An Accurate and Rapidly Calibrating Speech Neuroprosthesis

Nicholas S. Card, Ph.D.[1], **Maitreyee Wairagkar, Ph.D.**[1], **Carrina Iacobacci, B.S.**[1], **Xianda Hou, M.S.**[1,2], **Tyler Singer-Clark, B.S.**[1,3], **Francis R. Willett, Ph.D.**[4,8], **Erin M. Kunz, M.S.**[5,6], **Chaofei Fan, M.S.**[7], **Maryam Vahdati Nia, M.S.**[1,2], **Darrel R. Deo, Ph.D.**[4], **Aparna Srinivasan, M.S.**[1,3], **Eun Young Choi, Ph.D.**[4], **Matthew F. Glasser, M.D., Ph.D.**[9], **Leigh R Hochberg, M.D., Ph.D.**[10,11,12], **Jaimie M. Henderson, M.D.**[4,6], **Kiarash Shahlaie, M.D.**[1], **Sergey D. Stavisky, Ph.D.**[1], **David M. Brandman, M.D., Ph.D.**[1]

[1]Department of Neurological Surgery, University of California Davis, Davis, CA

[2]Department of Computer Science, University of California Davis, Davis, CA

[3]Department of Biomedical Engineering, University of California Davis, Davis, CA

[4]Department of Neurosurgery, Stanford University, Stanford, CA

[5]Department of Electrical Engineering, Stanford University, Stanford, CA

[6]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA

[7]Department of Computer Science, Stanford University, Stanford, CA

[8]Howard Hughes Medical Institute, Stanford University, Stanford, CA

[9]Departments of Radiology and Neuroscience, Washington University School of Medicine, Saint Louis, MO

[10]School of Engineering and Carney Institute for Brain Sciences, Brown University, Providence, RI

[11]VA RR&D Center for Neurorestoration and Neurotechnology, VA Providence Healthcare, Providence, RI

[12]Center for Neurotechnology and Neurorecovery, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA

Abstract

Background: Brain-computer interfaces can enable communication for people with paralysis by transforming cortical activity associated with attempted speech into text on a computer screen. Communication with brain-computer interfaces has been restricted by extensive training requirements and limited accuracy.

Methods: A 45-year-old man with amyotrophic lateral sclerosis (ALS) with tetraparesis and severe dysarthria underwent surgical implantation of four microelectrode arrays into his left precentral gyrus, which recorded neural activity from 256 intracortical electrodes 5 years after the

Corresponding authors: 1. Nicholas Card - nscard@ucdavis.edu, 2. Sergey Stavisky - sstavisky@ucdavis.edu, 3. David Brandman - dmbrandman@ucdavis.edu, Nicholas S. Card, Ph.D., Department of Neurological Surgery, University of California Davis, Davis, CA nscard@ucdavis.edu.
Sergey D. Stavisky and David M. Brandman contributed equally.

onset of his illness. We report the results of decoding his cortical neural activity as he attempted to speak in both prompted and unstructured conversational settings. Decoded words were displayed on a screen, then vocalized using text-to-speech software designed to sound like his pre-ALS voice.

Results: Twenty-five days after surgery, on the first day of system use and following 30 minutes of collection of cortical recordings and processing while the participant attempted to speak, the neuroprosthesis achieved 99.6% accuracy with a 50-word vocabulary. On the second day, after 1.4 additional hours of system training, the neuroprosthesis achieved 90.2% accuracy using a 125,000-word vocabulary. With further training data, the neuroprosthesis sustained 97.5% accuracy for self-paced conversations for over 248 cumulative hours over 8.4 months after surgical implantation.

Conclusions: In an individual with ALS and severe dysarthria, an intracortical speech neuroprosthesis reached a level of performance suitable to restore naturalistic communication after brief training (ClinicalTrials.gov number: NCT00912041).

Introduction:

Communication is a priority for people with dysarthria from neurological disorders such as stroke and amyotrophic lateral sclerosis (ALS)[1]. People with diseases that impair communication report increased rates of isolation, depression, and decreased quality of life[2,3]; losing communication may determine if a person will pursue or withdraw life-sustaining care in advanced ALS[4]. While existing augmentative and assistive communication technologies such as head or eye trackers are available, they have low information transfer rates and become increasingly difficult to use as patients lose voluntary muscle control[5]. Brain-computer interfaces are a promising communication technology that can directly decode the user's intended speech from cortical neural signals[6]. Efforts to develop a speech neuroprosthesis are built largely on studies using data that are retrospectively analyzed from able-bodied speakers undergoing electrophysiological monitoring for clinical purposes[7–16]. Several groups have performed real-time brain-computer interface studies to restore lost speech using implanted electrocorticography (ECoG)[17–20], including a report published in the Journal[17], or intracortical multielectrode arrays[21]. Two recent reports have established 'brain-to-text' speech performance[19,21] by decoding cortical neural signals generated by attempted speech into phonemes (the building blocks of words) and assembling these phonemes into words and/or sentences displayed on a computer screen. These studies achieved communication performance, quantified by word error rates, of 25.5% with a 1,024-word vocabulary[19] and 23.8% with a 125,000-word vocabulary[21] and required approximately 17 hours of recording to collect sufficient training data to obtain that level of performance.

We report an intracortical speech neuroprosthesis that ultimately provided access to a 125,000-word vocabulary, with low training data requirements in a participant with advanced ALS and severe dysarthria that achieved high accuracy with useful function beginning on the first day of use, 25 days after implantation.

