SESSION 14: NEW DIRECTIONS/APPLICATIONS

Richard Stern, Chair

Department of Electrical and Computer Engineering
and School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

As in recent years, the last session of the Workshop focussed on new directions and unusual applications of spoken language technology. Five papers were presented.

One of the highlights of the meeting was a presentation by Julie Payette, an astronaut for the Canadian Space Agency, who discussed the use of speech recognition in space travel. Some of the reasons why voice input and output are potentially valuable in space include the naturalness of the speech modality, the need to have hands and eyes available for performing applications tasks, and the overwhelming number of displays and controls to be interrogated and manipulated. The harsh environment of space and the severe consequences of failure demand that designers produce systems that are highly accurate, reliable, and robust.

Payette described a pilot experiment in which a commercial speaker-dependent isolated-word recognizer was used to control the position and functions of the closed-circuit television system. While subjective reports by the users were favorable, it is clear that accuracy and reliability will have to increase for speech to be used to control mission-critical functions. Recognition accuracy did not appear to be adversely affected by the microgravity environment in space.

The second paper, presented by Suzanne Liebowitz Taylor of Unisys, described recent progress in Unisys's ongoing work on intelligent document understanding. The long-term goal of this research has been to develop methodologies for analyzing, classifying, and summarizing text from printed documents for which on-line versions are not available.

This paper discussed three ways in which natural language understanding techniques was used to augment image analysis. First, a combination of string-matching techniques, simple grammars, and statistical analysis of syntactic structure were used to re-integrate text which had been fragmented into physically-separated segments, as is commonly the case for stories that are written for popular magazines. Second, the PUNDIT natural language system was employed to correct errors introduced in the optical character recognition process. The use of natural language reduced error rate by more than 15 percent in an ATIS-like task for scanned text, compared to the error rate obtained using spelling correction alone. Third, case-frame parsing was used to provide semantic analysis of scanned documents. Two applications were described in the areas of text retrieval and hypertext generation.

The next two papers, from SUNY Buffalo and BBN, each describe ways in which speech recognition technology has been applied to the automatic recognition of handwritten text. John Makhoul described an extremely interesting demonstration by Thad Starner and colleagues at BBN that "with essentially no modification, a speech recognition system can perform accurate on-line handwriting recognition".

Starner and colleagues used conventional HMM techniques to recognize continuous cursive writing on a writer-independent basis. The features used for classification included the temporal evolution of the writing angle and its derivative, as well as changes of pen position, and identified pen up/pen down events. Homologs for handwriting analysis were developed for the familiar phonetic models, representations of context dependencies, and statistical grammars. The system was trained and tested on written sentences derived from the ATIS and WSJ tasks. It was found that the use of both context and statistical grammar provided marked improvements to recognition accuracy. Observed word error rates were 1.1 percent for the 3050-word, 52-symbol ATIS task, and 4.1 percent for the 25,000-word, 86-symbol WSJ task, using a version of BYBLOS with virtually no fine tuning for cursive handwriting.

The work described by Rohini Srihari of SUNY Buffalo focuses on the improvement to recognition accuracy for handwriting that can be obtained by application of two rule-based and statistical syntactic analysis procedures. The first procedure is based on the mutual information associated with word collocations within a phrase. The use of collocation information increased provided a 16.7-percent relative decrease of errors among top-choice word candidates. Under some circumstances this information can also be used to insert highly-probable visually-confusible alternates to the hypothesized words. The second approach made use a statistical model of syntax based on part-of-speech (POS) tags. The results reconfirm that the use of statistical grammars can significantly improve recognition accuracy.

The final paper of the Workshop, presented by Steve Lowe of Dragon Systems, concerned the use of large-vocabulary speech recognition systems to perform language identification, based on cumulated likelihood scores. To reduce the confounding variability introduced by differences in the quality of the acoustic match, normalized scores are obtained by dividing by the best-possible acoustic score on a frame-by-frame basis. This strategy produced good results in performing English-Spanish discriminations for high-quality speech in WSJ-type domain, but the results of subsequent experiments using telephone speech and a less restricted domain were more ambiguous. It is believed that this approach remains a promising one and work is continuing on the topic.