SPEECH FOR EXPORT

Automating the translation of spoken words

By IVARS PETERSON

The caller at the other end of the telephone line speaks a language that's completely foreign to you, and you can't tell what she wants. For help, you could look for a co-worker fluent in the language, or you could turn to a commercial telephone service that connects you with an interpreter.

But Alex Waibel, a computer scientist at Carnegie Mellon University in Pittsburgh, has a more high-tech solution in mind. He envisions the development of a computer system that recognizes speech in one language, translates the spoken words into another language, and feeds the translated text into a speech synthesizer.

"Real-time translation of telephone conversations is an ambitious project," he admits. It requires the integration of three capabilities — speech recognition, machine translation, and speech synthesis — that by themselves present formidable difficulties.

In a demonstration staged last January, Waibel and his team, working with research groups in Germany and Japan, showed both the future promise of such technology and its present-day limitations. In a scripted, three-way conversation, independent systems — at Carnegie Mellon, at the Interpreting Telephony Research Laboratories of Advanced Telecommunications Research (ATR) in Kyoto, Japan, and at Siemens A.G. in Munich, Germany — went through the rigmarole of obtaining information and registering for an international conference.

Pronouncing his words distinctly and carefully, an English-speaking participant in the demonstration talked into a headset-mounted microphone connected to Carnegie Mellon's JANUS system. He said: "I would like to register for the conference." Seconds later, the voice synthesizer in Germany repeated: "Ich wuerde mich gerne zur Konferenz anmelden."

But if a speaker happened to stray from the script, going beyond the system's vocabulary of roughly 500 words, the computer would fail to produce a translation. "That's a problem," Waibel says. "We're now moving ahead, trying to break through these limitations."

Waibel seems just the right kind of person to be involved in this linguistic stew. He is fluent in both English and German, and his wife is Japanese. So he can readily check how well the machines are doing.

Beyond the intellectual challenge of a difficult research problem and his own interest in language understanding, Waibel also sees an unfulfilled need for technology that can aid communication among people speaking different languages. Although an increasing proportion of the world's population is learning English, these people are seldom really fluent in that language.

"Even in countries like Germany or Japan, people don't all speak English," Waibel notes. "There is actually a huge need for language translation."

As an example of the strong interest of users in having an interpreter available, Waibel cites the success of a relatively new enterprise known as AT&T Language Line Services. Since it started in the early 1980s, this telephone service — now offered 24 hours a day, seven days a week — has grown rapidly to encompass interpretation between English and 140 other languages. This requires a large staff of part-time interpreters, who work by telephone out of their homes at locations all over the United States.

Frequent customers of the service include hospitals, insurance companies, and all manner of government agencies — anyone regularly dealing with U.S. residents who do not speak English. The service is also used by large and small businesses interested in cracking international markets and even by individuals trying to communicate with foreign visitors. Spanish is the most requested language, followed by French, German, Italian, Chinese, Japanese, Korean, and Vietnamese.

At the same time, "human interpreters are very costly and may not be required for some routine things," Waibel notes. "If you want to talk poetry or do international peace talks, you would hire the best interpreter you can get. But if you want to register for a conference, reserve a room at a hotel, plan a trip to Japan, you don't necessarily want to go through an expensive interpreter. You want to have a box that helps you along."

Among the various projects at Carnegie Mellon devoted to speech recognition, natural language understanding, and machine translation, Waibel's group has the distinction of emphasizing the application of neural networks — computer systems intended to mimic the brain — to speech recognition. Programmed to modify itself according to whatever signals come into the system, the speech recognizer actually "learns" how to identify sounds and words.

"This allows a great deal of flexibility and robustness," Waibel says. "The technology has matured enough that we are in a position to produce a state-of-the-art speech recognizer comparable with the best based on any other technique."

In 1988, the Carnegie Mellon group teamed up with Japan's ATR — which was in the midst of a seven-year initiative devoted to speech translation — to provide the ATR system's English-language component. Meanwhile, Waibel started a speech translation laboratory at the University of Karlsruhe in Germany and then got the Munich-based Siemens company interested. The two together have developed a German-language counterpart.