Methods:

Study participant

A 45-year-old left-handed man with amyotrophic lateral sclerosis (ALS) had symptoms beginning 5 years before enrollment into this study. At the time of enrollment, he was non-ambulatory, dependent on others for controlling his electric wheelchair, dressing, eating, and hygiene, had severe dysarthria, and had an ALS Functional Rating Scale Revised (ALSFRS-R) score of 23 (range 0 to 48 with higher scores indicating better function). For 8 months following surgical placement of recording arrays, he has maintained a modified mini-mental status exam score of 27 (range 0 to 27, with 27 being the highest score attainable). At the time of this report, he retains eye and neck movements but has limited orofacial movement with a mixed upper- and lower-motor neuron dysarthria resulting in monotone, low-volume, nasal speech. He requires non-invasive respiratory support at night and does not have a tracheostomy. When his speech is being listened to by people who are not his regular care partner, he is unintelligible (Audio 1): his oral motor tasks on the Frenchay Dysarthria Assessment-2 were an "E" rating (a measure of several speech behaviors, range A to E, with A representing normal function and E, no function), representing profound dysarthria. When speaking to expert listeners, he has communicated at 6.8 ± 5.6 (mean ± standard deviation) correct words per minute (conversational English is approximately 160 words per minute[22]). His typing speed using a gyroscopic headmouse (Zono 2, Quha, Nokia, Finland) has been 6.3 ± 1.3 correct words per minute (Fig. S1). The severity of dysarthria has remained stable during the period of this report, including the immediate postoperative period. Additional participant details are in Section S1.01 of the Supplementary Appendix, available with the full text of this paper.

There have been 19 participants in the BrainGate and ongoing BrainGate2 clinical trials, which historically focused on decoding attempted arm and hand movements from related areas of cortex. Following the recent evolution of the trial to include recording from speech areas of cortex, results from one prior participant for a speech neuroprosthesis have been reported[21]; that participant had only two arrays implanted in precentral gyrus (and two in inferior frontal gyrus) rather than four arrays in precentral gyrus as in the current particpant.[21]

Surgical implantation

We implanted four microelectrode arrays (NeuroPort Array, Blackrock Neurotech, Salt Lake City, Utah, USA) into the left precentral gyrus, an important cortical region for coordinating motor activities related to speech[17,19,21]. Each microelectrode array is 3.2×3.2 mm, has 64 electrodes in an 8×8 grid arrangement inserted 1.5 mm into the cortex using a specialized high speed pneumatic inserter. Each electrode has one recording site of ~50 μm size and is designed to record from a single or a small number of cortical neurons. Implantation was through a left-sided 5×5 cm craniotomy under general anesthesia. Care was taken to avoid placing the microelectrode arrays through large vessels on the cortical surface that were identified by visual inspection. Two arrays are connected to one percutaneous connector ("pedestal") designed to transmit the neural recordings to external computers. Two percutaneous pedestals, each secured to the skull with titanium screws, provided for

recording from a total of 256 sites. Reference wires were placed in both the subdural and epidural spaces. The pedestals were connected by detachable connectors that used HDMI cables to transmit data to computers (Fig. 1a). These computers sat on a wheeled cart and were connected to standard electrical wall outlets.

The surgical implantation was in July 2023, had no serious adverse events, and the participant was discharged on postoperative day 3. Non-serious adverse events, including incisional pain and transient increased frequency of muscle spasms from spasticity that he had experienced for months before implantation, are listed in Section S1.01. From initial incision to closure, the operation took 5 hours. We began collecting data in August 2023, 25 days after surgery.

Recording array locations and decoding contributions

Prior to implanting arrays in the precentral gyrus, we identified the central sulcus by MRI and confirmed that the participant was left-hemisphere language dominant by functional MRI despite being left-handed, using standard clinical fMRI tasks (sentence completion, silent word generation, silent verb generation, and object naming). We refined the implantation targets using the Human Connectome Project's multi-modal MRI-derived cortical parcellation precisely mapped to the participant's brain[23] (Fig. 1b) (Supplementary Fig. S2, Section S1.02; Figure S.11 shows the estimated locations on the Montreal Neurological Institute template brain). We targeted language-related Broadmann area 55b[24] (an area identified in the Human Connectome Project as implicated in phonologic representation) and three areas in the precentral gyrus associated with speech production: dorsal and ventral aspects of the ventral premotor cortex (d6v, v6v, respectively), and primary motor cortex (Brodmann area 4; Fig. S2). Our choice of targeting speech motor cortex was informed by our previous study in another aforementioned individual that found two arrays in 6v provided informative signals for speech decoding[21].

Real-time acquisition and processing of cortical recordings

A signal processing system (NeuroPort System, Blackrock Neurotech) was used to acquire signals from the two connector pedestals (Fig. S3) and send them to a series of commercially available computers running our publicly available software[25] (Section S1.5) for real-time signal processing (Section S1.4) and decoding (Sections S2, S3). Blackrock Neurotech was not involved in the data collection or reporting in this study and had no oversight regarding the decision to publish these results. There were no agreements of any kind between the authors and the commercial entity. The devices used were purchased for research use and not provided by a commercial entity.

