate robust, high performance continuous speech recognition (CSR) techniques focussed on application in Spoken Language Systems (SLS) which will enhance the effectiveness of military and civilian computer-based systems. A key complementary objective is to define and develop applications of robust speech recognition and understanding systems, and to help catalyze the transition of spoken language technology into military and civilian systems, with particular focus on application of robust CSR to mobile military command and control. The research effort focusses on developing advanced acoustic modelling, rapid search, and recognition-time adaptation techniques for robust large-vocabulary CSR, and on applying these techniques to the new ARPA large-vocabulary CSR corpora and to military application tasks.

BACKGROUND

The Lincoln program began with a focus on improving speaker stress robustness for the fighter aircraft environment. A robust HMM isolated-word recognizer (IWR) was developed with very high performance under stress conditions. The robust HMM techniques were then developed and extended to large-vocabulary CSR with state-of-the-art performance for both speaker-dependent (SD) and speaker-independent (SI) tasks on the ARPA resource management (RM) database.

More recently, the HMM CSR has been extended to tasks with much larger vocabularies (5,000 - 64,000 words) and higher perplexities (50 - 250) with focus on the Wall Street Journal (WSJ) corpus. Improved acoustic modelling techniques for the tied-mixture CSR have been developed and applied, including multiple observation streams, semiphone models, sex- dependent acoustic models, cross-word triphone models, and improved duration modelling. For the large-vocabulary tasks, the Lincoln HMM CSR has been converted to use a stack decoder search algorithm with integrated acoustic fast match and detailed match algorithms.

RECENT ACCOMPLISHMENTS

Developed and improved the Lincoln large-vocabulary tied-mixture HMM CSR, including stack decoder search, acoustic fast-match, and cross-word and sex-dependent acoustic models, and applied this CSR in the November 1993 evaluation tests; the new system showed a 42 percent improvement in error rate compared to the November 1992 evaluation test system.

Developed and successfully tested recognition-time adaptation techniques for large-vocabulary CSR in the November 1993 evaluation tests.

Developed tests on data-driven allophonic tree clustering smoothing techniques for best use of available training data.

Developed Bayesian smoothing techniques for triphones and obtained promising initial results on CSR corpora.

Continued contributions to ARPA CSR corpus development and evaluation, including contribution of stochastic language models to all sites for the 1993 evaluation tests; provided the ARPA CSR community with text processing software tools for large-vocabulary corpus development.

Organized and chaired the ARPA Spoken Language Technology and Applications Day (SLTA 93), which has produced very promising results in catalyzing technology transition of spoken language technology into military and civilian applications.

PLANS

Continue to develop large-vocabulary stack decoder-based HMM CSR, with particular focus on improvement of acoustic fast-match techniques,

Develop advanced acoustic modelling techniques including data-driven decision-tree-based triphone smoothing.

Develop run-time adaptation techniques for both acoustic HMM parameters and for stochastic language model parameters; include adaptation to speaker, channel, environment, and task.

Continue to define and develop spoken language technology applications, with particular focus on recognition and understanding of spoken messages in a command and control environment; also continue follow-up on other application opportunities produced by SLTA 93.

Chair the 1994 ARPA Human Language Technology Workshop.

459