Computer system transcribes words users 'speak silently'

April 5 2018, by Larry Hardesty

Arnav Kapur, a researcher in the Fluid Interfaces group at the MIT Media Lab, demonstrates the AlterEgo project.. Credit: Lorrie Lejeune/MIT

MIT researchers have developed a computer interface that can transcribe words that the user verbalizes internally but does not actually speak

aloud.

The system consists of a [wearable device](#) and an associated computing system. Electrodes in the device pick up neuromuscular signals in the jaw and face that are triggered by internal verbalizations—saying words "in your head"—but are undetectable to the human eye. The signals are fed to a machine-learning system that has been trained to correlate particular signals with particular words.

The device also includes a pair of bone-conduction headphones, which transmit vibrations through the bones of the face to the inner ear. Because they don't obstruct the ear canal, the headphones enable the system to convey information to the user without interrupting conversation or otherwise interfering with the user's auditory experience.

The device is thus part of a complete silent-computing system that lets the user undetectably pose and receive answers to difficult computational problems. In one of the [researchers](#)' experiments, for instance, subjects used the system to silently report opponents' moves in a chess game and just as silently receive computer-recommended responses.

"The motivation for this was to build an IA device—an intelligence-augmentation device," says Arnav Kapur, a graduate student at the MIT Media Lab, who led the development of the new system. "Our idea was: Could we have a computing platform that's more internal, that melds human and machine in some ways and that feels like an internal extension of our own cognition?"

"We basically can't live without our cellphones, our digital devices," says Pattie Maes, a professor of media arts and sciences and Kapur's thesis advisor. "But at the moment, the use of those devices is very disruptive. If I want to look something up that's relevant to a conversation I'm

having, I have to find my phone and type in the passcode and open an app and type in some search keyword, and the whole thing requires that I completely shift attention from my environment and the people that I'm with to the phone itself. So, my students and I have for a very long time been experimenting with new form factors and new types of experience that enable people to still benefit from all the wonderful knowledge and services that these devices give us, but do it in a way that lets them remain in the present."

The researchers describe their device in a paper they presented at the Association for Computing Machinery's ACM Intelligent User Interface conference. Kapur is first author on the paper, Maes is the senior author, and they're joined by Shreyas Kapur, an undergraduate major in electrical engineering and computer science.

Subtle signals

The idea that internal verbalizations have physical correlates has been around since the 19th century, and it was seriously investigated in the 1950s. One of the goals of the speed-reading movement of the 1960s was to eliminate internal verbalization, or "subvocalization," as it's known.

But subvocalization as a computer interface is largely unexplored. The researchers' first step was to determine which locations on the face are the sources of the most reliable neuromuscular signals. So they conducted experiments in which the same subjects were asked to subvocalize the same series of words four times, with an array of 16 electrodes at different facial locations each time.

The researchers wrote code to analyze the resulting data and found that signals from seven particular electrode locations were consistently able to distinguish subvocalized words. In the conference paper, the

researchers report a prototype of a wearable silent-speech interface, which wraps around the back of the neck like a telephone headset and has tentacle-like curved appendages that touch the face at seven locations on either side of the mouth and along the jaws.

But in current experiments, the researchers are getting comparable results using only four electrodes along one jaw, which should lead to a less obtrusive wearable device.

Once they had selected the electrode locations, the researchers began collecting data on a few computational tasks with limited vocabularies—about 20 words each. One was arithmetic, in which the user would subvocalize large addition or multiplication problems; another was the chess application, in which the user would report moves using the standard chess numbering system.

Then, for each application, they used a neural network to find correlations between particular neuromuscular signals and particular words. Like most neural networks, the one the researchers used is arranged into layers of simple processing nodes, each of which is connected to several nodes in the layers above and below. Data are fed into the bottom layer, whose nodes process it and pass them to the next layer, whose nodes process it and pass them to the next layer, and so on. The output of the final layer yields is the result of some classification task.

The basic configuration of the researchers' system includes a neural network trained to identify subvocalized words from neuromuscular signals, but it can be customized to a particular user through a process that retrains just the last two layers.

Practical matters

Using the prototype wearable interface, the researchers conducted a usability study in which 10 subjects spent about 15 minutes each customizing the arithmetic application to their own neurophysiology, then spent another 90 minutes using it to execute computations. In that study, the system had an average transcription accuracy of about 92 percent.

But, Kapur says, the system's performance should improve with more training data, which could be collected during its ordinary use. Although he hasn't crunched the numbers, he estimates that the better-trained system he uses for demonstrations has an accuracy rate higher than that reported in the usability study.

In ongoing work, the researchers are collecting a wealth of data on more elaborate conversations, in the hope of building applications with much more expansive vocabularies. "We're in the middle of collecting data, and the results look nice," Kapur says. "I think we'll achieve full conversation some day."

"I think that they're a little underselling what I think is a real potential for the work," says Thad Starner, a professor in Georgia Tech's College of Computing. "Like, say, controlling the airplanes on the tarmac at Hartsfield Airport here in Atlanta. You've got jet noise all around you, you're wearing these big ear-protection things—wouldn't it be great to communicate with voice in an environment where you normally wouldn't be able to? You can imagine all these situations where you have a high-noise environment, like the flight deck of an aircraft carrier, or even places with a lot of machinery, like a power plant or a printing press. This is a system that would make sense, especially because oftentimes in these types of or situations people are already wearing protective gear. For instance, if you're a fighter pilot, or if you're a firefighter, you're already wearing these masks."

"The other thing where this is extremely useful is special ops," Starner adds. "There's a lot of places where it's not a noisy environment but a silent environment. A lot of time, special-ops folks have hand gestures, but you can't always see those. Wouldn't it be great to have silent-speech for communication between these folks? The last one is people who have disabilities where they can't vocalize normally. For example, Roger Ebert did not have the ability to speak anymore because lost his jaw to cancer. Could he do this sort of silent speech and then have a synthesizer that would speak the words?"

More information: Arnav Kapur et al. AlterEgo, *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18* (2018). DOI: 10.1145/3172944.3172977

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology