High-tech shatters trilingual barrier

By Byron Spice
Science Editor, Post-Gazette

It was the most perfunctory of conversations. One man asked how to get somewhere from Station Square. The other said to take a taxi.

What was remarkable was that one man spoke English and the other spoke German. And both understood the other perfectly. Later, a Japanese speaker joined in and nobody missed a beat.

This conversation yesterday through a video link between Carnegie Mellon University; Munich, Germany; and Kyoto, Japan, was the first public demonstration of a speech translation system. It is being jointly developed by Carnegie Mellon and its Japanese and German partners.

The vocabulary was limited to 500 words or so and the researchers all took pains to speak clearly and with proper grammar. The conversation nevertheless was a milestone in the system's development, said Alex Waibel, a researcher with appointments at both Carnegie Mellon and the University of Karlsruhe in Karlsruhe, Germany.

"As the world is growing more and more into a global village, we have a growing need to communicate with each other," he said. The hope is that someday English-speaking people will be able to pick up a phone and, without a bit of foreign language knowledge, talk with someone who doesn't know a bit of English.

That idea is still a bit outrageous, given the limitations of today's technology, Waibel said. But within five

Device lets three languages interplay

years, travelers may be able to carry foreign language dictionaries they can simply speak into to obtain a translation of a word or phrase.

In five or 10 years, commercial systems may allow people to conduct rudimentary business, such as making hotel or airplane reservations, he said. They might also be used as an aid in negotiating sales or to speed translation of television shows and movies.

Carnegie Mellon is spending about $350,000 a year to develop its version of the technology, which it dubbed Janus after the Roman god with two faces. But Japan's ATR Interpreting Telephony Laboratories in Kyoto is spending $15 million a year and the German government is launching an $8 million a year initiative.

A year ago, it took six computers about five minutes to translate a 2.5-second sentence, Waibel said, but a single high-powered personal computer now takes about eight seconds to translate the same sentence. A supercomputer can handle the task in about three seconds.

It's a three-stage process. First, the computer analyzes the sounds it hears and tries to match them with words it knows. It doesn't try to distinguish whether the speaker asked for "a new display," or "a nudist play." It just comes up with the 50 best acoustical matches.

At the second stage, the computer considers the context of the conversa-tion and decides whether the speaker most likely meant he can "recognize speech" or "wreck a nice beach." It then represents these thoughts in an intermediate language called Interlingua. From Interlingua, it can readily be translated into German, Japanese or English.

In the final stage, speech is synthesized.

Plenty of technical hurdles remain, Waibel emphasized. The translation must be done faster and involve a wider vocabulary. Speech synthesis must become more natural. But the system must also become more ro-bust, he said. That means it must be able to filter out background noises, such as coughs, laughs and dropped pens, and ignore the "uhs" and "ahs" that pepper most conversations. What's more, it has to decipher bad grammar and compensate for unusual cadences and pronunciations.

CMU programmer Arthur McNair adjusts controls during a demonstration of Janus. With him is researcher Alex Waibel.