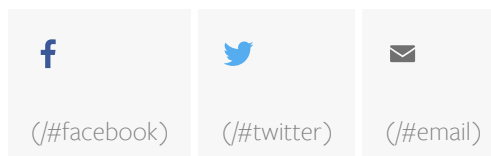## AI-equipped eyeglasses can read silent speech

By Louis DiPietro ,
Cornell Ann S. Bowers College of Computing and Information Science

April 6, 2023

It may look like Ruidong Zhang is talking to himself, but in fact the doctoral student in the field of information science is silently mouthing the passcode to unlock his nearby smartphone and play the next song in his playlist.

It's not telepathy: It's the seemingly ordinary, off-the-shelf eyeglasses he's wearing, called EchoSpeech – a silent-speech recognition interface that uses acoustic-sensing and artificial intelligence to continuously recognize up to 31 unvocalized commands, based on lip and mouth movements.

Developed by Cornell's **Smart Computer Interfaces for Future Interactions (SciFi) Lab** (https://www.scifilab.org/), the low-power, wearable interface requires just a few minutes of user training data before it will recognize commands and can be run on a smartphone, researchers said.

Zhang is the lead author of "**EchoSpeech: Continuous Silent Speech Recognition on Minimally-obtrusive Eyewear Powered by Acoustic Sensing** (https://ruidongzhang.com/files/papers/EchoSpeech_authors_version.pdf)," which will be presented at the Association for Computing Machinery Conference on Human Factors in Computing Systems (CHI) this month in Hamburg, Germany.

Provided
Ruidong Zhang, a doctoral student in the field of information science, wearing EchoSpeech glasses.

"For people who cannot vocalize sound, this silent speech technology could be an excellent input for a voice synthesizer. It could give patients their voices back," Zhang said of the technology's potential use with further development.

In its present form, EchoSpeech could be used to communicate with others via smartphone in places where speech is inconvenient or inappropriate, like a noisy restaurant or quiet library. The silent speech interface can also be paired with a stylus and used with design software like CAD, all but eliminating the need for a keyboard and a mouse.

Outfitted with a pair of microphones and speakers smaller than pencil erasers, the EchoSpeech glasses become a wearable AI-powered sonar system, sending and receiving soundwaves across the face and sensing mouth movements. A deep learning algorithm, also developed by SciFi Lab researchers, then analyzes these echo profiles in real time, with about 95% accuracy.

"We're moving sonar onto the body," said **Cheng Zhang** (https://infosci.cornell.edu/content/zhang-0), assistant professor of information science in the Cornell Ann S. Bowers College of Computing and Information Science and director of the SciFi Lab.

"We're very excited about this system," he said, "because it really pushes the field forward on performance and privacy. It's small, low-power and privacy-sensitive, which are all important features for deploying new, wearable technologies in the real world."

The SciFi Lab has developed several wearable devices that track **body** (https://news.cornell.edu/stories/2022/11/wrist-mounted-camera-captures-entire-body-3d), **hand** (https://news.cornell.edu/stories/2020/07/researchers-develop-3d-hand-sensing-wristband) and **facial** (https://news.cornell.edu/stories/2022/02/smart-necklace-recognizes-english-mandarin-commands) movements using machine learning and wearable, miniature video cameras. Recently, the lab has shifted away from cameras and toward acoustic sensing to track face and body movements, citing improved battery

life; tighter security and privacy; and smaller, more compact hardware. EchoSpeech builds off the lab's similar acoustic-sensing device called **EarIO** (https://news.cornell.edu/stories/2022/07/earable-uses-sonar-reconstruct-facial-expressions), a wearable earbud that tracks facial movements.

Most technology in silent-speech recognition is limited to a select set of predetermined commands and requires the user to face or wear a camera, which is neither practical nor feasible, Cheng Zhang said. There also are major privacy concerns involving wearable cameras – for both the user and those with whom the user interacts, he said.

Acoustic-sensing technology like EchoSpeech removes the need for wearable video cameras. And because audio data is much smaller than image or video data, it requires less bandwidth to process and can be relayed to a smartphone via Bluetooth in real time, said **François Guimbretière** (https://infosci.cornell.edu/content/guimbreti%C3%A8re), professor in information science in Cornell Bowers CIS and a co-author.

"And because the data is processed locally on your smartphone instead of uploaded to the cloud," he said, "privacy-sensitive information never leaves your control."

Battery life improves exponentially, too, Cheng Zhang said: Ten hours with acoustic sensing versus 30 minutes with a camera.

The team is exploring commercializing the technology behind EchoSpeech, thanks in part to **Ignite: Cornell Research Lab to Market gap funding** (https://ctl.cornell.edu/ignite/).

In forthcoming work, SciFi Lab researchers are exploring smart-glass applications to track facial, eye and upper body movements.

"We think glass will be an important personal computing platform to understand human activities in everyday settings," Cheng Zhang said.

Other co-authors were information science doctoral student Ke Li, Yihong Hao '24, Yufan Wang '24 and Zhengnan Lai '25. This research was funded in part by the National Science Foundation.

Louis DiPietro is a writer for the Cornell Ann S. Bowers College of Computing and Information Science.

MEDIA CONTACT

Becka Bowyer

rpb224@cornell.edu (mailto:rpb224@cornell.edu)

📞 6072204185 (tel:6072204185)

Get Cornell news delivered right to your inbox.

SUBSCRIBE (/ABOUT/SUBSCRIBE-NEWSLETTER)

YOU MIGHT ALSO LIKE

Wrist-mounted camera captures entire body in 3D (/stories/2022/11/wrist-mounted-camera-captures-entire-body-3d)

'Earable' uses sonar to reconstruct facial expressions (/stories/2022/07/earable-uses-sonar-reconstruct-facial-expressions)