



## Spoken Translation Commercializes Speech-to-Speech Translation

**Jackie Fenn**

*VP & Gartner Fellow, Lowell, MA, USA*

### **Update on Emerging Technologies, October 2005, 02-Nov-2005**

Spoken Translation, based in Berkeley, California, has launched the first commercial speech-to-speech translation system we have seen (that is, a process by which the user speaks in one language and the system translates and speaks the message in the target language). Until now, the capability has only been demonstrated in prototypes, such as the Multilingual Automatic Speech-to-Speech Translator (MASTOR) system from IBM Research, which translates between Chinese and English for a field medical worker (see "Speech-to-Speech Translation Moves Closer" in "[Update on Emerging Technologies, July 2004](#)").

The challenge regarding speech-to-speech translation is that it combines two less-than-perfect technologies — speech recognition and machine translation. (The third stage, speaking the translated text through text-to-speech technology, is more straightforward.) A successful strategy in this type of environment is to constrain the scope of the problem to raise the accuracy to acceptable levels. This has been used to good effect in the field of speech recognition for call centers, where domain-specific speech systems offer usable levels of accuracy for travel reservations, stock trading, call routing and ticket purchases. In the MASTOR prototype, IBM achieved impressive accuracy in speech-to-speech translation by limiting the scope of the dialogue to medical issues.

Although its first launch is a medical product, Spoken Translation's mission is to deliver general-purpose speech-to-speech translation. It uses third-party speech recognition, machine translation and text-to-speech technology, and has focused its capabilities on driving up accuracy levels through increasing user interaction rather than limiting the scope of coverage.

After a user speaks a sentence, Spoken Translation's system displays the output from the speech recognizer and enables the user to correct it, which is standard in complex speech applications. The novel part of Spoken Translation's approach is that the system translates to *and back from* the target language to create a paraphrase of the sentence before it is read aloud. For example, if a physician says "What seems to be the matter?" the system provides an on-screen translation, an English paraphrase and "meaning cues" indicating that the system has found two ways to interpret "matter" — one that means "substance" and one that means "problem." If the system has picked the wrong interpretation, the physician can select the correct meaning and ask the system to retranslate.

Given the additional correction stage required for the technology to work to full effect, Spoken Translation's approach will be best suited for applications where the speaker has the time and motivation to interact with the system. Suitable applications would include physicians, police, courts, border patrol or other situations where a government or other official needs to interact with a member of the public who does not speak the same language, and where a human translator is not available or not affordable. Further development or innovation will be required to support other, more-dynamic applications, such as international call centers, or situations where the speakers are not motivated to cooperate with the additional stages of the interactive approach.