Speech task designs

We collected data in 84 sessions over 32 weeks (Section S1.06; Table S2) in the participant's home. No more than one session was performed on any study day. Each study session consisted of a series of task blocks, lasting approximately 5–30 minutes, wherein the participant used the neuroprosthesis. Between blocks, he would take breaks, eat meals, etc. During each block, the participant used the system in two different ways: 1) an instructed-delay Copy Task (Videos 1, 2 and Section S1.07); and 2) a self-paced Conversation

Mode (Videos 4, 5 and Section S1.08). The instructed-delay task consisted of words being presented on a computer screen, and the participant attempting to say the words after a visual/audio cue[21]. The self-paced Conversation Mode involved the participant attempting to say whatever he wanted (although the computer outputs were limited to a 125,000-word dictionary) in an unstructured conversational setting. In both tasks, speech decoding occurred in real-time; as he spoke, the cortical activity at the four micro-electrode arrays were recorded and decoded, and the predicted words were presented on the screen. Completed sentences were read aloud by a computer program and, in later sessions, automatically punctuated (Sections S4 and S3.03). The neuroprosthesis could also send the sentence to the participant's personal computer by acting as a Bluetooth keyboard, which allowed him to use it for activities such as writing emails (Section S1.08). Sampled phoneme and words used for decoder training accumulated over the course of the study (Figure S4).

Decoding speech

No microphone input was used for decoding and we found no evidence of acoustic or vibration-related contamination in the recorded neural signals (Section S4.03, Fig. S7). Every 80 ms, the activity from the cortical recordings was used to predict the most likely English phoneme being attempted (Section S2, Figs. S5, S6). Phoneme sequences were then combined into words using an openly available language model[21]. Next, we applied two further open-source language models to translate the sequence of words initially predicted from the neural activity into the most likely English sentence (Section S3, Fig. S8), as described in a previous report[21]. Data from multiple days were combined to continuously calibrate the decoder (Section S2 and Figs. S9, S10).

Evaluation

We used two measures to analyze the speech decoding performance: phoneme error rate and word error rate, consistent with previous speech decoding studies[17,19,21]. These measures are the ratio of phonemic and word errors to the total number of phonemes or words expected to be decoded, respectively. An error is defined as the need for an insertion, deletion, or substitution to have the decoded sentence match the intended sentence (e.g., the prompted text in the Copy Task). The phoneme error rate can be understood as the system's ability to translate cortical neural activity into phonemes without language models ('raw' phoneme error rate in figures), and word error rate as an estimate of overall communication accuracy. When evaluating accuracy during the Copy Task, the correct text was the prompted text shown to the participant. When evaluating accuracy during self-initiated conversation, we used a combination of methods to identify the intended sentence, including asking the participant after the session what he meant to say (Supplementary Section S1.09). We also report estimated sentence-level accuracies by having the participant use an eye-tracker to select on-screen buttons corresponding to whether the preceding output text was "100% correct", "mostly correct", or "incorrect" (Supplementary Section S1.08). Data was collected in continuous blocks (lasting 5–30 minutes in length as described above), separated by short breaks. Blocks were either "training blocks" in which data were collected for decoder training and optimization or predetermined "evaluation blocks" used to measure and report performance. Error rates were aggregated over all evaluation sentences for each session (Section S1.09). The first-ever closed-loop block (session 1) was excluded from evaluation

because the participant cried with joy as the words he was trying to say correctly appeared on-screen; the research team paused the evaluation until after the participant and his family had a chance to celebrate the moment.

Statistical analyses

Results for each analysis are presented with 95% confidence intervals or as mean ± standard deviation. Confidence intervals were estimated by randomly resampling each dataset 10,000 times with replacement and have not been adjusted for multiplicity. The evaluation measures for decoding performance, phoneme error rate and word error rate, both measured as Levenshtein distance (an estimate of the minimum number of edits required to correct a sequence), were chosen before the start of data collection (Section S1.09).

Results:

Online decoding performance

In the first session, the participant attempted to speak prompted sentences constructed from a 50-word vocabulary[17]. We recorded 213 sentences over 30 minutes of the Copy Task, which were used to calibrate the speech neuroprosthesis. Next, we decoded his neural cortical activity in real-time as he tried to speak. The neuroprosthesis decoded his attempted speech with a word error rate of 0.44% (95% confidence interval [CI], 0.0% to 1.4%). We retested this result for 50-word vocabulary decoding in the second research session, in which all of the participant's attempted sentences were decoded correctly (0% word error rate; Fig. 2).

In this second research session, we expanded the vocabulary of the neuroprosthesis from 50 words to over 125,000 words, which encompasses the majority of the English language[26]. We collected an additional 260 sentences of training data over 1.4 hours. After being trained on these additional sentences, the neuroprosthesis decoded the participant's attempted speech with a Copy Task word error rate of 9.8% (95% CI, 4.1% to 16.0%; Fig. 2). Performance continued to improve in subsequent research sessions as we collected more training data and adapted innovations for incorporating new data more effectively[27] (Sections S2 & S3). The neuroprosthesis achieved a word error rate of 2.5% (95% CI, 1.0% to 4.5%) by session 15, and this approximate accuracy was maintained through session 84, more than eight months after implant. Average Copy Task decoding performance in the final 5 evaluation sessions had a word error rate of 2.5% (95% CI, 2.0% to 3.1%) at the participant's self-paced speaking rate of 31.6 words per minute (95% CI, 31.2% to 32.0%; Fig. S1), with individual day's average word error rates ranging from 1.0% to 3.3% (Table S3). The neuroprosthesis' communication rate exceeded the participant's standard means of communication using a head mouse or skilled interpreter (Fig. S1a).

