MACHINE-AIDED VOICE TRANSLATION (MAVT) ADVANCED DEVELOPMENT MODEL[1]

Christine A. Montgomery
Bonnie Glover Stalls
Robert E. Stumberger
Naicong Li
Robert S. Belvin
Alfredo R. Arnaiz
Susan Hirsh Litenatsky

Language Systems, Inc.
{chris,glover,res,naicong,robin,arnaiz,hirsh}@lsi.com

1   Introduction

Machine-Aided Voice Translation (MAVT) is a development begun in 1990 under contract to Rome Laboratory, AFMC, for a spoken language translation prototype to assist Air Force personnel in interacting with speakers of foreign languages. The initial phase of the project, which concluded in 1992, resulted in the development of a speaker-independent continuous speech, translation system for English => Spanish => English, using a vocabulary of about 500 words. Evaluation of the testbed system was described at the AMTA workshop in San Diego in 1992, and a system overview is presented in the workshop proceedings [I].

This demonstration presents a snapshot of the Phase II MAVT ADM system, which shows voice input and output for English => Spanish => English, and text translation for Arabic and Russian, using a small sample of test sentences from the development corpus.

Like the Phase I system, the current system is comprised of three subsystems: a speech recognition system, a natural language processing system, and speech generators. Speech recognition is accomplished via Entropic's HMM Toolkit, while speech synthesis for English and Spanish utilizes Entropic's *TrueTalk^tm*, licensed from AT&T. As in the Phase I system, natural language understanding and translation generation is achieved via LSI's DBG natural language processing system. These three subsystems are briefly described in the following paragraphs.

2   The DBG Natural Language Processing System

LSI's DBG system has served as the NLP engine for a variety of text understanding applications, focusing on information extraction for data base generation (from which the acronym DBG is derived) for a range of different types of text, and message fusion, based on a large sample of transcribed radiotelephone traffic. The components of the DBG system as configured for these applications include modules for lexical lookup and morphological analysis, full syntactic and semantic analysis, and discourse or text-level analysis. The analyzed content of a text is represented as a set of interconnected frame structures called templates, which reflect the entities and events described in a source text.

For the MAVT application, modules were added to generate the target language text. In the Phase I MAVT development, a direct transfer strategy was used to achieve translation, although many of the components—e.g., the principle-based parser—were designed for multilingual use. In the current MAVT development, we have adopted an interlingual approach to translation, introducing Lexical Conceptual Structures (LCS's) into the DBG system as the language-independent part of the content representation of a text. (A detailed discussion of this innovative development is presented in our conference paper [2].)

3   ASR via HTK: an HMM Software Toolkit

Entropic Research Laboratory licenses this technology from the Cambridge University Technology Transfer Company, and is responsible for ongoing support of HTK and future enhancements. HTK allows flexible development and modification of speaker models (e.g., recognizers for different languages and applications) based on Hidden Markov Model (HMM) principles, for isolated, connected, or continuous speech recognition. The recognizer is syntax-driven, via a finite state grammar which is customized for a particular recognition task. In recent ARPA testing of speech recognition systems developed by ARPA contractors and others, the HTK-based system performed comparably with those of ARPA contractors on dictation tasks involving a 5,000 word vocabulary and a 20,000 word vocabulary derived from Wall Street Journal texts. On the 5,000 word task, the recognizer developed with HTK performed at 95% accuracy, performing at 87% for the complex 20,000 word dictation task. HTK is written in ANSI C, and runs on Sun, H-P, DEC, or SGI workstations under Unix.

In the initial version demonstration of the MAVT ADM, speaker-independent, continuous speech recognizers for a limited mission-oriented vocabulary have been developed for English and Latin American Spanish.

4   *TrueTalk^{tm}* **Text-to-Speech (TTS) Software**

TrueTalk^{tm} is an advanced software-only TTS system that converts digitized text into speech, with a word intelligibility rate of approximately 97%. Entropic licenses this technology from AT&T, where it has been in development over the past 10 years. *TrueTalk^{tm}* features a variety of user controls, including pitch, word duration, intonation, and speaking rate. For English, *TrueTalk^{tm}* uses a primary dictionary of 166,000 words, and a secondary dictionary to assist in accurate pronunciation of proper names, such as location designations. The Spanish vocabulary is of a comparable size. *TrueTalk^{tm}* runs on Sun, H-P, or SGI workstations under Unix.

References

[1] C. A. Montgomery, B.G. Stalls, R.E. Stumberger, N. Li, R. S. Belvin, and A. R. Arnaiz. Evaluation of the Machine-Aided Voice Translation (MAVT) System. In *Proceedings of the Workshop on MT Evaluation: Basis for Future Directions*. AMTA, 1992.

[2] B.G. Stalls, R. S. Belvin, A. R. Arnaiz, C. A. Montgomery, and R.E. Stumberger. An adaptation of lexical conceptual structure to multilingual processing in an existing text understanding system. In *Proceedings of the AMTA 94 Conference*. AMTA, 1994.