Google's DeepMind Achieves Speech-Generation Breakthrough

Jeremy Kahn
jeremyakahn
September 9, 2016 — 6:29 AM CDT

Google's DeepMind unit, which is working to develop super-intelligent computers, has created a system for machine-generated speech that it says outperforms existing technology by 50 percent.

U.K.-based DeepMind, which Google **acquired** for about 400 million pounds ($533 million) in 2014, developed an artificial intellig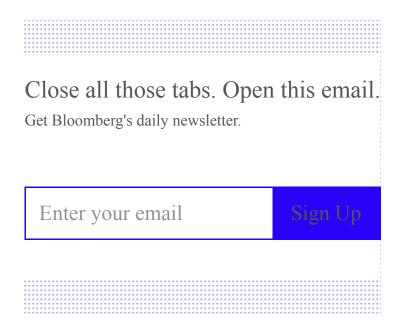ence called WaveNet that can mimic human speech by learning how to form the individual sound waves a human voice creates, it said in a **blog post** Friday. In blind tests for U.S. English and Mandarin Chinese, human listeners found WaveNet-generated speech sounded more natural than that created with any of Google's existing text-to-speech programs, which are based on different technologies. WaveNet still underperformed recordings of actual human speech.

QUICKTAKE
Artificial Intelligence

Many computer-generated speech programs work by using a large data set of short recordings of a single human speaker and then combining these speech fragments to form new words. The result is intelligible and sounds human, if not completely natural. The drawback is that the sound of the voice cannot be easily modified. Other systems form the voice completely electronically, usually based on rules about how the certain letter-combinations are pronounced. These systems allow the sound of the voice to be manipulated easily, but they have tended to sound less natural than computer-generated speech based on recordings of human speakers, DeepMind said.

WaveNet is a type of AI called a neural network that is designed to mimic how parts of the human brain function. Such networks need to be trained with large data sets.

'Challenging Task'

WaveNet won't have immediate commercial applications because the system requires too much computational power: it has to sample the audio signal it is being trained on 16,000 times per second or more, DeepMind said. And then for each of those samples it has to form a prediction about what the soundwave should look like based on each of the prior samples. Even the DeepMind researchers acknowledged in their blog post that this "is a clearly challenging task."

Still, tech companies are likely to pay close attention to DeepMind's breakthrough. Speech is becoming an increasingly important way humans interact with everything from mobile phones to cars. Amazon.com Inc., Apple Inc., Microsoft Inc. and Alphabet Inc.'s Google have all invested in personal digital assistants that primarily interact with users through speech. Mark Bennett, the international director of Google Play, which sells Android apps, told an Android developer conference in London last week that 20 percent of mobile searches using Google are made by voice, not written text.

And while researchers have made great strides in getting computers to understand spoken language, their ability to talk back in ways that seem fully human has lagged.

Strategy Game

WaveNet is yet another coup for DeepMind, which is best known for creating **AlphaGo**, an AI system that beat the world's top ranked human player in the strategy game Go this year.

Still, Google has disclosed little about how DeepMind's research has helped it commercially, although the company has **revealed** that it has used DeepMind's technology to reduce the power demands of its data centers by 40 percent, saving enough money to justify the amount Google spent to buy the London AI company. It has also said that DeepMind has helped achieve "substantial improvements to a set of services from YouTube and Google Play to Google's advertising products."