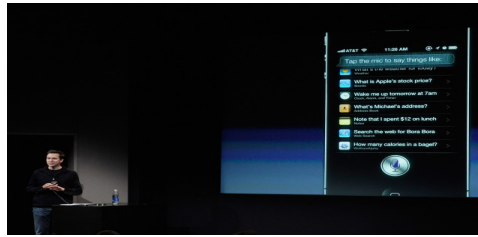


Human-machine cooperation next phase of voice recognition



Apple's Senior Vice President of iOS Scott Forstall speaks about the new voice recognition app called Siri at the event introducing the new iPhone 4s at the company's headquarters October 4, 2011 in Cupertino, California. Getty/ Kevork Djansezian

In the past year, many resources have been spent on speech recognition technologies. Siri was upgraded in 2012. IBM launched Watson, and Google upgraded its Voice Search application. Microsoft created a service that spoke in Mandarin in the user's voice. Japan telecom company DoCoMo unveiled a real-time translation app at the Barcelona Mobile World Congress at a week-long yearly trade show held in Barcelona this year.

But as users can attest, automated voice-to-text and translation programs often deliver a less than seamless result. Companies looking for accurate, professional transcriptions are left to seek out small-scale services that compose a fragmented industry. Several companies are popping up and looking to grow big in that space, including NextWave, QuickTate, Speechpad, TranscribeMe, and Amazon's CastingWords. Of those, some such as Nexiwave and QuickTate pride themselves on figuring out how machines can be just as good, if not better and faster, than humans in transcription. Others, such as Speechpad and CastingWords insist on human-sourced translations. The problem with human-based translations is that they take longer, especially for large bulks of text. TranscribeMe, a transcription company with \$1.5 million in venture capital is somewhere in between, crowd-sourced and cloud-sourced, using a mix of algorithms and human.

TranscribeMe works like this: The user records an audio file via computer or phone, and receives an email about payment almost immediately. Once the payment information is filled, the transcription results will be emailed in less than 24 hours on average, depending on the length.

TranscribeMe is cloud-based and crowd-sourced. Once the audio file is uploaded, an algorithm does a rough conversion to text. Then, the file is broken into sixty-second pieces (to ensure privacy of subject matter) and sent to different transcribers around

the world who can transcribe in their free time. Each segment transcribed is recorded in a cloud, which contains data used for the algorithms for automated transcription. Which means the more TranscribeMe transcribes, the closer it is to reliance on machines.

In the perfect scenario TranscribeMe CEO Dunayev's described in his interview with Blouin News, transcriptions are automated and edited by humans. Such a model could definitely work in the medical and legal fields, where accuracy and speed are both important. TranscribeMe is not real-time yet, but Dunayev said they plan to make it real time this year. The technology is there, he said, but the next step is gathering a critical mass of transcribers that will access and start transcribing a file immediately. For example, if a 5-minute transcription is split up into 100 pieces and distributed to transcribers that are available immediately, the turnaround time can be as low as three seconds.

Crowdsourcing to increase speed is not limited to transcription. The logical next step of transcription is the information research model, similar to IBM's Watson. Most information retrieval is done through machines, but how can human critical thinking be added to the model, while maintaining real-time speed? Dunayev envisioned a scenario where a complicated question is asked, broken down into smaller pieces, each piece given to a group of humans to work together and solve. Amazon's Mechanical Turk is already starting a process like that, even though it could be more streamlined. Those who work for Mechanical Turk take on small tasks such as describing the color and shape of a pair of ballet shoes for rates as cheap as \$0.04 cents. Often, the payment varies based on quality. For example, if a person's answer on a question-answer sight receives positive feedback, their payment becomes higher.

Despite the excitement surrounding machines' ability to understand and use human speech, doubts are rising about how accurate technology can be in that field, or any field that requires analyzing data. Cooperation between humans and machines could be the next phase that gives rest to the accuracy question.

TAGS:amazonAmazon.comAmazon's CastingWordsGoogleIBMNextWaveTranscribeMe

Link: <http://blogs.blouinnews.com/blouinbeattechnology/2013/02/27/a-new-formula-for-voice-recognition-human-machine-cooperation/>