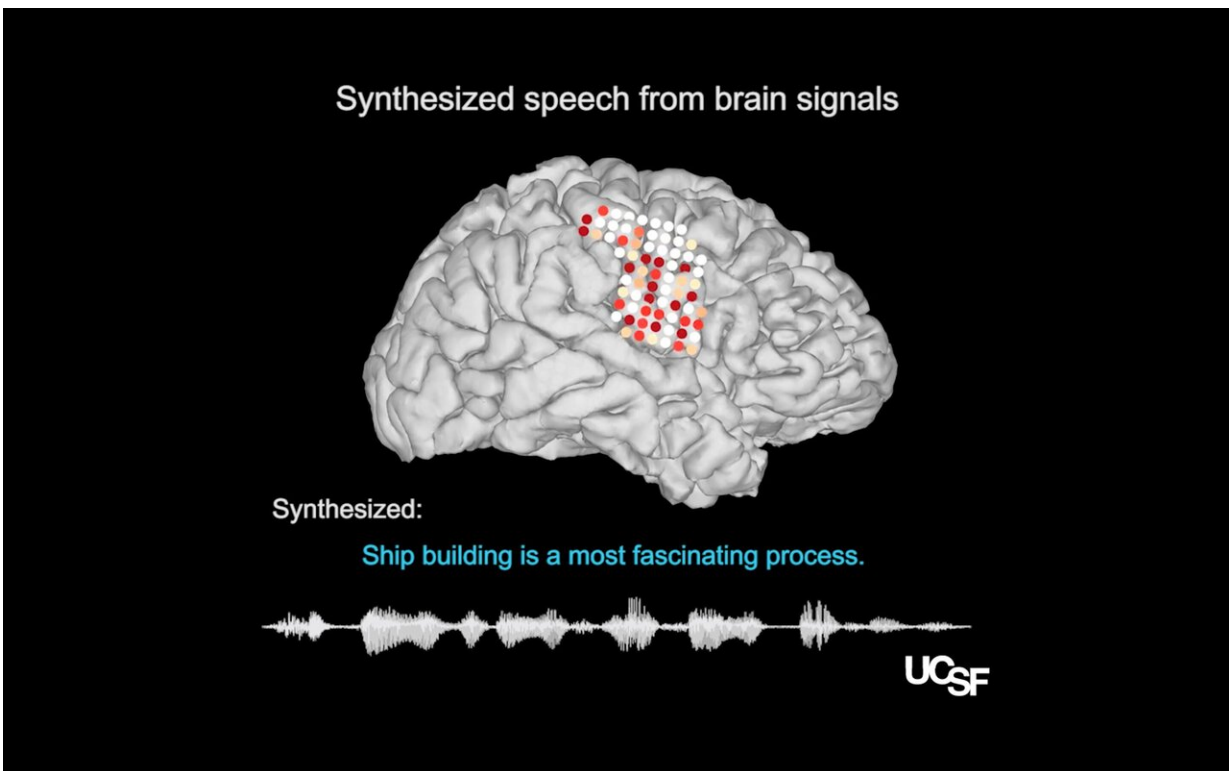


# Synthetic speech generated from brain recordings

April 24 2019

---



Illustrations of electrode placements on the research participants' neural speech centers, from which activity patterns recorded during speech (colored dots) were translated into a computer simulation of the participant's vocal tract (model, right) which then could be synthesized to reconstruct the sentence that had been spoken (sound wave & sentence, below). Credit: Chang lab / UCSF Dept. of Neurosurgery

A state-of-the-art brain-machine interface created by UC San Francisco neuroscientists can generate natural-sounding synthetic speech by using brain activity to control a virtual vocal tract—an anatomically detailed computer simulation including the lips, jaw, tongue, and larynx. The study was conducted in research participants with intact speech, but the technology could one day restore the voices of people who have lost the ability to speak due to paralysis and other forms of neurological damage.

Stroke, [traumatic brain injury](#), and [neurodegenerative diseases](#) such as Parkinson's disease, multiple sclerosis, and amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease) often result in an irreversible loss of the ability to speak. Some people with severe speech disabilities learn to spell out their thoughts letter-by-letter using assistive devices that track very small eye or facial muscle movements. However, producing text or synthesized speech with such devices is laborious, error-prone, and painfully slow, typically permitting a maximum of 10 words per minute, compared to the 100-150 words per minute of natural speech.

