



Published in final edited form as:

JAMA. 2020 February 04; 323(5): 413–414. doi:10.1001/jama.2019.19813.

Toward a Speech Neuroprosthesis

Edward F. Chang, MD,

Department of Neurological Surgery, University of California, San Francisco.

GopalaK. Anumanchipalli, PhD

Department of Neurological Surgery, University of California, San Francisco.

Spoken communication is a basic human function. As such, loss of the ability to speak can be devastating for affected individuals. Stroke or neurodegenerative conditions, such as amyotrophic lateral sclerosis, can result in paralysis or dysfunction of vocal structures that produce speech. Current options are assistive devices that use residual movements, for example, cheek twitches or eye movements, to navigate alphabet displays to type out words.¹ While some users depend on these alternative communication approaches, these devices tend to be slow, error-prone, and laborious. A next generation of rehabilitative technologies currently being developed, called brain-computer interfaces (BCIs), directly read out brain signals to replace lost function. The application of neuroprostheses to restore speech has the potential to improve the quality of life of patients with neurological disease, but also including patients who have lost speech from vocal tract injury (eg, from cancer or cancer-related surgery).

Many potential approaches exist for reading out brain activity toward restoring communication through a neuroprosthesis. While both noninvasive and intracranial approaches are being explored, an approach using neurophysiological recordings of neuronal activity measured from electrodes either directly on the brain surface or from thin microwire electrode arrays inserted into the cortex has provided encouraging results. Most approaches have adopted the traditional augmentative and alternative communication strategy by restoring communication using the neuroprosthesis to control a computer cursor, usually by decoding neural signals associated with arm movements, to type out letters one by one. However, the best rates for spelling out words are still under 10 words per minute, despite rapid cursor control by some individuals.² This may represent fundamental limitations in the approach of using a single cursor to spell out words, rather than the ability to accurately read out brain activity. There is a need to substantially improve the accuracy and speed of BCIs to begin to approach natural speaking rates (120–150 words per minute in healthy speakers). The Figure compares the communication rates across various modalities.^{2,3}

Speech is among the most complex motor behaviors and has evolved for efficient communication that is unique to humans. A defining aspect of speech is the rapid transmission of information, ranging from brief, informal conversations to communicating complex ideas, such as in a formal presentation. One reason speech can carry so much

information is that the speech signal is generated by the precise and coordinated movements of approximately 100 muscles throughout the vocal tract, giving rise to the repertoire of speech sounds that make up a given language.

The key to improving communication BCIs is in the neuroscientific understanding of how the brain controls the vocal tract during speech. For example, the motor map of the human homunculus contains neuronal populations that are involved in executing voluntary control of the larynx, lips, jaw, and tongue. While these representations underlie many functions, such as swallowing and kissing, they are specialized in humans for producing speech features, such as consonants, vowels, and prosodic intonation. In recent years, understanding has significantly deepened from a general idea of the brain location in which functions are located to the more fundamental question of how these patterns are generated by neural substrates.

Recent discoveries have been enabled by high-resolution (eg, millimeter and millisecond) neurophysiological recordings in humans, for example, in patients with epilepsy who volunteered for research studies involving implanted brain electrodes for localizing a seizure focus. These rare opportunities have yielded discoveries that neural commands produce vocal tract “gestures,” low-dimensional coordinative patterns of movement.⁴ Gestures produce specific shapes in the vocal tract, for example, the closure of the lips and jaw to make a “p” sound. These gestures are sequenced together to produce fluent sentences. A natural application of these insights is to decode speech from brain activity.

A recent report indicated that it is possible to synthesize speech by decoding directly from human cortex while study participants spoke full sentences. Brain signals drove the gestural movements of a computational “virtual vocal tract” to generate audible speech (Video).⁵ It has also been shown to be possible to translate brain signals into text in real time.⁶ While these developments are promising, several challenges and opportunities exist in realizing high-performance speech BCIs.

Most demonstrations of successful speech decoding have been carried out among study participants with intact speech function. In such contexts, actual speaking was used to train decoding algorithms. A major challenge is how to achieve similar performance in people who are paralyzed and no speech data are available. Imagined speech does not appear to be sufficient for decoding, and the neural code for inner speech or pure thoughts is not clear at this time. Learning to control a speech neuroprosthesis may be possible, but would be akin to relearning how to speak, if not much more difficult. As a result, one potential option is to use a person’s native neural code for speech, which is presumably dormant in paralyzed individuals. Further, closed loop real-time feedback has demonstrated promise in other neuroprosthetic applications and might also have a critical role for speech neuroprostheses. While previous work has focused on the motor cortical areas for speech articulation, ongoing neuroscience studies may provide insights into how other brain regions contribute to speaking. For example, the Broca area is thought to be involved in high-order aspects of language processing such as grammatical sequencing and speech planning. Sampling neural activity in these brain areas may provide additional information to complement or even bypass articulatory commands.

Another major avenue for improving BCI performance is artificial intelligence (AI). Breakthroughs in AI, such as human-level speech recognition and naturalistic speech generation in computers, are made possible by deep learning.⁷ Deep learning involves computational modelling of multiple layers of abstractions in relating diverse data representations. Traditional deep learning relies on large scale, high-quality data sets to train state-of-the-art decoding models. For example, commercial speech recognizers (eg, Apple's Siri) are trained on more than 10 000 hours of speech and text data to match human-level transcription of speech. While this would be prohibitive for paralyzed individuals, biomimetic architectures that computationally simulate natural physiology of speech movements and enforce meaningful neural representations have reduced the need for large amounts of training data.⁵

Another promising direction involves “transfer-learning” protocols, whereby models trained on healthy individuals can be used to initialize neural decoders for paralyzed patients. It may be possible to create models with shared latent or abstracted representations so that data may be more efficiently (re)used across individuals. Further gains may be obtained by predictive text modeling to fix spelling mistakes or complete words and sentences before they are typed out like the autocorrect function on a smartphone.

