Research aims to improve speech recognition software

EurekAlert

[1]

BINGHAMTON, NY ? Anyone who has used an automated airline reservation system has experienced the promise ? and the frustration ? inherent in today's automatic speech recognition technology. When it works, the computer "understands" that you want to book a flight to Austin rather than Boston, for example. Research conducted by Binghamton University's Stephen Zahorian aims to improve the accuracy of such programs.

Zahorian, a professor of electrical and computer engineering, recently received a grant of nearly half a million dollars from the Air Force Office of Scientific Research. The funds will support the two-year development of a multi-language, multi-speaker audio database that will be available for spoken-language processing research. Zahorian and his team plan to gather and annotate recordings of several hundred speakers each in English, Spanish and Mandarin Chinese.

"The challenge," he said, "is to get speech recognition working better in real-life situations."

That's why the samples in the new database will come from publicly available sources such as YouTube.

Zahorian's team will annotate each sample, creating a more detailed version of closed captioning, including time stamps and descriptions of background sounds. Once the human listener has finished with the transcription, automatic speech recognition algorithms will be used to align the recording with the captions. Next, software will be developed to verify and correct errors in the time alignment.

"Speech-recognition algorithms begin by mimicking what your ear does," Zahorian said. "But we want the algorithms to extract just the most useful characteristics of the speech, not all of the possible data. That's because more detail can actually hurt performance, past a certain point."

The field of automatic speech recognition has a long history, dating back to projects at Bell Labs before the computer age. These days, much of the technology relies on algorithms that convert sounds into numbers.

In Zahorian's research, he represents speech as a picture in a time-frequency plane. He then uses image-processing techniques to extract features of the speech, which has led him to focus more on time than on frequency.

When researchers are ready to test an algorithm, they rely on a common set of databases held by the Linguistic Data Consortium. Zahorian's unusual image-based approach has given his team some of the best results ever reported for automatic speech recognition experiments using two of the consortium's best-known databases.

The database Zahorian develops with the new funding will join these others, offering researchers around the world a new way to test their theories with samples of real-life speech.

Some mistakes are inevitable, given the variations in pitch, tone and pronunciation from person to person.

Still, the field does have a clear standard, Zahorian said: "In order to be useful, a system should have a word-error rate of no more than 10 percent."

Zahorian is interested in language modeling ? if someone has said these three words, what's the fourth word likely to be? ? as well as conversation modeling ? that is, predicting when the speakers will switch. He's also intrigued by the potential to make advances by using established methods from other fields, including the neural networks developed by researchers working in artificial intelligence.

He sees a future in which automatic speech recognition will enable technology to extract the meaning of speech as well as the words.

"The dream," Zahorian said, "is that someday travelers will be able to speak into a little gadget that will translate what they've said into another language instantly and accurately."

SOURCE [2]

Source URL (retrieved on *03/16/2014 - 9:57am*):
http://www.ecnmag.com/news/2010/08/research-aims-improve-speech-recognition-software

Links:
[1] http://www.eurekalert.org/multimedia/pub/24694.php?from=166544
[2] http://www.eurekalert.org/pub_releases/2010-08/bu-rat081110.php