Multilingual Conversational Speech Research
9807-11

Progress Report: January 1, 2001—June 30, 2001

James Glass and Stephanie Seneff

Project Overview

The long-term goals of this research project are to foster collaboration between MIT and NTT speech and language researchers and to develop language-independent approaches to speech understanding and generation. We will initiate this effort by developing the necessary human language technologies that will enable us to port our conversational interfaces from English to Japanese. The Jupiter weather information system will be used as the basis of this porting process. This work will involve the close collaboration with NTT researchers both in Japan and at MIT.

Progress Through June 2001

Over the last six months we have continued to refine the speech and language components of the Mokusei system and have deployed the first of two Mokusei systems to be located at NTT for data collection purposes. The following sections describe our activities in more detail.

System Deployment and Data Collection

In order to make the system more easily accessible to users in Japan, and thus collect more data from native Japanese speakers, we have planned to deploy the Mokusei system at NTT. The first of two such systems, including both hardware and software, was set up at the NTT Atsugi R&D Center. The hardware consists of two machines; the first (Windows) is used for the telephony interface, while the second (Linux) is used for the remainder of Galaxy servers. A harvesting program was set up to gather weather information from the same web sites used by the Jupiter systems deployed at MIT. In preparation for public deployment of the system, Mokusei now makes use of information that is commercially distributed by Weather News Inc., Japan. The system can now provide information for approximately 150 Japanese cities. All weather information is stored in an Oracle relational database residing at NTT.

The Japan-based Mokusei system is currently used by NTT employees. Thus far, there have been more than 500 calls that comprise 2,600 utterances, including expert users' data, which have been collected. We expect that the second NTT system will be set up during the summer of 2001 at the NTT Kyoto Lab. Researchers at NTT have

also initiated an effort to set up a toll-free number in Japan. A toll-free number in the United States will also be set up to allow Japanese speakers in the United States to talk to Mokusei.

Speech Recognition

We have continued to the refine speech recognizer by augmenting training data and improving the quality of the transcriptions. Currently, the acoustic and language models have been trained with 8,038 naive user utterances, 1,900 read speech utterances, and 2,592 expert user utterances. The recognizer has a vocabulary of 1,151 words and a trigram test set perplexity of 13.0. On average, there are 2.6 morae per word. The test set consisted of 2,442 utterances recorded from naive users. On the in-vocabulary test data of 1,745 utterances containing no artifacts, the word error rate is 8.5%, with a sentence error rate of 33.1%. Each sentence contains an average of 5.8 words. On the complete test set the word and sentence error rates increased to 19.0% and 45.9%, respectively.

Language Understanding

During this period, we have continued to improve the grammar coverage for language understanding without increasing the parsing time. We also incorporated the robust parsing technique, used by the MIT systems, that finds a partial parse when there is no full parse. The language understanding grammar currently has more than 500 categories and nearly 2,000 vocabulary entries. It provides either a full parse or partial parse for 79% of the naive user utterances that do not include artifacts. Overall understanding was measured on the 1,515 test set utterances whose transcriptions can be fully parsed. The concept error rate on these data is 12.0%, which is comparable to the Jupiter English counterpart.

Language Generation

Although response generation was not improved very much during this period, the spoken responses for the manually transcribed test set utterances were manually checked, and more than 90% of them were considered fluent.

Research Plan for the Next Six Months

This is the final report for this project. A new project will be initiated on July 1, 2001 on research and development of multi-lingual, multi-modal conversational interfaces.