The performance of neural decoding was maintained across days (Figs. S14, S15). The arrays in the ventral premotor cortex and middle precentral gyrus contributed most to decoding accuracy (Fig. S16). The neuroprosthesis decoded words it was not explicitly trained on (Fig. S18), and worked across different attempted speaking amplitudes, including non-vocalized speech (Fig. S19).

Conversational speech using the brain-computer interface

During conversational speech, the neuroprosthesis detected when the participant started or stopped speaking (Section S1.08, Fig. S21). When evaluated offline on Copy Task sentences (where we knew when he was or was not trying to speak), the system falsely detected that he wanted to speak on less than 1% of sentences (Fig. S22). Additionally, the participant had the option to use an eye tracker for selecting actions (Fig. 3a) to finalize and read aloud a sentence, to indicate whether the neuroprosthesis output was correct, or to initiate a mode where he could spell out words letter-by-letter by attempting to say those letters. This was useful for situations where words were not correctly predicted by the decoder, for example, because they were not in the vocabulary, such as certain proper nouns.

The participant's first use of the neuroprosthesis for naturalistic communication with his family is shown in Video 3 (Fig. S23; Table S4 provides additional transcripts). In subsequent sessions, he utilized the neuroprosthesis for personal use (e.g., Videos 4–5), communicating a total of 22,679 sentences during 72 (out of 84 total) sessions over 8.4 months (248.3 cumulative hours; Fig. 4a). The word error rate during selected Conversation Mode sessions was 3.7% (95% CI, 3.3% to 4.3%; Fig. 4b), and the participant self-reported that 52.9% and 32.3% of all Conversation Mode sentences were decoded correctly or mostly correctly, respectively (Fig. 4c). The longest almost continuous use of the speech neuroprosthesis in Conversation Mode was 7.7 hours. Across the 29 session days, during which the participant used the neuroprosthesis solely for personal use, he requested to recalibrate and update the decoder on three occasions because there were more errors than he was accustomed to. Each calibration took approximately 7.5 minutes, during which twenty Copy Task sentences were displayed to provide training labels and thereby rapidly update the decoder.

The participant used Conversation Mode to perform activities ranging from talking to the research team, family and friends, to performing his occupation by participating in videoconferencing meetings and writing documents and emails. Using the neuroprosthesis, the participant told the research team, "I hope that we are very close to the time when everyone who is in a position like me has the same option to have this device as I do" (Table S4).

Discussion:

Beginning on the first day of device use, 25 days after implantation, a brain-to-text speech neuroprosthesis with 256 cortical recording sites in the left precentral gyrus accurately decoded intended speech in a man with severe dysarthria due to ALS. He communicated using a 125,000-word vocabulary on the second day of use. Within 16 cumulative hours of use, the neuroprosthesis correctly identified 97.3% of attempted words. To contextualize this error rate, the state-of-the-art for English automated speech recognition (e.g., smartphone dictation) has an approximate 5% word error rate[28] and able-bodied speakers have a 1–2% word error rate[29] when reading a paragraph aloud.

The study participant used the speech neuroprosthesis to converse with family, friends, healthcare professionals, and colleagues. His regular means of communication without a

neuroprosthesis involved either (1) having expert caregivers interpret his severely dysarthric speech, or (2) using a head-mouse with point-and-click selections on a computer screen. The investigational system became his preferred way to communicate with our research team, and he used it on his own time, however, a researcher's assistance was required to connect and launch the system. The participant and family indicated that the system's voice resembled his own.

This study demonstrated a reduction in the quantity of training data required to achieve high accuracy decoding compared to our previous study,[21] in which performance was tested starting 113 days post-implant, and used 16.8 hours of training data collected over 15 days, to achieve a word error rate of 23.8%. Another previous speech neuroprosthesis required 17.7 hours of training data collected over 13 days, to reach a word error rate of 25.5%[19].

In addition to recording from two arrays in the putative ventral portion of the speech premotor cortex as in a previous report[21], we also targeted one array each into two areas which, to our knowledge, have not previously been recorded from with multielectrode arrays: Broadmann area 4 (primary motor cortex, which in humans is often in the central sulcus[23] and thus largely not accessible with microelectrode arrays) and area 55b.

This study involved a single participant and it is uncertain if similar results can be expected in future users. The durability of the system as ALS progresses has not been extensively studied and we cannot comment on the use of this system in other disorders.

In an individual with ALS, a rapidly usable and accurate restoration of speech-based communication with an extensive vocabulary was enabled by an intracortical neuroprosthesis. Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
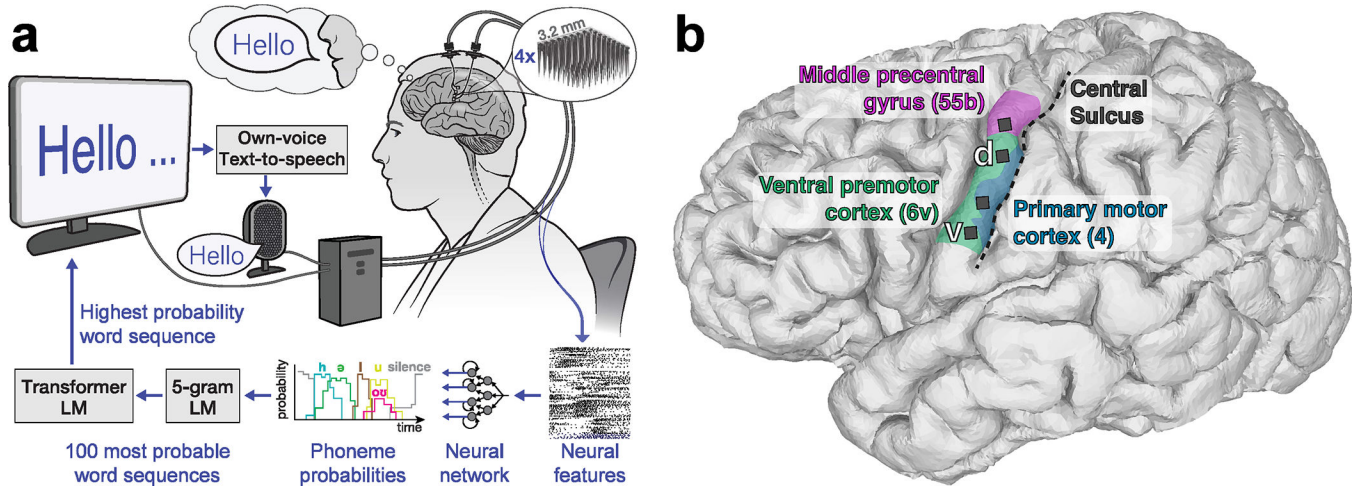
Funding:

References

1. Coppens P. Aphasia and Related Neurogenic Communication Disorders. Jones & Bartlett Publishers; 2016.

2. Katz RT, Haig AJ, Clark BB, DiPaola RJ. Long-term survival, prognosis, and life-care planning for 29 patients with chronic locked-in syndrome. Arch Phys Med Rehabil 1992;73(5):403–8. [PubMed: 1580764]

3. Lulé D, Zickler C, Häcker S, et al. Life can be worth living in locked-in syndrome [Internet]. In: Laureys S, Schiff ND, Owen AM, editors. Progress in Brain Research. Elsevier; 2009 [cited 2023 Dec 11]. p. 339–51. Available from: https://www.sciencedirect.com/science/article/pii/S0079612309177233

4. Bach JR. Communication Status and Survival with Ventilatory Support. Am J Phys Med Rehabil 1993;72(6):343. [PubMed: 8260126]

5. 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 Baltim Md 1985 2019;35(1):13–25.

6. Luo S, Rabbani Q, Crone NE. Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication. Neurotherapeutics 2022;19(1):263–73. [PubMed: 35099768]

7. Herff C, Heger D, de Pesters A, et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. Front Neurosci [Internet] 2015 [cited 2023 Dec 11];8. Available from: https://www.frontiersin.org/articles/10.3389/fnins.2015.00217

8. Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. J Neural Eng 2010;7(5):056007. [PubMed: 20811093]

9. Mugler EM, Patton JL, Flint RD, et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng 2014;11(3):035015. [PubMed: 24836588]

10. Ramsey NF, Salari E, Aarnoutse EJ, Vansteensel MJ, Bleichner MG, Freudenburg ZV. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. NeuroImage 2018;180(Pt A):301–11. [PubMed: 28993231]

11. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature 2019;568(7753):493–8. [PubMed: 31019317]

12. 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):3096. [PubMed: 31363096]

13. Stavisky SD, Willett FR, Wilson GH, et al. Neural ensemble dynamics in dorsal motor cortex during speech in people with paralysis. eLife 2019;8:e46015. [PubMed: 31820736]

14. Stavisky SD, Willett FR, Avansino DT, Hochberg LR, Shenoy KV, Henderson JM. Speech-related dorsal motor cortex activity does not interfere with iBCI cursor control. J Neural Eng 2020;17(1):016049. [PubMed: 32023225]

15. Berezutskaya J, Freudenburg ZV, Vansteensel MJ, Aarnoutse EJ, Ramsey NF, Van Gerven MAJ. Direct speech reconstruction from sensorimotor brain activity with optimized deep learning models. J Neural Eng 2023;20(5):056010.

16. Guenther FH, Brumberg JS, Wright EJ, et al. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. PLOS ONE 2009;4(12):e8218. [PubMed: 20011034]

17. Moses DA, Metzger SL, Liu JR, et al. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. N Engl J Med 2021;385(3):217–27. [PubMed: 34260835]

18. Metzger SL, Liu JR, Moses DA, et al. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. Nat Commun 2022;13(1):6510. [PubMed: 36347863]

19. Metzger SL, Littlejohn KT, Silva AB, et al. A high-performance neuroprosthesis for speech decoding and avatar control. Nature 2023;620(7976):1037–46. [PubMed: 37612505]

20. Luo S, Angrick M, Coogan C, et al. Stable Decoding from a Speech BCI Enables Control for an Individual with ALS without Recalibration for 3 Months. Adv Sci 2023;n/a(n/a):2304853.

21. Willett FR, Kunz EM, Fan C, et al. A high-performance speech neuroprosthesis. Nature 2023;620(7976):1031–6. [PubMed: 37612500]

22. Yuan J, Liberman M & Cieri C. Towards an integrated understanding of speaking rate in conversation. In 9th Intl Conf. on Spoken Language Processing 10.21437/Interspeech.2006-204 (2006).

23. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. Nature 2016;536(7615):171–8. [PubMed: 27437579]

24. Silva AB, Liu JR, Zhao L, Levy DF, Scott TL, Chang EF. A Neurosurgical Functional Dissection of the Middle Precentral Gyrus during Speech Production. J Neurosci 2022;42(45):8416–26. [PubMed: 36351829]

25. Ali YH, Bodkin K, Rigotti-Thompson M, et al. BRAND: A platform for closed-loop experiments with deep network models. Journal of Neural Engineering, 2024.

26. Godfrey JJ, Holliman EC, McDaniel J. SWITCHBOARD: telephone speech corpus for research and development [Internet]. In: [Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing. 1992 [cited 2023 Dec 11]. p. 517–20 vol.1.Available from: https://ieeexplore.ieee.org/document/225858

27. Fan C, Hahn N, Kamdar F, et al. Plug-and-Play Stability for Intracortical Brain-Computer Interfaces: A One-Year Demonstration of Seamless Brain-to-Text Communication. Adv Neural Inf Process Syst 2023;36:42258–70. [PubMed: 38738213]

28. Tüske Z, Saon G, Kingsbury B. On the limit of English conversational speech recognition [Internet]. 2021 [cited 2023 Dec 11]; Available from: http://arxiv.org/abs/2105.00982

29. Thomson D, Besner D, Smilek D. In pursuit of off-task thought: mind wandering-performance trade-offs while reading aloud and color naming. Front Psychol [Internet] 2013 [cited 2023 Dec 11];4. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00360

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 1. Electrode locations and speech decoding setup.

a, Diagram of the brain-to-text speech neuroprosthesis. Cortical neural activity is measured using four 64-electrode arrays. Machine learning techniques decode the cortical neural activity into an English phoneme every 80 ms (see also Section S5). **b**, Approximate microelectrode array locations (gray squares) superimposed on a 3d reconstruction of the participant's brain. Colored regions correspond to cortical areas[22] aligned to the participant's brain using the Human Connectome Project's MRI protocol scans before implantation.
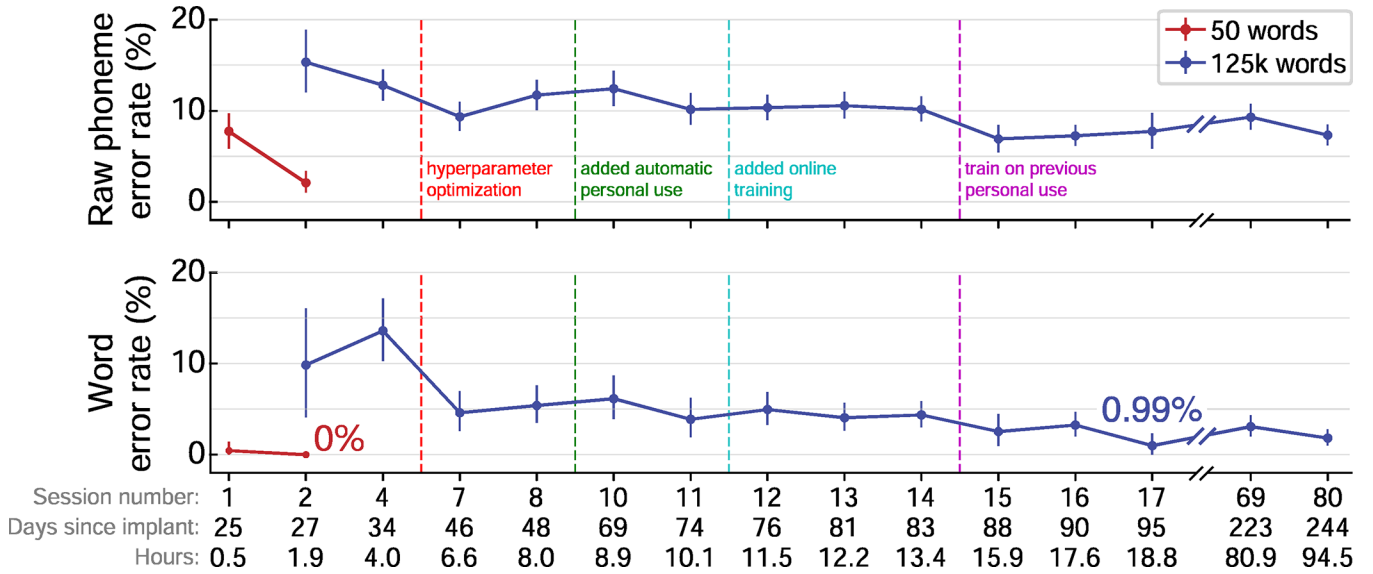
Figure 2. Online speech decoding performance.

