

SESSION 7: DEMONSTRATIONS

Victor Abrash, Chair

SRI International
Speech Research and Technology Program
333 Ravenswood Avenue
Menlo Park, CA, 94025

This year, seven sites presented ten demos and one video, showing the continuing progress of speech and natural language technology research and application. Papers were optional for this session.

Jack Mostow, from CMU, began the evening by showing a prototype reading coach designed to help children read by listening to them reading aloud. The system follows along as the student reads, detecting when words are not read correctly or when the reader gets stuck. Reading disfluencies are evaluated by a rule-based pedagogical evaluation component, which chooses appropriate interventions to help the reader. These interventions include asking the reader to re-read a word or a phrase, pronouncing a difficult word to the reader, ignoring the error, giving the reader a chance to correct the mis-reading in the context of a fluent reading of the initial part of the sentence, and reading the whole phrase to the reader. The system currently runs on a combination of DEC-Alpha and NeXT computers, with adult voices reading and mis-reading children's stories.

Alex Hauptmann, also from CMU, presented the latest demonstration version of their ATIS system. Based on the general purpose speech recognition system Sphinx-II, it ran completely in software on a DEC Alpha 3600 workstation. CMU claims that their implementation of the ATIS task is unusually robust to the quirks of spontaneous speech.

Lyn Bates of BBN demonstrated speech-driven database access in a non-ATIS domain. This application illustrated the ability to quickly port BBN's SLS technology to a domain completely unrelated to ATIS. The new domain consisted of a Sybase database containing information about Air Force facilities, equipment, and their features and status. Images as well as text were shown in response to user queries. This led to some discussion on useful tools to facilitate porting both the speech and language understanding components of SLS systems.

BBN also showed a speech interface to the Navy's JOTS system, developed to demonstrate the use of ARPA speech recognition technology in an operational military setting. JOTS is an application that provides the ability to enter, update, and display tracks of interesting objects, such as ships or aircraft. This demo provides a speech-recognition upgrade to JOTS, specifically in terms of entering and updating track position reports, the operations which consume the largest fraction of the operator's workload. The system does not remove any keyboard or mouse functionality, but simply adds the ability to set any field by

speaking the field name and value, using an active vocabulary of a few hundred words. This illustrated that speech can be used as a useful adjunct to an interface that operators already understand; BBN reports that it was well received by JOTS operators.

Richard Schwartz demonstrated the BBN Wall Street Journal dictation system. This year, BBN increased the size of their vocabulary to over 40,000 words. They also upgraded their grammar training by adding three more years of WSJ text (1990-1992), the spontaneous journalist training data, and other words currently in the news. According to John Makhoul, these steps greatly reduced the frequency of out-of-vocabulary words, making it more likely that the language model would cover stories about recent events. As in the past, their recognizer performs an approximate fast match in the forward direction, followed by a fast, accurate backwards pass.

Patti Price of SRI demonstrated a spoken language translation system for English to Swedish translation in the air travel planning domain. This demo, using a modular spoken language translation architecture, was developed under sponsorship from Swedish Telecom by a collaboration between SRI International, the Swedish Institute of Computer Science (SICS), and Telia Research. This system produces a synthesized Swedish translation of ATIS utterances, and consists of an English language speech recognizer (SRI's DECIPHER ATIS system), a natural language component which produces a language-independent intermediate form, and a translator to produce Swedish text which is sent to a text-to-speech synthesizer. State-of-the-art performance was demonstrated after only modest development effort. Although current technology does not provide for unconstrained translation between languages, this demo indicates that recent advances allow us to envision realistic translation applications in specific domains.

Victor Zue and the MIT-LCS Spoken Language Systems Group demonstrated GALAXY, a system enabling users to access and manipulate various sources of information using spoken input in order to solve specific tasks. GALAXY focused in general on the travel domain, including air travel, local navigation, and weather, building on MIT's past experience in the ATIS, VOYAGER, and PEGASUS domains. The demo made use of several real databases distributed on the information highway, including American Airlines' EAASY SABRE, the NYNEX Yellow Pages, the US Census Bureau's map database, and the National Weather Service database, and was designed to promote modularity, flexi-

bility, portability, and scalability. MIT believes that their strategy of integrating many different knowledge sources will allow them to uncover critical new technical issues, including new word detection and learning, portability across domains, and dialogue modeling. Furthermore, they believe that working on real applications has the potential benefit of shortening the interval between demonstration of a technology and its ultimate use, and that real applications which help people solve problems will be used by real users, thus providing a rich and continuing source of useful data.

Bill Ogden and Jim Cowie, from New Mexico State University, presented Tabula Rasa, an interactive design tool and code generator for machine assisted human information extraction. Intended to put state-of-art TIPSTER technology into the hands of today's analysts, this tool allows them to define a new domain and to produce the matching information extraction tool in minutes. It is an attempt to reduce two of the major bottlenecks of information extraction: the definitions of the text extraction task and the production of automatic extraction tools to aid the human analyst in the production of structured data. In the demo, they showed how easily the system could be used to generate new information extraction tools incorporating good human factors with existing automatic extraction capabilities.

Carl Weir of Unisys demonstrated ASWIS, a system for message processing and data fusion. This illustrated the use of message processing technology for monitoring the movements of submarines, along with other types of naval platforms. Normally, message traffic pertinent to an object being tracked is ignored in the data fusion process, even though it could be a valuable source of information. ASWIS demonstrates the use of a text understanding (data extraction) sensor which detects references in the message traffic to submarines arriving or departing from ports, providing a fix on their locations.

Suzanne Taylor, also from Unisys, presented IDUS, an intelligent document understanding program. Starting from the bit-mapped image of a document, IDUS creates the data for a text retrieval application. This demo employed many different technologies, ranging from image understanding and optical character recognition (OCR), to document structural analysis and text understanding, in a knowledge-based cooperative fashion. The IDUS system showed the power and utility of an integrated set of document interpretation modules, all accessible via the standard X-Windows/Motif user interface.

Finally, Xuedong Huang from Microsoft showed a video of Microsoft Whispers, their speech recognizer, running under Windows on an Intel 486-based PC. The video showed two separate applications: a voice command-and-control system for a Windows-based PC, and a speech recognition interface to an intelligent agent system. The first points out an important application which will undoubtedly help introduce the public at large to speech recognition. The intelligent agent interface, complete with an animated talking bird rendered in high-resolution computer graphics, combined speech synthesis, recognition, and understanding to enable users to negotiate with the agent to ask for and play songs stored in a jukebox.