

SESSION 6: DEMONSTRATIONS AND VIDEOTAPES OF SPEECH AND NATURAL LANGUAGE TECHNOLOGIES

Mari Ostendorf

Boston University
ECS Department
44 Cummington Street
Boston, MA 02215

In this session, several sites presented videos or demos to illustrate research progress and demonstrate operation of their spoken language system. Since papers were optional for this session, the proceedings do not completely reflect the accomplishments that were reviewed in this session.

Richard Lyon, from Apple, showed a video imaging speech through a correlogram: a signal representation based on a cochlear model. Though the video is based on simulations on a Cray, the algorithm is currently being implemented in analog VLSI for real-time signal processing. The associated paper in this proceedings describes the VLSI circuit implementation.

Harvey Silverman, from Brown, showed a video illustrating signal and noise separation using microphone arrays. Such algorithms can improve speech recognition performance in noisy environments and free the user from the close-talking microphone. Not surprisingly, algorithm performance in a realistic environment with reverberation noise is not as good as the theory predicts, and much research remains in this area.

Paul Bamberg of Dragon Systems demonstrated their connected word recognition system in two domains: radiology and Resource Management. The system runs on a PC with a special purpose signal processing board and was trained on a database which includes speech from very diverse sources.

Pat Peterson from BBN showed a video illustrating: 1) their real-time spoken language system HARC which uses Byblos, the speech recognition system, to provide the top N sentence hypotheses for natural language processing; 2) dialect normalization through speaker adaptation which results in dramatic recognition performance improvements for non-native English speakers and native speakers with strong accents; and 3) demonstrating how integration of HARC into the BBN DART (Dynamic Analytical Replanning Tool) project can allow faster user access to information than through a mouse alone. A paper in this proceedings describes the use of spoken language in the DART system.

Mitch Weintraub of SRI International demonstrated his noise robust signal processing algorithm in a digit recognition task. He was able to switch between three different microphones - a close-talking mic, a hand-held mic and a table-top mic - with no loss in recognition performance. Patti Price and John Butzberger demonstrated the SRI ATIS system. The system uses

a PC with a DSP board for signal processing, a Sparc station for HMM speech recognition with includes a bigram Markov language model, and a second Sparc station for natural language processing using the template matching grammar. The system used in this demo runs in real time using a perplexity 10 grammar; the benchmark system has a higher perplexity with a 1-2 minute response time.

Victor Zue and Stephanie Seneff demonstrated the MIT ATIS system as used for data collection, specifically operating the system in the flight booking mode. The system involves cooperative computer/human interaction working toward the goal of filling out information in a ticket in data collection mode. They pointed out that we do not yet know how to collect spontaneous speech data and that we should experiment with different procedures. A paper describing their data collection procedure and analysis of different ATIS corpora appears in the proceedings.

Alex Rudnicky showed a video illustrating the CMU Office Management spoken language system, based on the Sphinx recognition system and a frame-based parser. The system uses multi-modal input (mouse, text, and different modes of voice input) to control various tools including a personal information database, voice mail, an appointment calendar and a calculator. The goal of working with this task domain is to study a large user population and a complete human/machine interface. CMU considers task completion time is an important measure of system performance.

Ralph Weischedel, from BBN, showed a video produced to illustrate the DARPA Program on Natural Language Processing which is aimed at developing technology for enabling machines to process text intelligently. Because of the tremendous growth in volume of data, the ability to automatically extract and process relevant information in messages is becoming an important technology. Natural language processing offers the potential for automatic database update, query and retrieval, and message routing, prioritization, fusion and alerts. The video showed that, although today's natural language systems are limited to constrained domains, they are quite successful within those constraints.

Papers were optional in this session, due to the difficulties in translation from the different media.