Working largely independently but sharing ideas, the three groups used their own approaches to build somewhat different systems. Nonetheless, charged with the common task of facilitating conference registration, all three systems also had to work together.

Carnegie Mellon's component, JANUS, translates English-language speech into German or Japanese text. To begin with, a person talks into a microphone. The resulting signal is converted into digits and parceled into 10-millisecond segments. Each of these speech fragments is then converted into 16 numbers, representing the signal's strength in 16 frequency ranges.

A speech recognizer analyzes the segments, identifying the particular language sounds, or phonemes, involved. Looking for patterns, it works out possible word combinations that seem to fit the identified sequence of phonemes and produces a list of candidate sentences, starting with the most likely possibility.

The translation part of the system then parses the top candidate, or works out its grammar in detail. Using this information, it converts the sentence into a special, intermediate language. The appropriate language generator then translates this intermediate form into either Japanese or German. Finally, the text is transmitted to computers in Japan or

Germany, where speech synthesizers complete the process.

The Carnegie-ATR-Siemens/Karlsruhe collaboration is not the only speech translation effort under way. Last year, scientists at AT&T Bell Laboratories in Murray Hill, N.J., collaborated with researchers at Telefonica Investigación y Desarollo in Spain to create a translator that could handle a 450-word vocabulary in Spanish and English. The system determined which language was spoken, translated the sentence into the other language, and "spoke" the new sentence, typically taking less than two seconds to complete the process.

To achieve this speed, the researchers found a way to use the same language model for both speech recognition and grammatical analysis, saving a potentially time-consuming step. Moreover, the system – known as VEST for Voice English/Spanish Translator – handled sentences dealing only with currency exchange and routine banking transactions.

Indeed, the most successful systems now in use all have strictly limited vocabularies and topics of conversation. "If you have an expert system that knows all about currency exchange, then it's [easy] to translate sentences back and forth between languages – so long as they deal only with currency exchange," says David Roe of Bell Labs. "What is hard is if you say 'bank' and you don't mean financial institution, you mean 'snowbank.'"

"That is where text translation machines usually run into problems," he adds. "They see a word and they cannot tell from the context what the sense of the word is."

One particularly successful system used in Canada translates weather forecasts with better than 99 percent accuracy between French and English. "Its saving grace is that it always deals only with weather forecasts," Roe comments.

The VEST system, demonstrated at Expo '92 in Seville, Spain, is part of an ongoing research effort at Bell Labs and Telefonica. The Spanish company already offers customers a system that recognizes the spoken words "uno," "dos," "tres," and so on, allowing someone using a dial telephone to make the same kinds of choices possible on a push-button phone.

At Bell Labs, Roe and his colleagues are working to improve speech translation systems by going back to the basics – looking for a superior method of speech recognition and for a better mathematical way of telling whether a given sequence of words is a valid sentence. "We also want to have translation from English into any of eight languages," Roe says.

Improved speech recognition remains one of the keys to better translation. A number of groups have recently demonstrated systems that indicate how far this technology has progressed in recent years (SN: 4/3/93, p.222). In one impressive showing, John Makhoul and his co-workers at BBN Systems and Technologies in Cambridge, Mass., showed that a speech recognition system running on an ordinary workstation could readily handle a 20,000-word vocabulary, no matter who the speaker is and without unnatural pauses between the spoken words.

But it's still a giant leap from speech recognition to accurate, rapid translation of speech – especially as the vocabulary gets larger and speakers are no longer restricted to grammatical sentences. It would also be nice if the translation system could somehow provide feedback concerning what it doesn't understand about any particular utterance.

"A human interpreter will carry on a dialog with a speaker in one language until the concept is clear before generating a message in the other language," Waibel remarks. "That's one of the things we're attempting to do in the second phase of our project."