The new system being developed in the laboratory of Edward Chang, MD—described April 24, 2019 in *Nature*—demonstrates that it is possible to create a synthesized version of a person's voice that can be controlled by the activity of their [brain](#)'s speech centers. In the future, this approach could not only restore fluent communication to individuals with severe speech disability, the authors say, but could also reproduce some of the musicality of the human voice that conveys the speaker's emotions and personality.

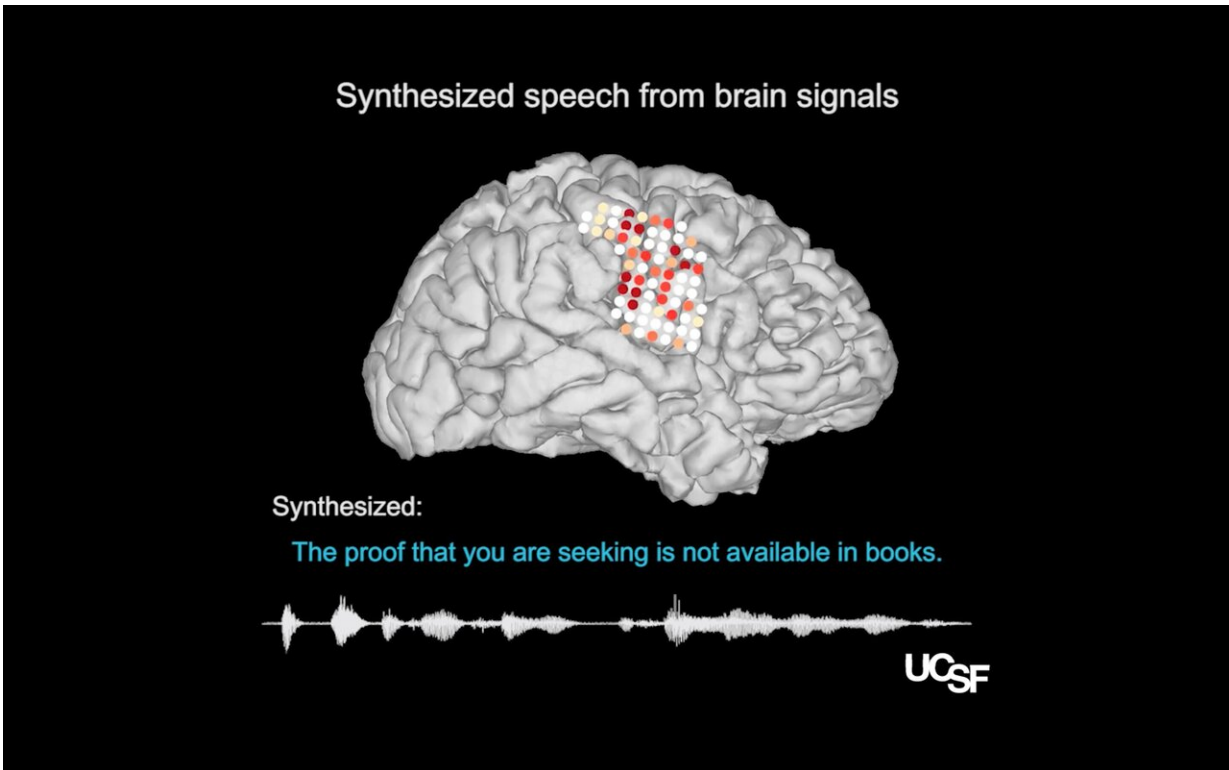
"For the first time, this study demonstrates that we can generate entire spoken sentences based on an individual's brain activity," said Chang, a professor of neurological surgery and member of the UCSF Weill Institute for Neuroscience. "This is an exhilarating proof of principle that with technology that is already within reach, we should be able to build a device that is clinically viable in patients with speech loss."

## Virtual Vocal Tract Improves Naturalistic Speech Synthesis

The research was led by Gopala Anumanchipalli, Ph.D., a speech scientist, and Josh Chartier, a bioengineering graduate student in the Chang lab. It builds on a [recent study](#) in which the pair described for the first time how the human brain's speech centers choreograph the movements of the lips, jaw, tongue, and other vocal tract components to produce fluent speech.

