Voice copying algorithms found able to dupe voice recognition devices

October 13 2021, by Bob Yirka

Credit: Pixabay/CC0 Public Domain

A team of researchers at the University of Chicago has found that voice-copying algorithms have advanced to the point that they are now capable of fooling voice recognition devices, and in many cases, people listening

to them. The group has posted a paper on the arXiv preprint server that describes two well-known voice copying algorithms.

Deepfake videos are well-known; many examples of what only appear to be celebrities can be seen regularly on YouTube. But while such videos have grown lifelike and convincing, one area where they fail is in reproducing a person's voice. In this new effort, the team at UoC found evidence that the technology has advanced. They tested two of the most well-known voice copying algorithms against both human and voice recognition devices and found that the algorithms have improved to the point that they are now able to fool both.

The two algorithms—SV2TTS and AutoVC—were tested by obtaining samples of voice recordings from publicly available databases. Both systems were trained using 90 five-minute voice snippets of people talking. They also enlisted the assistance of 14 volunteers who provided voice samples and access to their voice recognition devices. The researchers then tested the two systems using the open-source software Resemblyzer—it listens and compares voice recordings and then gives a rating based on similar two samples are. They also tested the algorithms by using them to attempt to access services on voice recognition devices.

The researchers found the algorithms were able to fool the Resemblyzer nearly half of the time. They also found that they were able to fool Azure (Microsoft's cloud computing service) approximately 30 percent of the time. And they were able to fool Amazon's Alexa voice recognition system approximately 62% of the time.

Two hundred volunteers also listened to pairs of recordings and tried to determine if the voices were from the same person—the results were mixed, but overall, the algorithms were able to fool the volunteers more often than not—and especially so when the voice samples were of famous people.

More information: Emily Wenger et al, "Hello, It's Me": Deep Learning-based Speech Synthesis Attacks in the Real World. arXiv:2109.09598v1 [cs.CR], arxiv.org/abs/2109.09598