That could be a handy capability when the system encounters the ill-formed sentences typical of spontaneous speech. "You want to allow people to speak spontaneously, without having to make sure they are speaking grammatically correct sentences, using only certain words, or not coughing in the middle of a sentence," Waibel says.

"But we're biting off a big chunk in going to spontaneous speech," adds Arthur E. McNair, a research programmer with the JANUS project.

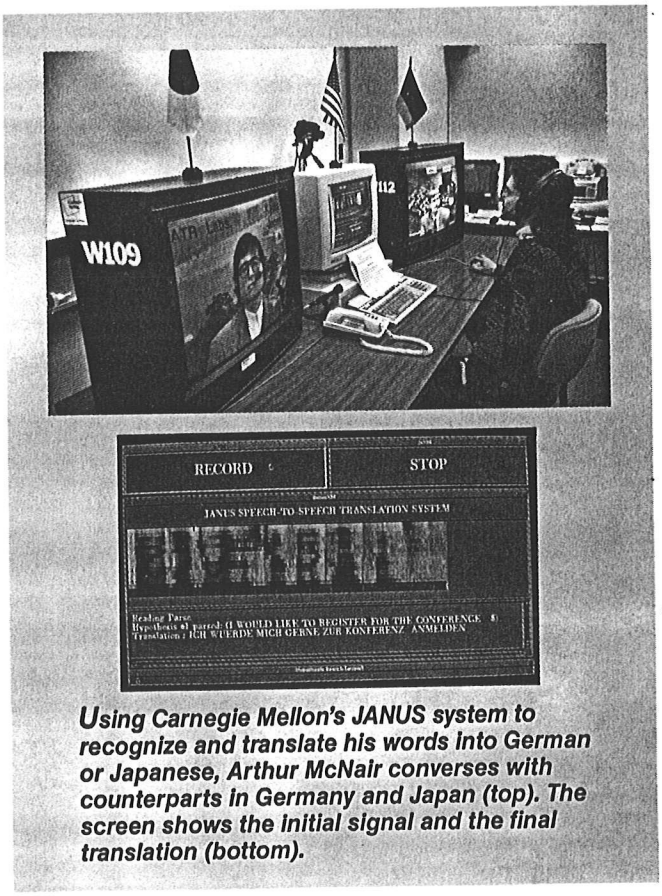
People in conversation naturally drift from topic to topic. Even in the seemingly benign realm of conference registration, a speaker may easily slip into subjects outside a system's expertise. When Waibel and his team recorded actual registration dialogs at a real conference, they found that some people stuck to the topic, while others wandered off on tangents. One woman went into a lengthy discussion of her recent divorce as a reason for asking the conference organizers to waive the registration fee.

Hence, most research groups will continue to concentrate on the translation of small vocabularies restricted to a certain domain. "The system isn't going to let you talk about anything under the sun," Waibel says. In its new effort, Waibel's group will focus on the task of scheduling a meeting as the topic of conversation.

Spoken language also has subtleties that seem almost impossible to capture by machine: the tone of a remark, the level of politeness, even the latest terms in an ever-changing body of slang expressions. "You need to get a lot more information out of the input data than what's required for a simple data-retrieval task," Waibel notes. "That makes speech translation much more challenging than speech recognition."

All this puts true "translating telephones" into the distant future. "A number of corporate managers have become very interested in the dream – and it really is a dream – of having telephone conversations between people speaking different languages," Roe says. "There's no doubt that this provides some of AT&T's corporate incentive for keeping the project going. But we have to work very hard to keep from overselling the technology."

At the same time, speech recognition shows enough promise that the German government has just launched a major initiative – an eight-year project dubbed Verbmobil – to develop a portable speech translator. And Japan's ATR is gearing up for the second phase of its effort. □



Using Carnegie Mellon's JANUS system to recognize and translate his words into German or Japanese, Arthur McNair converses with counterparts in Germany and Japan (top). The screen shows the initial signal and the final translation (bottom).

Photos: Carnegie Mellon