Speaking is an ability most humans take for granted, until it is lost by injury or surgery. For example, severe dysarthria precludes communication because the transmission of speech commands or the vocal tract is impaired. Neuroprosthetic technologies have the potential to help reduce some of the negative aspects related to the disability and improve quality of life by potentially enabling independence, social interactions, and community involvement. The current convergence of AI, speech neuroscience, and neural interface technologies may allow speech neuroprostheses to achieve naturalistic communication rates and accuracies in the future.

Conflict of Interest Disclosures:

Drs Chang and Anumanchipalli reported receiving research support from National Institutes of Health grant U01 NS098971-01, Facebook Reality Labs, New York Stem Cell Foundation, the Howard Hughes Medical Institute, the McKnight, Foundation, the Shurl and Kay Curci Foundation, and the William K. Bowes Foundation. They have patents pending for technology related to speech decoding.

REFERENCES

1. Koch Fager S, Fried-Oken M, Jakobs T, Beukelman DR. New and emerging access technologies for adults with complex communication needs and severe motor impairments: state of the science. *Augment Altern Commun.* 2019;35(1):13–25. doi:10.1080/07434618.2018.1556730 [PubMed: 30663899]
2. Pandarinath C, Nuyujukian P, Blabe CH, et al. High performance communication by people with paralysis using an intracortical brain-computer interface. *Elife.* 2017;6:e18554. doi:10.7554/eLife.18554 [PubMed: 28220753]
3. Mugler EM. Investigation of Speech for Communicative BrainComputer Interface [doctoral dissertation]. Chicago, IL: University of Illinois at Chicago; 2014.
4. Chartier J, Anumanchipalli GK, Johnson K, Chang EF. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron.* 2018;98(5):1042–1054.e4. doi:10.1016/j.neuron.2018.04.031 [PubMed: 29779940]

5. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. *Nature*. 2019;568(7753):493–498. doi:10.1038/s41586-019-1119-1 [PubMed: 31019317]
6. Moses DA, Leonard MK, Makin JG, Chang EF. Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nat Commun*. 2019;10(1):1–14. doi:10.1038/s41467-019-10994-4 [PubMed: 30602773]
7. Hinton G. Deep learning: a technology with the potential to transform health care. *JAMA*. 2018;320(11):1101–1102. doi:10.1001/jama.2018.11100 [PubMed: 30178065]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

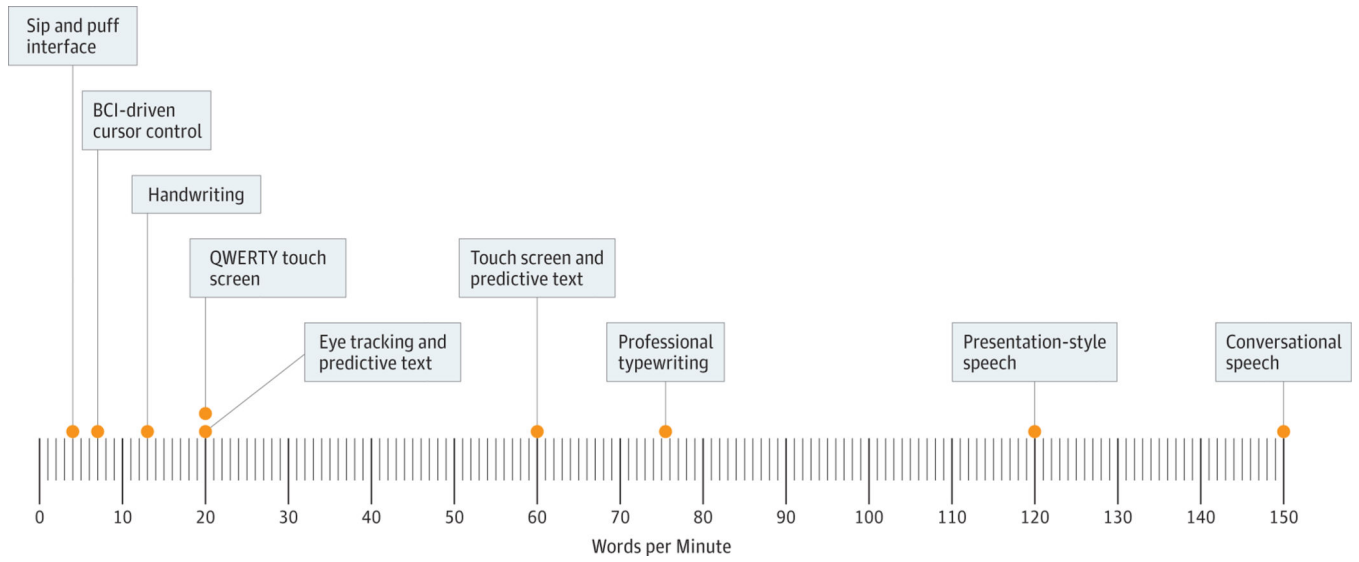


Figure. Comparison of Communication Rates Across Various Modalities, Measured as the Average Number of Words per Minute in a Typical Scenario
BCI indicates brain-computer interface.

Video.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript