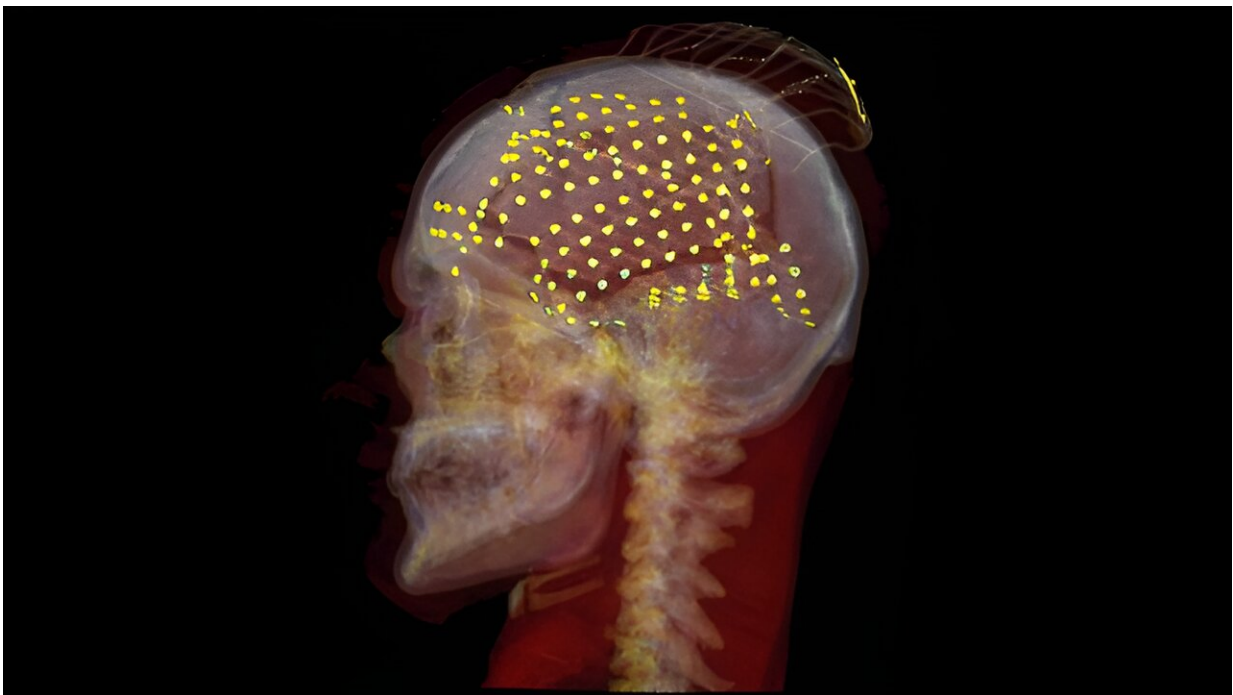# Researchers reconstruct speech from brain activity, illuminating complex neural processes

October 11 2023

CT reconstruction of an ECoG electrode grid placement (yellow) in one patient. Credit: NYU Tandon School of Engineering

Speech production is a complex neural phenomenon that has left researchers explaining it tongue-tied. Separating out the complex web of neural regions controlling precise muscle movement in the mouth, jaw

and tongue with the regions processing the auditory feedback of hearing your own voice is a complex problem, and one that has to be overcome for the next generation of speech-producing protheses.

Now, a team of researchers from New York University have made key discoveries that help untangle that web, and are using it to build vocal reconstruction technology that recreates the voices of patients who have lost their ability to speak.

The team, co-led by Adeen Flinker, Associate Professor of Biomedical Engineering at NYU Tandon and Neurology at NYU Grossman School of Medicine, and Yao Wang, Professor of Biomedical Engineering and Electrical and Computer Engineering at NYU Tandon, as well as a member of NYU WIRELESS, created and used complex neural networks to recreate speech from brain recordings, and then used that recreation to analyze the processes that drive human speech.

They have detailed their new findings in a new paper published in the *Proceedings of the National Academy of Sciences (PNAS)*.

Human speech production is a complex behavior that involves feedforward control of motor commands as well as feedback processing of self-produced speech. These processes require the engagement of multiple brain networks in tandem. However, it has been hard to dissociate the degree and timing of cortical recruitment for motor control versus sensory processing generated by speech production.

In a new paper, the researchers have successfully disentangled the intricate processes of feedback and feedforward during speech production. Utilizing an innovative deep learning architecture on human neurosurgical recordings, the team employed a rule-based differentiable speech synthesizer to decode speech parameters from cortical signals.

By implementing neural network architectures that distinguish between causal (using current and past neural signals to decode current speech), anticausal (using current and future neural signals), or a combination of both (noncausal) temporal convolutions, the researchers were able to meticulously analyze the contributions of feedforward and feedback in speech production.
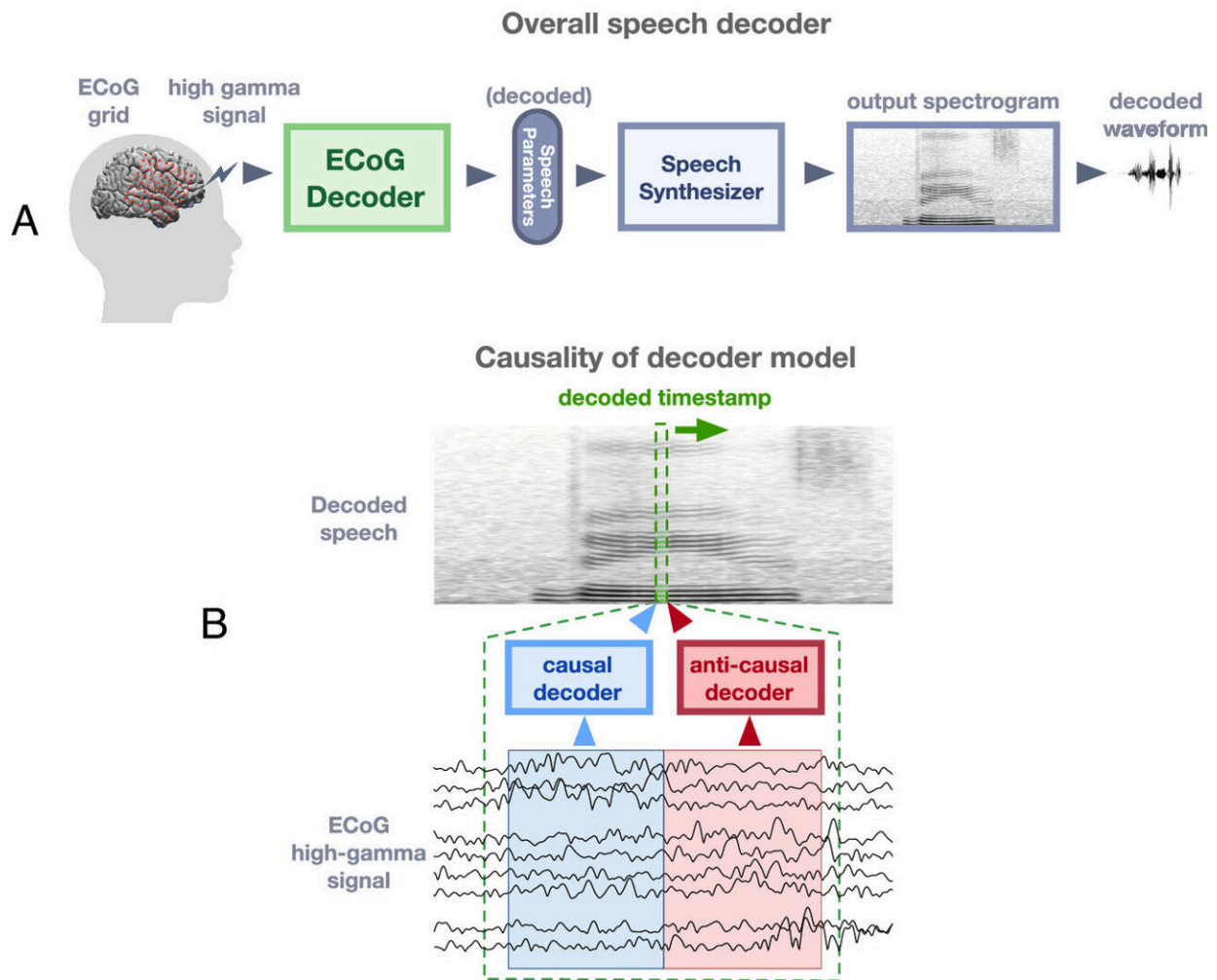
Speech decoding framework. (A) The overall structure of the decoding pipeline. ECoG amplitude signals are extracted in the high gamma range (i.e., 70–150 Hz). The ECoG Decoder translates neural signals from the electrode array to a set of speech parameters. This parameter representation is used to drive a speech

synthesizer which creates a spectrogram (and associated waveform). (B) Illustration of causal versus anticausal neural decoding, which is leveraged to disentangle feedforward and feedback contributions. In order to reconstruct a speech timestamp, the causal model only uses feedforward neural signals in the past (blue), whereas the anticausal model only uses the neural signals that occur after the timestamp (red). The noncausal model is allowed to use both past and future neural signals to decode the current timestamp. Credit: *Proceedings of the National Academy of Sciences* (2023). DOI: 10.1073/pnas.2300255120

"This approach allowed us to disentangle processing of feedforward and feedback neural signals that occur simultaneously while we produce speech and sense feedback of our own voice," says Flinker.

This cutting-edge approach not only decoded interpretable speech parameters but also provided insights into the temporal receptive fields of recruited cortical regions. Remarkably, the findings challenge prevailing notions that segregate feedback and feedforward cortical networks.

The analyses unveiled a nuanced architecture of mixed feedback and feedforward processing, spanning frontal and temporal cortices. This novel perspective, combined with exceptional speech decoding performance, marks a significant leap forward in our understanding of the intricate neural mechanisms underlying speech production.

The researchers have used this new perspective to inform their own development of prostheses that can read brain activity and decode it directly into speech. While many researchers are working on developing such devices, the NYU prototype has one key difference—it is able to recreate the voice of the patient, using only a small data set of recordings, to a remarkable degree.

The result may be that patients do not get just any voice back after losing it—they will get their own voice back. This is thanks to a deep neural network that takes into account a latent auditory space, and can be trained on just a few samples of an individual voice, like a YouTube video or Zoom recording.

In order to collect the data, researchers turned to a group of patients who have refractory epilepsy, currently untreatable by medication. These patients had a grid of subdural EEG electrodes implanted on their brains for a one week period to monitor their conditions, and consented to an additional 64 smaller electrodes interleaved between the regular clinical electrodes. They provided the researchers with key insights to brain activity during the act of speech production.

In addition to Flinker and Wang, the researchers include Ran Wang, Xupeng Chen and Amirhossein Khalilian-Gourtani from NYU Tandon's Electrical and Computer Engineering Department, Leyao Yu from the Biomedical Engineering Department, Patricia Dugan, Daniel Friedman, and Orrin Devinsky from NYU Grossman's Neurology Department, and Werner Doyle from the Neurosurgery department.

Provided by NYU Tandon School of Engineering