COMBINING LINGUISTIC AND STATISTICAL TECHNOLOGY FOR IMPROVED SPOKEN LANGUAGE UNDERSTANDING

Michael Cohen and Robert Moore, Principal Investigators

SRI International
Menlo Park, CA 94025

PROJECT GOALS

The goal of this project is to develop technology for spoken language understanding which is highly accurate, robust, and fast, is easily ported to new domains, environments, and channels, and quickly adapts to new speakers. The system combines the DECIPHER speech recognition system with the Gemini natural language understanding system.

RECENT RESULTS

SRI has developed a spoken language interface to the Official Airline Guide (OAG). Despite a funding gap for more than four months of the year, substantial improvements have been made in the component technologies. On recent ARPA benchmarks, SRI achieved 5.5% word error on the ATIS speech recognition task, 18.2% utterance error on the natural-language understanding task, and 20.7% utterance error on the spoken-language understanding task. Other recent results include:

- Investigated several speaker-adaptation algorithms for both native and non-native speakers of English. The resulting techniques can match speaker-dependent performance (trained on 650 sentences) using 100 adaptation sentences, and outperforms the speaker-dependent system when more than 100 adaptation sentences are used.

- Developed an approach for constructing acoustic models for telephone applications using high-quality recordings, resulting in a substantial savings in effort when porting the ATIS application to a telephone environment.

- Developed methods to discriminate "hesitation" from "end-of-utterance" silent pauses based on durational and f0 correlates of preceding syllables. This can have important implications for the design of end-pointing algorithms.

- Performed a study of filled pauses which showed that they occur almost exclusively in between words in low-probability word sequences.

- Improved the modeling of out-of-vocabulary words and word-fragments.

- Developed a class-trigram grammar for ATIS, resulting in a 30% decrease in word error compared to a word-bigram grammar. Approximately half the improvement was due to the trigrams, and half to the classes.

- Developed methods to incorporate natural-language constraints supplied by the Gemini parser into the DECIPHER recognition search.

- Increased the speed of the Gemini parser by a factor of four by improved handling of semantic selectional restrictions.

- Expanded the scope of the SRI ATIS system for ATIS3.

- Assumed leadership of the effort to define a semantic evaluation methodology for spoken language systems, working out a detailed framework for annotation of the predicate-argument structure of utterances.

- Collected a total of 2863 ATIS3 training and test utterances (speech, transcriptions, and log files).

PLANS FOR THE COMING YEAR

In the following year we plan to continue to explore methods for integrating natural-language constraints into speech recognition systems, develop rapid speaker adaptation methods, improve the portability and scalability of the technology, and complete the development of a telephone-based ATIS system.