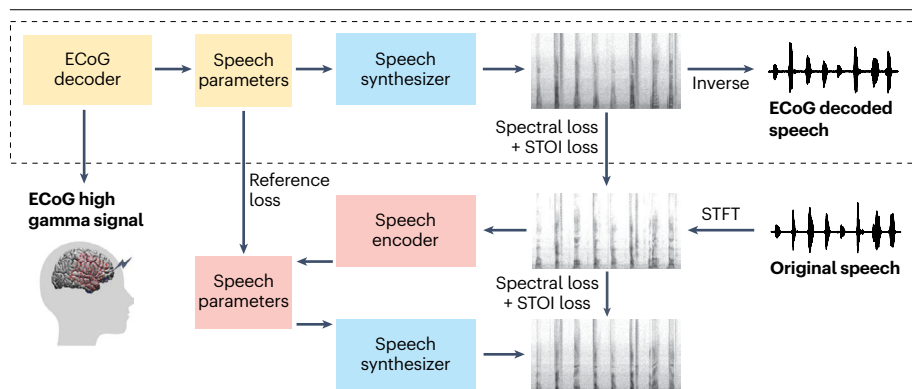


An electrocortigraphy-based speech decoder for neural speech prostheses



Neural speech prostheses based on brain-computer interface (BCI) systems and machine learning technologies are fundamental tools for helping restore speech loss due to neurological deficits. Despite the progress, advances in speech decoding are still hampered by the limited timeframe of data available for training personalized models and the inherent variability in speech production. Now, writing in *Nature Machine Intelligence*, Adeen Flinker and colleagues introduce an electrocortigraphy (ECoG)-to-speech framework that overcomes data scarcity and enhances natural-sound speech synthesis.

‘Our labs work together at the interface of machine learning and neurosurgical recordings in patients undergoing treatment for refractory epilepsy,’ says Flinker. The study involved participants who completed a series of speech tasks while their neural and acoustic data were synchronized and recorded using clinically implanted electrodes that were in place to monitor and localize the seizures. The open-source

framework comprises an ECoG decoder that translates neural signals into interpretable acoustic speech parameters (for example, pitch, voicing and formant frequencies) and a differentiable speech synthesizer that converts these parameters into spectrograms. ‘In this way, we managed to deal with the limitations of scarce data and leveraged an interpretable and compact acoustic space that can be used to investigate and decode both perceived and produced speech,’ continues the researcher.

The approach is based on a specific acoustic space that can synthesize speech using differentiable digital signal processing and an acoustic pre-training that is blind to the neural data and enable to decode realistic vocalizations with limited data. The training process involves semi-supervised learning that initially uses only speech signals, followed by supervised learning based on ground-truth spectrograms (using measures of spectrogram difference and short-time objective intelligibility, STOI). The team used various deep learning architectures

(including 3D ResNet, 3D Swin Transformer and LSTM) that reply on different causality directions for the ECoG decoder and trained the models using a portion of the data and validated them on the remaining set. Robust decoding performance were achieved across participants with both high-density and low-density electrode sampling on the cortex and high accuracy was demonstrated across different models, with causal decoding with convolutional architectures showing superior performance.

The team also conducted a contribution analysis to identify which cortical regions are most involved in speech decoding. They used an occlusion approach to evaluate the change in when signals from specific electrodes were removed. This analysis highlighted similar contributions from both hemispheres across sensorimotor areas, particularly in ventral regions and comparable decoding from both left and right hemispheres were demonstrated.

Future work will focus on refining these techniques for broader clinical application, including testing in patients with damaged left hemispheres to confirm right-hemisphere decoding efficacy. ‘The immediate next step is to create new models that are topology agnostic and are viable for new patients that the model has never encountered. We are also working towards expanding the vocabulary to more words and sentences,’ concludes Flinker.

Silvia Conti

Original article: Chen, X. et al. A neural speech decoding framework leveraging deep learning and speech synthesis. *Nat. Mach. Intell.* **6**, 467–480 (2024)