Regaining your voice

AI speech neuroprostheses can restore day-to-day communication after neurological injury

Sergey Stavisky

Speech is one of the most remarkable and essential human behaviors. We precisely control nearly 100 muscles to produce an exquisitely timed series of breaths and rapid articulator movements that vibrate air in just the right way so that another person can understand our verbalized thoughts. The loss of this ability owing to neurological injuries such as stroke or amyotrophic lateral sclerosis (ALS) is devastating. Yet our understanding of the neural basis of speech—and our efforts to restore this ability—have been limited because of the complexity of speaking and the lack of animal models.

Despite these challenges, neuroscientists have started to unravel the neural code for speech production thanks to opportunities to directly measure people's brain activity during medical procedures such as epilepsy treatment (1). Such studies revealed that computer algorithms could predict the content of speech from simultaneously recorded neural signals (2).

These discoveries, when combined with 20 years of brain-computer interface (BCI) research showing that the brain still encodes movement intentions even after years of paralysis (3), suggested that it might be possible to build a speech neuroprosthesis: a device that directly translates brain activity into words, bypassing injured parts of the nervous system (4). By combining neuroscience and artificial intelligence (AI), this effort has made swift progress in just a few years (5, 6).

Our team recently demonstrated an AI-based speech neuroprosthesis capable of restoring day-to-day communication (7). We worked with a 45-year-old man who could no longer speak intelligibly because of ALS. He volunteered for our academic clinical trial and had 256 microelectrodes implanted across three speech-related regions of his brain (cortical areas 4, 6, and 55b). We found that across the population of measured neurons, distinct spatiotemporal patterns of neural activity occurred for all 39 English phonemes. This finding enabled us to train deep learning algorithms to predict what phoneme the participant was trying to say at any given moment. Language models then strung these phoneme sequences together into words and sentences.

On the first day of its use, the neuroprosthesis achieved 99.6% accuracy with a limited 50-word vocabulary. Given the high signal-to-noise ratio of our intracortical sensors, which capture individual neuronal action potentials, the initial calibration required only 30 min of recording while the participant attempted to speak. On the second day, we expanded the available vocabulary to the whole English dictionary (125,000 words) with 90.2% accuracy. The participant's words appeared on-screen as soon as he tried to say them. When he finished a phrase, it was vocalized by text-to-speech software trained on old audio recordings to sound like his pre-ALS voice. That very day, the first thing he did was use the neuroprosthesis to speak to his 4-year-old daughter, who until then had no memories of being able to understand her dad.

After further training, the speech neuroprosthesis sustained 97.5% accuracy (a 90% reduction in word error rates compared with previous studies), month after month. Each completed sentence was added to a database used for continuous online learning by the AI models, thereby maintaining performance over time. This growing dataset recently allowed us to deploy more powerful (but data-hungry) algorithms to achieve >99% word accuracy (8). Two years in, the participant has used this technology almost daily to say >2.7 million words. The neuroprosthesis has allowed him to chat with friends and family and even resume working full time in climate advocacy by videoconferencing with and writing to his colleagues.

In marked contrast to earlier neuroprostheses, speech decoding would not have been possible without AI. BCIs initially benefited from a bit of beginner's luck: Simple algorithms that assumed each neuron's activity just represented a preferred movement direction—a schema long known to be grossly incomplete—nonetheless proved surprisingly effective at restoring people's ability to control a computer cursor or robotic arm (9, 10). But those methods broke down for the more complex mapping between neural activity and speech. Fortunately, progress in deep learning, especially as applied to automated speech recognition and machine translation, arrived just in time. AI provided powerful tools that could map inputs (neural measurements) to outputs (phonemes) without precise time alignment, exploit longer-timescale context, and leverage the predictive structure of language (11–13).

Our speech neuroprosthesis provides a life-changing way to communicate, but it lacks the full range of human expression afforded by our voices. Even when the resulting text is read aloud by a computer, this approach does not capture the prosody that we use to convey emotional nuance. It also does not enable the quick back-and-forth of interactive conversation or the vocal flexibility needed for singing. Fortunately, these paralinguistic elements of voice can also be decoded from cortical activity (14). Thus, the next frontier for restoring speech is to develop a BCI that essentially acts as a digital vocal tract, transforming neural activity into sounds almost instantaneously.

Real-time voice restoration is very challenging. The need for low latency and vocal flexibility precludes the accumulation of neural measurements over longer time windows or the use of language models to boost accuracy. Nonetheless, our group has made substantial progress. We first characterized how intonation is encoded in the brain of this same participant and then developed a state-of-the-art causal Transformer-based BCI to decode both his phonetic and paralinguistic voice features with a latency of just 30 ms from action potential to sound (15). This BCI has enabled him to not only hear himself speak in real time but also modulate his synthesized voice's intonation to ask questions or emphasize specific words. He even sang a simple melody.

Further work is needed to establish whether such performance will replicate for patients with a total inability to speak and to make immediately synthesized voice consistently intelligible. AI will undoubtedly play a crucial and rapidly improving role in this quest to restore the full speed, precision, and expressive richness of patients' lost voices. In the future, AI (in

particular, the inner structure of large language models) might even provide us with clues about how to decode meaning, rather than speech, for patients with more extensive damage to the brain's speech formation system.

REFERENCES AND NOTES

1. K. E. Bouchard, N. Mesgarani, K. Johnson, E. F. Chang, *Nature* **495**, 327 (2013).
2. S. Kellis *et al*., *J. Neural Eng.* **7**, 056007 (2010).
3. L. R. Hochberg *et al*., *Nature* **442**, 164 (2006).
4. D. A. Moses *et al*., *N. Engl. J. Med.* **385**, 217 (2021).
5. S. L. Metzger *et al*., *Nature* **620**, 1037 (2023).
6. F. R. Willett *et al*., *Nature* **620**, 1031 (2023).
7. N. S. Card *et al*., *N. Engl. J. Med.* **391**, 609 (2024).
8. N. S. Card *et al*., *Nat. Med.* 10.1038/s41591-026-04414-6 (2026).
9. C. Pandarinath *et al*., *eLife* **6**, e18554 (2017).
10. J. L. Collinger *et al*., *Lancet* **381**, 557 (2013).
11. A. Graves, S. Fernández, F. Gomez, J. Schmidhuber, in *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA (Association for Computing Machinery, 2006), pp. 369–376.
12. A. Graves, A. Mohamed, G. Hinton, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing,* Vancouver, BC (2013), pp. 6645–6649.
13. A. Vaswani *et al*., in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA (2017), pp. 6000–6010.
14. B. K. Dichter, J. D. Breshears, M. K. Leonard, E. F. Chang, *Cell* **174**, 21 (2018).
15. M. Wairagkar *et al*., *Nature* **644**, 145 (2025).