SPOKEN LANGUAGE SYSTEMS

PI: John Makhoul
makhoul@bbn.com

BBN Systems and Technologies, 10 Moulton St., Cambridge, MA 02138

OBJECTIVES

The objective of this project is to develop a real-time spoken language system capable of understanding and responding to spoken English commands and queries for interactive human-machine applications, such as battle management, command and control, and training of personnel on complex tasks. The system will also include a capability to adapt to new speakers and a capability to detect when a user says a new word, and to allow the user to add the word to the system.

ACCOMPLISHMENTS

Work in this area requires the integration of three technologies: large-vocabulary continuous speech recognition, natural language understanding, and system integration. In our work at BBN, we have integrated our BYBLOS continuous speech recognition technology with a new natural language understanding component, DELPHI, resulting in a complete spoken language system, called HARC (Hear And Respond to Continuous speech). A major accomplishment of this project has been the development of a real-time version of HARC, implemented completely on commercially-available hardware. An N-best version of BYBLOS (see below) running in real-time has been implemented on a Sun 4 with a Sky Challenger signal processing board. The DELPHI natural language component also runs on a Sun 4. The complete system has been interfaced recently to a DARPA-sponsored military logistical planning system, called DART (Dynamic Analysis Replanning Tool).

The DELPHI natural language component uses a Unification formalism for describing the syntax and semantics of English and for enforcing syntactic and semantic constraints. It uses a higher-order intensional logic for representing the meaning of a sentence. The system provides for the incremental application of syntax and semantics; advantages of this approach are that unproductive search paths are cut off more quickly, and any improvements in unification parsing apply automatically to semantics as well as syntax. We have implemented unification semantics for our grammar rules in four task domains: battle management, personnel information retrieval, airline travel information retrieval, and military logistical planning. We have interfaced and extended the JANUS discourse module, developed under an earlier DARPA effort, to the HARC system. We also developed a method for rapid porting of the natural language component to new task domains using the Parlance Learner^{TM} knowledge acquisition tool. Recent accomplishments in DELPHI include parsing speedups, streamlining the unification grammar, and introducing mapping units into the semantic processing. Syntactic and semantic parsing of a sentence now takes less than one second on average on a Sun 4.

One important contribution has been the development of the N-best search strategy for integrating speech and natural language components. This method produces the N highest scoring sentences that match an input utterance, aided by a statistical language model. The natural language component then searches these N sentences for the highest scoring sentence for which the system can produce a semantic interpretation. The N-best paradigm, by providing a clean and simple interface between speech and natural language, has found immediate acceptance as the method of choice in spoken language integration. An efficient two-pass (forward-backward) algorithm for obtaining the N-best sentences has allowed the implementation of the algorithm in real-time on a Sun 4.

In this project, we have been instrumental in the design of methodologies for the collection of spoken language data and the objective evaluation of spoken language systems. We previously helped specify the DARPA Resource Management Corpus that is now in common use for speech recognition evaluation. More recently, our proposals for the evaluation of spoken language systems have been adopted by the DARPA community.

We have developed what we believe to be the first successful method for the automatic detection of out-of-vocabulary words. This is an important problem for any realistic system with a large vocabulary. Initial results show a 70% detection rate with only 1% false alarm. We have recently developed a capability for the addition of new words to a speech recognition system.