AI Revolution: Paralyzed Woman 'Speaks' via Digital Avatar

Featured   Neuroscience   · August 23, 2023

For weeks, the participant worked with the team to train the system's artificial intelligence algorithms to recognize her unique brain signals for speech. Credit: Neuroscience News

Summary: Researchers created a groundbreaking brain-computer interface (BCI) that allows a paralyzed woman to communicate through a digital avatar. This advancement marks the first-ever synthesis of speech or facial expressions directly from brain signals.

The system can convert these signals to text at an impressive rate of nearly 80 words per minute, surpassing existing technologies. The study presents a significant leap towards restoring comprehensive communication for paralyzed individuals.

Key Facts:

1. The BCI developed decodes brain signals into synthesized speech and facial expressions, enabling paralyzed individuals to communicate more naturally.
2. Instead of recognizing whole words, the system identifies phonemes, the sub-units of speech, enhancing speed and accuracy.
3. The digital avatar's voice was personalized to mirror the user's voice pre-injury, and facial animations were driven by software that interpreted the brain's signals for various facial expressions.

Source: UCSF

Researchers at UC San Francisco and UC Berkeley have developed a brain-computer interface (BCI) that has enabled a woman with severe paralysis from a brainstem stroke to speak through a digital avatar.

It is the first time that either speech or facial expressions have been synthesized from brain signals. The system can also decode these signals into text at nearly 80 words per minute, a vast improvement over commercially available technology.

Edward Chang, MD, chair of neurological surgery at UCSF, who has worked on the technology, known as a brain computer interface, or BCI, for more than a decade, hopes this latest research breakthrough, appearing Aug. 23, 2023, in *Nature*, will lead to an FDA-approved system that enables speech from brain signals in the near future.

"Our goal is to restore a full, embodied way of communicating, which is really the most natural way for us to talk with others," said Chang, who is a member of the UCSF Weill Institute for Neuroscience and the Jeanne Robertson Distinguished Professor in Psychiatry.

"These advancements bring us much closer to making this a real solution for patients."

Chang's team previously demonstrated it was possible to decode brain signals into text in a man who had also experienced a brainstem stroke many years earlier. The current study demonstrates something more ambitious: decoding brain signals into the richness of speech, along with the movements that animate a person's face during conversation.

Chang implanted a paper-thin rectangle of 253 electrodes onto the surface of the woman's brain over areas his team has discovered are critical for speech. The electrodes intercepted the brain signals that, if not for the stroke, would have gone to muscles in her, tongue, jaw and larynx, as well as her face. A cable, plugged into a port fixed to her head, connected the electrodes to a bank of computers.

For weeks, the participant worked with the team to train the system's artificial intelligence algorithms to recognize her unique brain signals for speech. This involved repeating different phrases from a 1,024-word conversational vocabulary over and over again, until the computer recognized the brain activity patterns associated with the sounds.

Rather than train the AI to recognize whole words, the researchers created a system that decodes words from phonemes. These are the sub-units of speech that form spoken words in the same way that letters form written words. "Hello," for example, contains four phonemes: "HH," "AH," "L" and "OW."

Using this approach, the computer only needed to learn 39 phonemes to decipher any word in English. This both enhanced the system's accuracy and made it three times faster.

"The accuracy, speed and vocabulary are crucial," said Sean Metzger, who developed the text decoder with Alex Silva, both graduate students in the joint

Bioengineering Program at UC Berkeley and UCSF. "It's what gives a user the potential, in time, to communicate almost as fast as we do, and to have much more naturalistic and normal conversations."

To create the voice, the team devised an algorithm for synthesizing speech, which they personalized to sound like her voice before the injury, using a recording of her speaking at her wedding.

The team animated the avatar with the help of software that simulates and animates muscle movements of the face, developed by Speech Graphics, a company that makes AI-driven facial animation.

The researchers created customized machine-learning processes that allowed the company's software to mesh with signals being sent from the woman's brain as she was trying to speak and convert them into the movements on the avatar's face, making the jaw open and close, the lips protrude and purse and the tongue go up and down, as well as the facial movements for happiness, sadness and surprise.

"We're making up for the connections between the brain and vocal tract that have been severed by the stroke," said Kaylo Littlejohn, a graduate student working with Chang and Gopala Anumanchipalli, PhD, a professor of electrical engineering and computer sciences at UC Berkeley.

"When the subject first used this system to speak and move the avatar's face in tandem, I knew that this was going to be something that would have a real impact."

An important next step for the team is to create a wireless version that would not require the user to be physically connected to the BCI.

"Giving people the ability to freely control their own computers and phones with this technology would have profound effects on their independence and social interactions," said co-first author David Moses, PhD, an adjunct professor in neurological surgery.

Authors: Additional authors include Ran Wang, Maximilian Dougherty, Jessie Liu, delyn Tu-Chan, and Karunesh Ganguly of UCSF, Peter Wu and Inga Zhuravleva of UC Berkeley, and Michael Berger of Speech Graphics.

About this AI and neurotech research news

Abstract

An analog-AI chip for energy-efficient speech recognition and transcription

Models of artificial intelligence (AI) that have billions of parameters can achieve high accuracy across a range of tasks, but they exacerbate the poor energy efficiency of conventional general-purpose processors, such as graphics processing units or central processing units.

Analog in-memory computing (analog-AI) can provide better energy efficiency by performing matrix–vector multiplications in parallel on 'memory tiles'.

However, analog-AI has yet to demonstrate software-equivalent (SW_{eq}) accuracy on models that require many such tiles and efficient communication of neural-network activations between the tiles.

Here we present an analog-AI chip that combines 35 million phase-change memory devices across 34 tiles, massively parallel inter-tile communication and analog, low-power peripheral circuitry that can achieve up to 12.4 tera-operations per second per watt (TOPS/W) chip-sustained performance.

We demonstrate fully end-to-end SW_{eq} accuracy for a small keyword-spotting network and near-SW_{eq} accuracy on the much larger MLPerf recurrent neural-network transducer (RNNT), with more than 45 million weights mapped onto more than 140 million phase-change memory devices across five chips.