From that work, Anumanchipalli and Chartier realized that previous attempts to directly decode speech from brain activity might have met with limited success because these brain regions do not directly represent the acoustic properties of speech sounds, but rather the instructions needed to coordinate the movements of the mouth and throat during speech.

"The relationship between the movements of the vocal tract and the speech sounds that are produced is a complicated one," Anumanchipalli said. "We reasoned that if these speech centers in the brain are encoding movements rather than sounds, we should try to do the same in decoding those signals."



Illustrations of electrode placements on the research participants' neural speech centers, from which activity patterns recorded during speech (colored dots) were translated into a computer simulation of the participant's vocal tract (model, right) which then could be synthesized to reconstruct the sentence that had been spoken (sound wave & sentence, below). Credit: Chang lab / UCSF Dept. of Neurosurgery

In their new study, Anumancipali and Chartier asked five volunteers being treated at the UCSF Epilepsy Center—patients with intact speech who had electrodes temporarily implanted in their brains to map the source of their seizures in preparation for neurosurgery—to read several hundred sentences aloud while the researchers recorded activity from a brain region known to be involved in language production.

Based on the audio recordings of participants' voices, the researchers

used linguistic principles to reverse engineer the vocal tract movements needed to produce those sounds: pressing the lips together here, tightening vocal cords there, shifting the tip of the tongue to the roof of the mouth, then relaxing it, and so on.

This detailed mapping of sound to anatomy allowed the scientists to create a realistic virtual vocal tract for each participant that could be controlled by their brain activity. This comprised two "[neural network](#)" machine learning algorithms: a decoder that transforms brain activity patterns produced during speech into movements of the virtual vocal tract, and a synthesizer that converts these vocal tract movements into a synthetic approximation of the participant's voice.

The synthetic speech produced by these algorithms was significantly better than synthetic speech directly decoded from participants' brain activity without the inclusion of simulations of the speakers' vocal tracts, the researchers found. The algorithms produced sentences that were understandable to hundreds of human listeners in crowdsourced transcription tests conducted on the Amazon Mechanical Turk platform.