Phoneme error rates (top) and word error rates (bottom) are shown for each session for two vocabulary sizes (50 versus 125,000 words). The 'hours' row of the horizontal axis reports the cumulative hours of neural data used to train the speech decoder for that session. Aggregate error rates across all evaluation sentences are shown for each session (mean ± 95% confidence interval). Vertical dashed lines represent when decoder improvements were introduced. Fig. S20 shows phoneme and word error rates for individual blocks.
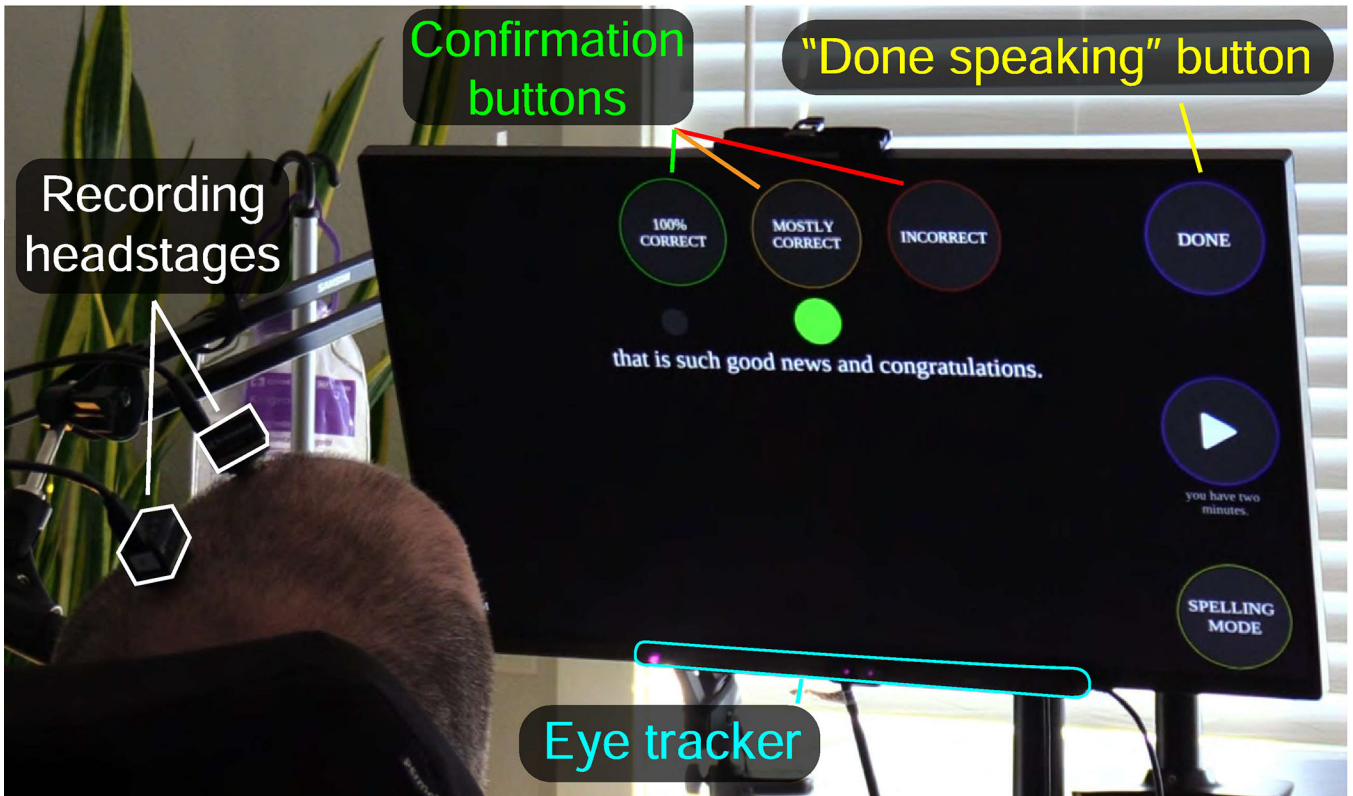
Figure 3. Conversation Mode user interface.
Photograph of the participant and speech neuroprosthesis in Conversation Mode. The neuroprosthesis detected when he was trying to speak solely based on neural activity, and concluded either after 6 seconds of speech inactivity, or upon his optional activation of an on-screen button via eye tracking. After the decoded sentence was finalized, the participant selected on-screen confirmation buttons via eye tracking to indicate if the decoded sentence was correct.
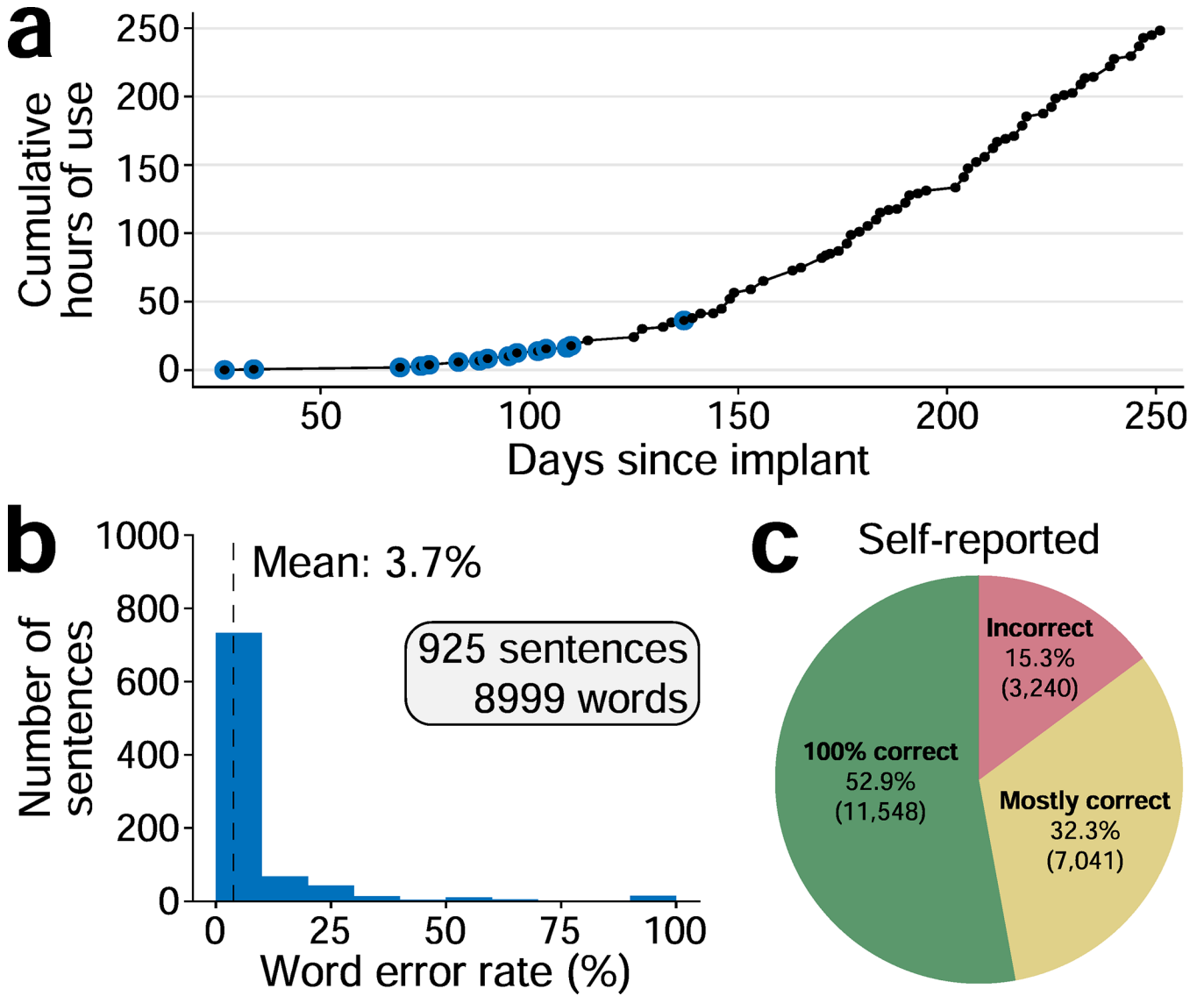
Figure 4. Use of the neuroprosthesis for self-initiated speech.
a, Cumulative hours that the participant used the speech neuroprosthesis to communicate during structured research sessions and personal use. For the sessions outlined in blue, Conversation Mode decoding accuracy was quantified in (**b**). **b,** Histogram evaluating speech decoding accuracy in conversations for the n = 925 sentences with known true labels (Section S1.09). The average word error rate was 3.7% (95% CI, 3.3% to 4.3%). c, Self-reported decoding accuracy for each sentence across all Conversation Mode data (n = 21,829).

