Microsoft wins applause for tone-preserving translation (w/ Video)

November 10 2012, by Nancy Owano

(Phys.org)—Speech recognition in computers is an ongoing story with years of little progress in between. Even such programs as Siri have inspired derisive tales of how Siri renders flubs. Microsoft Chief Research Officer Rick Rashid recently presented an overview of where

speech recognition at Microsoft stands today. His talk, delivered in October at the Tianjin, China at Microsoft Research Asia's 21st Century Computing, has captured the attention of technology watchers globally, as it makes the point that progress really is on a roll. Rashid made it clear, through his summary timeline of milestones and direct demo of text to spech capabilities, that the newer signs of progress are substantial and impressive.

Following the overview, he said he wanted to address the audience in Chinese, using a text to speech system. He showed "how we take the text that represents my speech and run it through translation.-It required a text to speech system that Microsoft researchers built using a few hours' speech of a native Chinese speaker and properties of my own voice taken from about one hour of prerecorded (English) data, in this case recordings of previous speeches I'd made." The speech synthesis software that was put to use was able to preserve his very own cadence. The audience expressed delighted applause to see how much the translated speech still sounded like the voice of the original speaker. Rashid's words were almost instantly turned into Chinese, via the translation system, maintaining his speaking style.

In brief, the demo indicates that the technology world has taken a three-step turn where (1) spoken English can undergo machine translation and (2) spoken back in another language, with (3) the second-language translation retaining the speaker's cadence and tone.

This caps the last 60 years or so, where computer scientists have been working to build systems that can understand what a person says when they talk. The reason why scientists found it tough going at first was because of the imperfect approach used, as simple pattern matching. The computer would examine the waveforms produced by human speech and try to match them to waveforms associated with particular words. Everyone's voice is different, however, and even the same person can

say the same word in different ways.

Another milestone came in the late 1970s, with researchers at Carnegie Mellon focusing on speech recognition using a technique that could make use of training data from many speakers to build statistical speech models. Over the years that followed, speech systems advanced more and more, thanks in part to faster computers and the ability to process more data.

Just over two years ago, he continued, researchers at Microsoft Research and the University of Toronto reported a speech-recognition breakthrough. They were utilizing the Deep Neural Networks technique, patterned after the behavior of the human brain, recognizing sound the way the brain does. The result has been better recognition rates.

As for machine translation of text, capabilities have improved for translating web pages from one language to another. In Rashid's demo, he said words in English, sent through the translator system, and his words were played in Chinese. There were two steps put in play. "The first takes my words and finds the Chinese equivalents, and while non-trivial, this is the easy part," he said. "The second reorders the words to be appropriate for Chinese, an important step for correct translation between languages."

Rashid said results are still not perfect. Much work remains but the technology is promising enough to raise hopes that systems to break down language barriers are years, not centuries, off.

Rashid is not the first, however, to showcase instant translation technologies. Earlier this year, Microsoft Chief Research and Strategy Officer Craig Mundie captured imaginations of the audience at TechFest 2012, when he presented a bilingual talking head. Called "Monolingual TTS," the Microsoft software at play similarly was able to translate the

user's speech into another language and in a voice that sounded like the original user's.

The tool involved speech recognition, followed by translation, followed by a final text-to-speech output in a different language. The demo used an avatar of Mundie. A synthetic version of Mundie's voice, in English, welcomed the audience to Microsoft Research. Then the voice shifted to the same phrase in Mandarin. The words in Mandarin were reported to be recognizably Mundie's voice. Mundie said the dream was to be able to sit in an office and send an avatar to meet somebody in Beijing, speaking in English while the avatar speaks in Mandarin, realtime. "We want the computer to be a simultaneous translator."

More information: blogs.technet.com/b/next/archi … gy.aspx#.UJ7uVs3Aerh