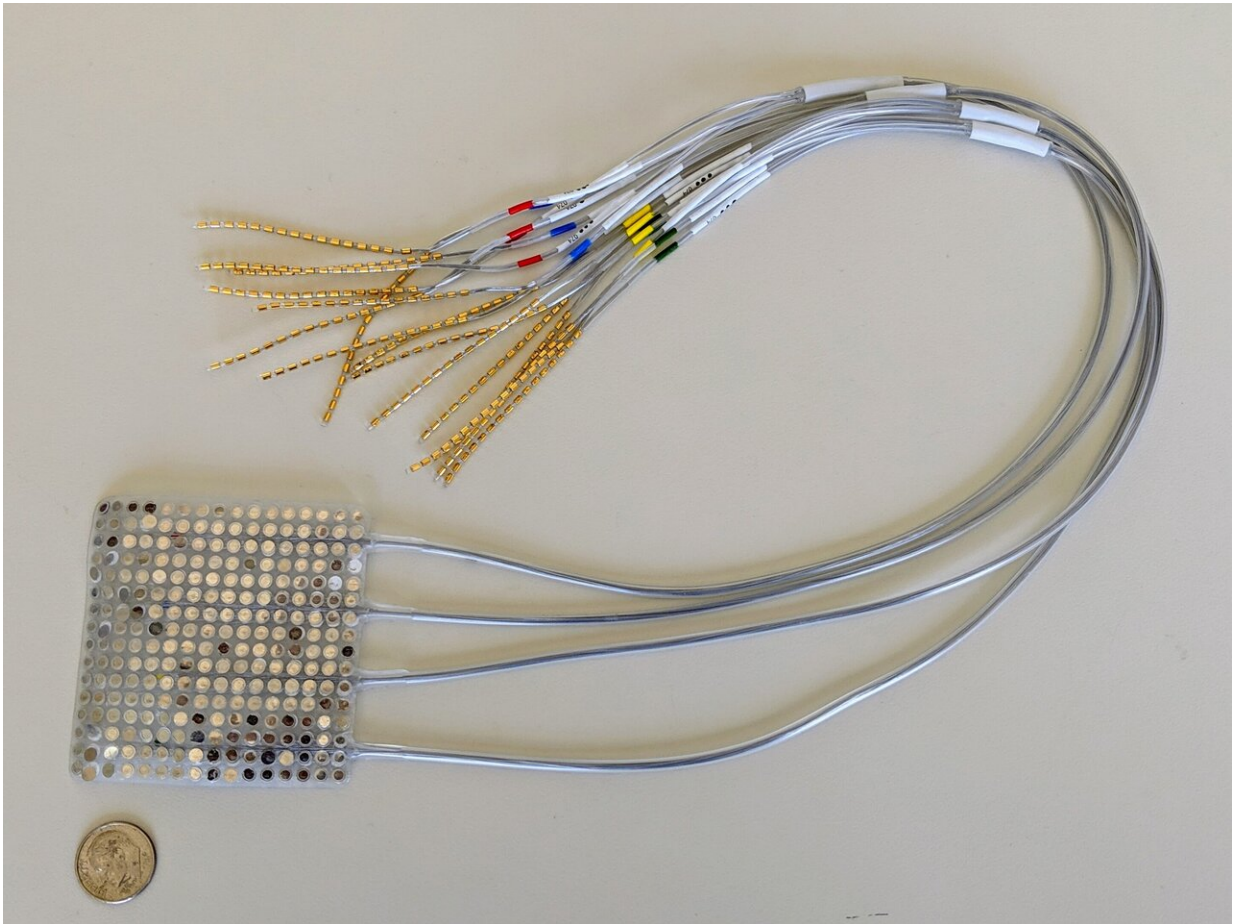


Image of an example array of intracranial electrodes of the type used to record brain activity in the current study. Credit: UCSF

As is the case with natural speech, the transcribers were more successful when they were given shorter lists of words to choose from, as would be the case with caregivers who are primed to the kinds of phrases or requests patients might utter. The transcribers accurately identified 69 percent of synthesized words from lists of 25 alternatives and transcribed 43 percent of sentences with perfect accuracy. With a more challenging 50 words to choose from, transcribers' overall accuracy dropped to 47 percent, though they were still able to understand 21 percent of synthesized sentences perfectly.

"We still have a ways to go to perfectly mimic spoken language," Chartier acknowledged. "We're quite good at synthesizing slower [speech sounds](#) like 'sh' and 'z' as well as maintaining the rhythms and intonations of speech and the speaker's gender and identity, but some of the more abrupt sounds like 'b's and 'p's get a bit fuzzy. Still, the levels of accuracy we produced here would be an amazing improvement in real-time communication compared to what's currently available."

## **Artificial Intelligence, Linguistics, and Neuroscience Fueled Advance**

The researchers are currently experimenting with higher-density electrode arrays and more advanced machine learning algorithms that they hope will improve the synthesized speech even further. The next major test for the technology is to determine whether someone who can't speak could learn to use the system without being able to train it on their own voice and to make it generalize to anything they wish to say.