Audio 1 - Demonstration of the participant's unintelligible dysarthric speech.
The participant is attempting to say prompted sentences aloud in an instructed delay Copy Task displayed on the screen in front of him (session 10; see Video 2). He retains intact eye movement and limited orofacial movement with the capacity for vocalization, but is unable to produce intelligible speech. At the end of each sentence, the decoded sentence is read aloud by a text-to-speech algorithm that sounds like his pre-ALS voice.

Video 1 - Copy Task speech decoding (session 10).

This video shows the same speech decoding trials as in Audio 1 (session 10). Prompted sentences appear on the screen in front of the participant. When the red square turns green, he attempts to say the prompted sentence aloud while the speech decoder predicts what he is saying in real time. In this video, he is signaling the end of a sentence by using an eye tracker to hit an on-screen "done" button. The participant could also end trials by attempting to squeeze his right hand into a fist, the neural correlates of which were decoded (see Video 2; Section S6). At the end of each trial, the decoded sentence is read aloud by a text-to-speech algorithm that sounds like his pre-ALS voice.

Video 2 - Copy Task speech decoding (session 17).

Another example of Copy Task speech decoding from session 17. In this video, he is signaling the end of a sentence by attempting to squeeze his right hand into a fist, the neural correlates of which are decoded (Section S6).

Video 3 - First self-directed use of the speech neuroprosthesis (session 2).
The participant uses the speech neuroprosthesis to say whatever he wants for the first time (125,000-word vocabulary; session 2). He chose to speak to his daughter; the transcript of what he said is in Fig. S23. Because the dedicated Conversation Mode (and speech detection) had not yet been developed during session 2, the participant waits for the onscreen square to turn from red to green before attempting to speak.

Video 4 - Conversation Mode speech decoding (session 31).

The participant is using the speech decoder in Conversation Mode to engage in freeform conversation with those around him. The audio is muted while conversation partners are speaking for privacy reasons. The speech neuroprosthesis reliably detects when the participant begins attempting to speak, and shows the decoded words on-screen in real time. He can signal the end of a sentence using an on-screen eye tracker button ("DONE" button in the top-right of the screen), or by not speaking for 6 seconds (as he does in this video), after which the neuroprosthesis finalizes the sentence. At the end of each sentence, the decoded sentence is read aloud by a text-to-speech algorithm that sounds like his pre-ALS voice. Finally, the participant uses the eye tracker to confirm whether the output sentence was correct or not. Correctly decoded sentences are used to fine-tune the neural decoder online.

Video 5 - Conversation Mode speech decoding (session 80).
Another example of the participant using the speech decoder in Conversation Mode.
The participant is speaking to a researcher about movies. The participant interface for
Conversation Mode was updated in session 72 to add new features (Section S1.09),
including the ability for him to control when text-to-speech audio is played. The own-voice
text-to-speech model was updated in session 55 to sound more lifelike and closer to the
participant's pre-ALS voice (Section S5.02).