

Speech Synthesizer Helps Movie Critic

June 15 2010, By Phillip F. Schewe



Matthew Aylett and colleagues are aiming to capture a more conversational style of synthesized speech.

The voices you hear on message services are often created artificially by fitting together short audio snippets from a large library of vocalized words and sounds. Scientists are now moving beyond the older generic voices to produce what, to many listeners, sounds like an "actual" person speaking.

Voice synthesis technology had a recent success when a personalized textto-speech system was crafted for the movie critic Roger Ebert, who suffers from thyroid cancer and is unable to speak any longer. The synthetic Ebert speech was created from many hours of his <u>voice</u> recorded during past television programs. The company that made Ebert's system, <u>CereProc</u>, is able to make TTS conversions quickly, allowing them to provide more than just a generic-voice synthesis.



Matthew Aylett, one of the founders of CereProc, said that he and his colleagues do not necessarily aim for smooth voicings. Instead, he said, "We want the variation which gives a voice a fresh and natural feel to it. This means not getting our voice talents to speak in a boring and neutral way but capturing their more conversational speech style."

At CereProc they are even able to build a certain amount of emotion into their voices as well creating regional accents. The company, located in Edinburgh, Scotland, has been successful, for example, in reproducing Scottish and Irish accented speech.

For "cloning" voices of specific people (they did one for President George Bush, for example), he said they sometimes have resort to "found" recordings. The trouble with using snippets of voice is that these recordings were often made under a variety of acoustic conditions, which then have to be corrected in making a final voice.

Roger Ebert learned about the President Bush "voice" and asked if CereProc could assemble a voice for him using the large inventory of <u>audio recordings</u> of his television show. So far they have used about five hours of Ebert's voice to produce a voice bank of about 300,000 phonetic sounds.

Timothy Bunnell, a scientist at Nemours Biomedical Research in Wilmington, Del., is working to make personalized voices accessible to everyone, especially for those with neurodegenerative diseases.

Bunnell's voice synthesis system can be prepared for people who still have the power of speech. By contrast, Aylett's synthetic speech program, such as that for Roger Ebert, is based on recordings of people who can no longer speak.

Carrying out text-to-speech synthesis for children is more difficult. "It is



difficult for young children to record enough speech with the required degree of consistency and precision needed to build a high-quality synthetic voice," said Bunnell. The main problem, he said, is not with the nature of the utterances, but with the amount of variability in children's <u>speech</u>, much more than for adults.

Aylett enjoys synthesis research. "It's fun to use the synthesizer, but it's even better to see it helping people who really need it."

Provided by Inside Science News Service

Citation: Speech Synthesizer Helps Movie Critic (2010, June 15) retrieved 27 April 2024 from <u>https://phys.org/news/2010-06-speech-movie-critic.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.