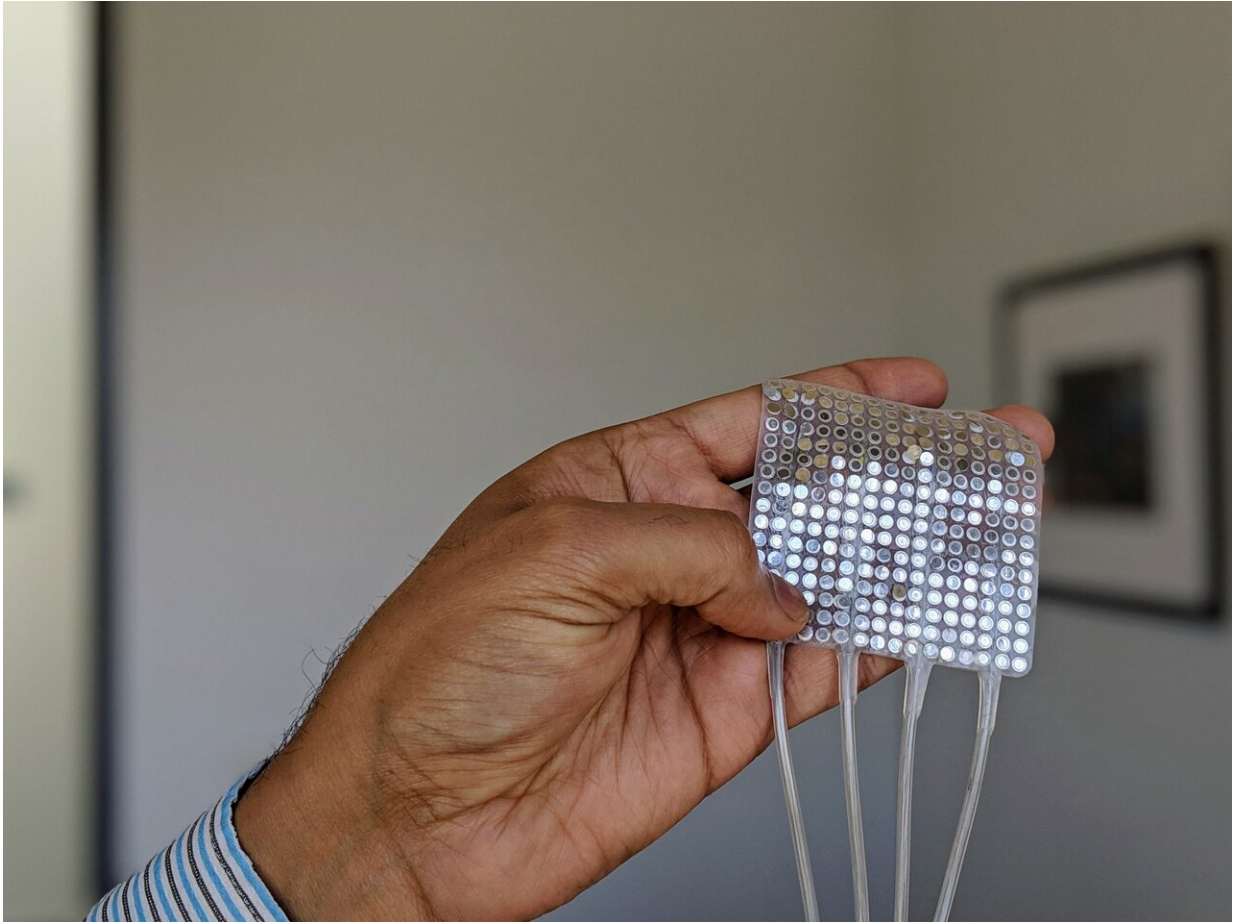


Image of an example array of intracranial electrodes of the type used to record brain activity in the current study. Credit: UCSF

Preliminary results from one of the team's research participants suggest that the researchers' anatomically based system can decode and synthesize novel sentences from participants' brain activity nearly as well as the sentences the algorithm was trained on. Even when the researchers provided the algorithm with [brain activity](#) data recorded while one participant merely mouthed sentences without sound, the system was still able to produce intelligible synthetic versions of the mimed sentences in the speaker's voice.



The researchers also found that the neural code for vocal movements partially overlapped across participants, and that one research subject's vocal tract simulation could be adapted to respond to the neural instructions recorded from another participant's brain. Together, these findings suggest that individuals with speech loss due to neurological impairment may be able to learn to control a speech prosthesis modeled on the voice of someone with intact speech.

"People who can't move their arms and legs have learned to control robotic limbs with their brains," Chartier said. "We are hopeful that one day people with [speech](#) disabilities will be able to learn to speak again using this brain-controlled artificial vocal tract."

Added Anumanchipalli, "I'm proud that we've been able to bring together expertise from neuroscience, linguistics, and machine learning as part of this major milestone towards helping neurologically disabled patients."

**More information:** Speech synthesis from neural decoding of spoken sentences, *Nature*, 2019. [DOI: 10.1038/s41586-019-1119-1](https://doi.org/10.1038/s41586-019-1119-1) , [www.nature.com/articles/s41586-019-1119-1](https://www.nature.com/articles/s41586-019-1119-1)

Provided by University of California, San Francisco

Citation: Synthetic speech generated from brain recordings (2019, April 24) retrieved 2 May 2024 from <https://medicalxpress.com/news/2019-04-synthetic-speech-brain.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--