Connecting the dots between voice and a human face
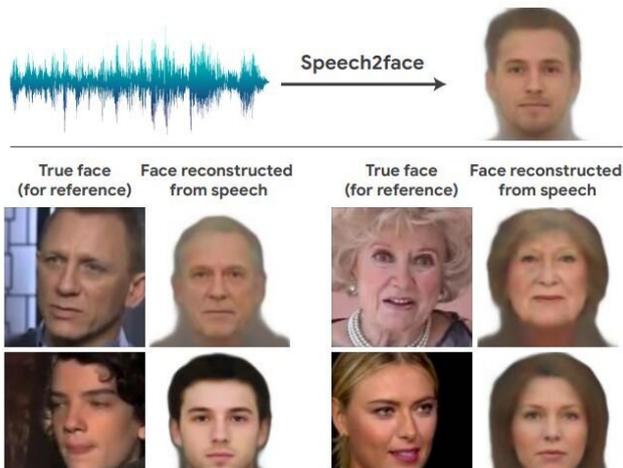
13 June 2019, by Nancy Cohen

Once again, artificial intelligence teams tease the realm of the impossible and deliver surprising results. This team in the news figured out what a person's face may look like just based on voice. Welcome to Speech2Face. The research team found a way to reconstruct some people's very rough likeness based on short audio clips.

The paper describing their work is up on arXiv, and is titled "Speech2Face: Learning the Face Behind a Voice." Authors are Tae-Hyun Oh, Tali Dekel, Changil Kim, Inbar Mosseri, William Freemany, Michael Rubinstein and Wojciech Matusiky. "Our goal in this work is to study to what extent we can infer how a person looks from the way they talk."

They evaluate and numerically quantify how, and in what way, their Speech2Face reconstructions from audio resemble the true face images of the speakers.

The authors apparently wanted to make sure their intent was clear, not as some attempt to link voices with images of the specific people who actually spoke, as "our goal is not to predict a recognizable image of the exact face, but rather to capture dominant facial traits of the person that are correlated with the input speech."

The authors on GitHub said that they also felt it important to discuss in the paper ethical considerations "due to the potential sensitivity of facial information."

They said in their paper that their method "cannot recover the true identity of a person from their voice (i.e., an exact image of their face). This is because our model is trained to capture visual features (related to age, gender, etc.) that are common to many individuals, and only in cases where there is strong enough evidence to connect those visual features with vocal/speech attributes in the data."

They also said the model will produce average-looking faces—only average looking faces— with characteristic visual features correlated with the input speech.

Jackie Snow, *Fast Company*, wrote about their method. Snow said the dataset that they took was made up of clips from YouTube. Speech2Face was trained by scientists on videos from the internet that showed people talking. They created a neural network-based model that "learns vocal attributes associated with facial features from the videos."

Snow added, "Now, when the system hears a new sound bite, the AI can use what it's learned to guess what the face might look like."

Neurohive discussed their work: "From the videos, they extract speech-face pairs, which are fed into two branches of the architecture. The images are encoded into a latent vector using the pre-trained face recognition model, whilst the waveform is fed into a voice encoder in a form of a spectrogram, in order to utilize the power of convolutional architectures. The encoded vector from the voice

encoder is fed into the face decoder to obtain the final face reconstruction."

One can also get a precise report on their method and how they tested with an article on *Packt*:

"They said they further evaluated and numerically quantified how their Speech2Face reconstructs, obtains results directly from audio, and how it resembles the true face images of the speakers. For this, they tested their model both qualitatively and quantitatively on the AVSpeech dataset and the VoxCeleb dataset."

How might their findings help realworld applications? They said, "we believe that predicting face images directly from voice may support useful applications, such as attaching a representative face to phone/video calls based on the speaker's voice."

Why their work matters: Think patterns. "Previous research has explored methods for predicting age and gender from speech," said Snow, "but in this case, the researchers claim they have also detected correlations with some facial patterns too."

More information: Speech2Face: Learning the Face Behind a Voice, arXiv:1905.09773 [cs.CV]
arxiv.org/abs/1905.09773

© 2019 Science X Network
APA citation: Connecting the dots between voice and a human face (2019, June 13) retrieved 26 November 2020 from https://techxplore.com/news/2019-06-dots-voice-human.html