N94-32437

AUTOMATIC TRANSLATION AMONG SPOKEN LANGUAGES

Sharon M. Walter and Lt. Kelly Costigan
Rome Laboratory/IRAA
Griffiss Air Force Base, NY 13441-4114

ABSTRACT

The Machine Aided Voice Translation (MAVT) system was developed in response to the shortage of experienced military field interrogators with both foreign language proficiency and interrogation skills. Combining speech recognition, machine translation, and speech generation technologies, the MAVT accepts an interrogator's spoken English question and translates it into spoken Spanish. The spoken Spanish response of the potential informant can then be translated into spoken English. Potential military and civilian applications for automatic spoken language translation technology are discussed in this paper.

THE MACHINE AIDED VOICE TRANSLATION (MAVT) SYSTEM

During times of military conflict it is important to acquire intelligence information quickly. The best sources of timely information, however, are often foreign-language speaking people: defectors from the opposition's camp, Prisoners of War, and civilians from the conflict area. Whenever and wherever conflict occurs, military linguists who are versed in the particular language and dialect of potential informants, familiar with the Commander's military strategy, and knowledgeable about interrogation techniques are a valuable asset --- and an extremely rare commodity.

The Machine Aided Voice Translation (MAVT) system is an early prototype demonstration of the application of current speech processing technology to help compensate for the shortage of suitably trained and experienced linguists. It allows a less skilled interrogator to "screen" potential informants. When an interrogator with little or no foreign language skills asks questions by speaking into the microphone in English, the MAVT translates the questions into machine-spoken Spanish. Upon hearing each question in his/her native language, the potential informant responds by speaking into the microphone and the system translates the response into spoken English. Based on the interrogator's perception of an informant's cooperativeness, reliability of information, and relevance of the information to the Commander's intelligence requirements, the potential informant can be referred to a more skilled interrogator for further, deeper questioning.

The MAVT display is shown in Figure 1 on the next page. The user selects 'male' or 'female' and 'English' or 'Spanish' to indicate to the system the gender of the speaker and whether he/she is an English-speaking interrogator or a Spanish-speaking interrogatee. Providing the MAVT with knowledge of the gender of the speaker allows better speech recognition due to the more appropriate use of either a male or female speech "model."

Inputs to the system must be restricted to those that use words from the system's Spanish and English dictionaries (lexicons) and those which use a word order allowed by its grammars. There are two MAVT grammars. One allows questions and answers about biographical information, so examples in English are: "What is your name?" and "Indicate your unit designation."; examples in Spanish are: "Mi rango es teniente general" and "Naci' en Santa Clara."* The second grammar focuses on mission-related information such as (in English:) "Why was your unit moving out to the south?" and "Is the main force heading in that direction?", and (in Spanish:) "Proteger el puesto de comando del regimiento" and "Su misio'n es encontrar unidades americanas." The display includes a scrollable list of examples of the inputs that are accepted by the speech recognizer. When a Spanish speaker is anticipated the display lists examples of acceptable Spanish responses.

The MAVT prototype is serving as the design basis for a follow-on development that will extend the English-Spanish translation vocabulary and expand system capabilities to include English-Arabic and English-Russian spoken language translations. The follow-on system will be completed in late 1996.

Language Systems Incorporated, of Woodland Hills, California, developed the MAVT and is the primary contractor for the advanced development model.

Hardware/Software Architecture, Briefly

The speech recognition system of the MAVT is the PE200 Phonetic Engine produced by Speech Systems, Inc. (SSI). The Phonetic Engine accepts speaker-independent, continuous speech. That is, it does not have to be trained to recognize any particular voice, and users can speak quite naturally and fluidly without pausing between each word as would be required by an isolated word speech recognizer.

---

* English and Spanish examples are not intended to represent questions and their respective answers.

