

Software Is Called Capable of Copying Any Human Voice

By LISA GUERNSEY
Published: July 31, 2001

AT&T Labs will start selling speech software that it says is so good at reproducing the sounds, inflections and intonations of a human voice that it can recreate voices and even bring the voices of long-dead celebrities back to life. The software, which turns printed text into synthesized speech, makes it possible for a company to use recordings of a person's voice to utter things that the person never actually said.

The software, called Natural Voices, is not flawless -- its utterances still contain a few robotic tones and unnatural inflections -- and competitors question whether the software is a substantial step up from existing products. But some of those who have tested the technology say it is the first text-to-speech software to raise the specter of voice cloning, replicating a person's voice so perfectly that the human ear cannot tell the difference.

"If ABC wanted to use Regis Philbin's voice for all of its automated customer-service calls, it could," said Lawrence R. Rabiner, vice president for AT&T Labs Research.

Potential customers for the software, which is priced in the thousands of dollars, include telephone call centers, companies that make software that reads digital files aloud, and makers of automated voice devices.

James R. Fruchterman, the chief executive of Benetech, a nonprofit organization that uses technology in social-service projects, tested the software along with a dozen people who evaluate technology for blind people, and they were impressed. "Natural Voices gets into the gray area," he said, "where there is plausible deniability that it is a machine."

Dr. Rabiner said he was excited about the possibility of resurrecting renowned voices, like that of Harry Caray, the Chicago Cubs announcer who delivered rousing play-by-play broadcasts. "There are probably hours of recordings in archives," he said. Wouldn't it be great, he asked, if Harry Caray's voice could again be broadcasting in Wrigley Field?

The advances raise several sticky issues. Who, for example, owns the rights to a celebrity's voice? (Dr. Rabiner predicts that new contracts will be drawn that include voice-licensing clauses.) With virtual characters already appearing in place of real ones in some movies, will synthesized voices compete with those of live actors as well? And although scientists say the technology is not yet good enough to perpetrate fraud, could the synthesized voices eventually be capable of tricking people into thinking that they were getting phone calls or digital audio recordings from people they know?

For now, technical limitations may temper any worries that a person's voice could be lifted without permission. To build the software that recreates unique voices -- which AT&T Labs is calling its "custom voice" product -- a person must first go to a studio where engineers record 10 to 40 hours of readings. Texts range from business news reports to nonsense babble.

The recordings are then chopped into fragments of sounds and sorted into databases. When the software processes a text, it retrieves the sounds and re-assembles them instantly to form entirely new sentences. In the case of long-dead celebrities, archival recordings could be used in the same way.

Other companies and research centers, like I.B.M. Research and Lernout & Hauspie Speech Products, are also experimenting with this technique -- which is called concatenative speech synthesis -- to improve the quality of text-to-speech software. It is a big step up, engineers say, from the speech engines that were built from whole words that had been pre-recorded. And it is also a vast improvement, some say, from the entirely computer-generated and therefore robotic sounds that are used in many versions of text-to-speech software now on the market.

Now, aided by the declining cost and increasing speed of microprocessors, far smoother sentences are possible, Dr. Rabiner said. He said that the speech team at AT&T Labs, led by Dr. Juergen Schroeter, an expert in speech synthesis, had created a more refined form of the concatenative technique by breaking a person's voice into "the smallest number of units possible."

A demonstration of the technology will be available on the Web beginning today at www.naturalvoices.att.com, said Michael Dickman, a spokesman for AT&T Labs.

Still, many engineers are skeptical of claims of a completely simulated voice that is almost indistinguishable from that of a human. "The methods and algorithms that we know of, they still need a lot more work," said P. S. Gopalakrishnan, the manager of the pervasive speech technologies group at I.B.M. Research, which competes with AT&T Labs in the field.

Now the pressure is on to perfect the technology. Analysts at McKinsey & Company, a management consultant, have predicted that the market for text-to-speech software will reach more than \$1 billion in the next five years. In addition to customers like call centers and manufacturers of automated voice systems, the software could also be used by publishers of video games and books-on-tape and automobile manufacturers whose cars are equipped with software that gives driving directions. In the near future, engineers expect that people will want high-end speech technology that enables them to interact at length with their cell phones and Palm organizers, instead of typing on and squinting at a tiny screen.

AT&T Labs' speech technology will be the first product that is actually sold by the laboratory, which is typically a research and development division. So far, the laboratory has hired three actors -- two male and one female -- to provide the voices that it will sell separately from the "custom voice" option. Mr. Dickman said that the company planned to recreate other voices, too, such as that of a child and a grandmother. Spanish-language voices are expected in a few months.

One of the voices is based on that of an African-American actor from New Jersey. (He and AT&T have requested that his name not be published because a clause in his contract stipulates that his identity is a company secret based on years of research and auditions.) He said the experience of being a "voice donor," as he called it, was both stimulating and unsettling.

"It's been for me exciting because I know there is an end product that will have my voice carried on forever," he said. At the same time, he said, "I have a lot of dread, or at least concern, of whether I'm contributing to the demise of the live actor."

Even Mr. Fruchterman, one of AT&T Lab's potential first clients, said he wondered what the new technology might bring.

"Just like you can't trust a photograph anymore," he said, "you won't be able to trust a voice either."

[Home](#) | [Times topics](#) | [Member Center](#)

[Copyright 2013](#) | [The New York Times Company](#) | [Privacy Policy](#) | [Help](#) | [Contact Us](#) | [Work for Us](#) | [Site Map](#) | [Index by Keyword](#)