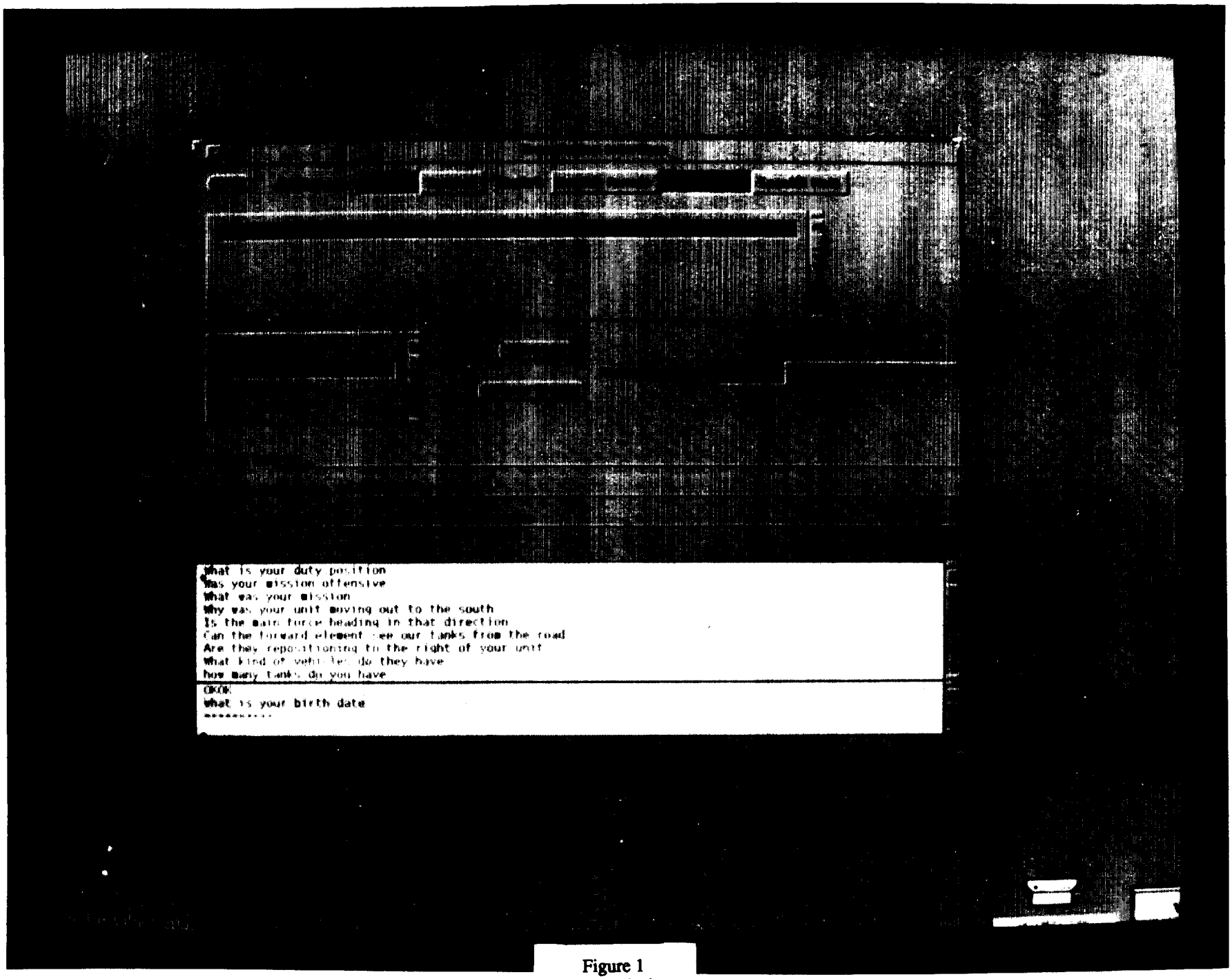

Figure 1
MAVT Display

MAVT speech output is provided by a DECtalk DTC01 speech generator from the Digital Equipment Corporation (DEC). The DECtalk accepts text output and converts it into spoken words. The pitch and speaking rate of the DECtalk may be modified.

The core, language translation software of the MAVT is LSI's DBG natural language processing system hosted on a Sun Microsystems Workstation (SPARCstation). DBG was extended for this project with a multilingual lexicon, a multilingual morphological component, and a language-independent syntactic parser. DBG works by deriving semantic (meaning) representations of inputs and translating them into the language which is to be output.

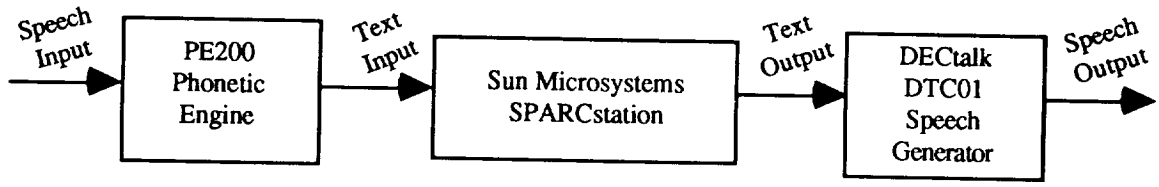

Figure 2
MAVT Architecture

DUAL-USE OF MAVT TECHNOLOGY

Other potential military applications of automatic spoken language translation include its use for multi-national military operations - facilitating communication among cooperative multi-lingual forces, and its use for deriving intelligence information from speech communications. This latter application requires a burdensome mix of knowledge-intensive skills similar to those involved in military interrogation since the analyst, or listener, must not only be able to listen to multiple lines of communication input and responsibly select and record information pertinent to the mission, but must also possess foreign-language skills when monitoring foreign-language communications.

The application of computer-based spoken language translation technology to language training can provide cost-effective augmentation and reinforcement of foreign language skills for inexperienced military linguists and civilian students. Students could request foreign language equivalents for spoken expressions in their native tongue or be encouraged to repeat phrases displayed on a screen or 'voiced' by the computer. The correctness of a student's attempt to speak in a foreign language would be determined by matching his utterance to the expected input. Instant reinforcement can be provided and intelligent computer prompting could guide students through difficult phrases. Adding other media, video for instance, to such a training system further extends the value of the technology as a training resource because students, cued by visual prompts, can then devise their own wording. Further, computer-based language training allows lessons to be varied depending on a student's level of competence.

The MAVT prototype has attracted the interest of law enforcement organizations and emergency room medical personnel. Most large metropolitan areas in the United States have many non-English speaking inhabitants, making communication and acquisition of information difficult in some cases and almost impossible in others. As a result of these communication problems law enforcement organizations such as the Los Angeles Police Department have expressed an interest in the use of automatic spoken language translation technology for interviewing crime witnesses, victims, and suspects. Hospital staffers have noted the value the technology holds for emergency room admittance of foreign-speaking patients. In a manner similar to the military application for interrogation, language translation technology would allow law enforcement personnel to communicate with citizens, and hospital personnel to communicate with patients, in their own languages without the time delay involved in locating an interpreter.

MAVT technology also finds application as a diplomatic, or business, briefing aid. Lack of a common language need not stand in the way, in the future, of international visits and communications among those with common political or business activities. Interest in the technology has been shown by a Texas organization eager to facilitate diplomatic interactions with representatives of the Mexican government.

Lastly, consider the value of the technology as a tourist travel aid. Currently, words or simple phrases can be typed into hand-held instruments that provide foreign-language translations, but how much better it would be to speak into a similar instrument and have it vocalize our intent in the language, or languages, of others.

TECHNOLOGY LIMITATIONS

The current state of the component technologies of the MAVT places limitations on the near-term application of computer-based spoken language translation. Some of the deficiencies of those technologies are identified here.

Speech recognition technology has improved substantially in just the last three years. Until very recently speaker independent continuous speech was still an out-of-range laboratory research goal, and the available isolated word speech recognition was not appropriate for very many applications. Current recognition capabilities will suffice, however, only for those applications for which every spoken input can be anticipated. Broader application of the technology will at least require the allowance of less restrictive, or more relaxed, grammars. In the meantime, the technology continues to advance rapidly.

Limitations set by the state of speech recognition technology on the breadth of the grammar actually make its use with text-based language translation viable since automatic translation technology is not yet capable of handling free-form text. Since every input is anticipated, correct translation of every possible input can fairly well be assured. Improvements in machine translation of text will be required in the future.

Broader application of speech translation technology will necessitate an improvement in intonation features offered by language generation systems. Computer-generated speech is currently robotic and monotone.

And finally, MAVT components are specifically tailored to operate with a specific language pair and within a specific domain. Tools to port spoken language translation technology to new domains and to new languages are needed.

REFERENCES

Montgomery, Christine A, Bonnie G. Stalls, Robert E. Stumberger, Naicong Li, Robert S. Belvin, Alfredo Arnaiz, Philip C. Shinn, Armand G. DeCesare, and Robert D. Farmer. Machine-Aided Voice Translation (MAVT). To be published as a Rome Laboratory Final Technical Report.

Montgomery, Christine A, Bonnie Glover Stalls, Robert E. Stumberger, Naicong Li, Sharon Walter, Robert S. Belvin, and Alfredo R. Arnaiz (May 24-27, 1993). Machine-Aided Voice Translation. In Proceedings of the 3rd Annual IEEE Mohawk Valley Section Dual-Use Technologies and Applications Conference, pp. 96-101.