***Title**:* An accurate and rapidly calibrating speech neuroprosthesis

Authors: 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.[4,5,9], Erin M. Kunz, B.S.[5,6], Chaofei Fan, B.S.[7], Maryam Vahdati Nia, M.S.[1,2], Darrel R. Deo, Ph.D.[4], Eun Young Choi, Ph.D.[4], Matthew F. Glasser, Ph.D.[10], Leigh R. Hochberg, M.D., Ph.D.[11,12,13], Jaimie M. Henderson, M.D.[4,8], Kiarash Shahlaie, M.D.[1], David M. Brandman, M.D., Ph.D.*[1], Sergey D. Stavisky, Ph.D.*[1]
* Denotes equal contribution

Corresponding authors:
1. Nicholas Card - nscard@ucdavis.edu
2. Sergey Stavisky - sstavisky@ucdavis.edu

Author affiliations:
[1]Departments of Neurological Surgery, [2]Computer Science, [3]Biomedical Engineering, University of California Davis, Davis, CA, USA; [4]Departments of Neurosurgery, [5]Electrical Engineering, [6]Mechanical Engineering, [7]Computer Science, [8]Wu Tsai Neurosciences Institute, [9]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA; [10]Departments of Radiology and Neuroscience, Washington University School of Medicine, Saint Louis, MO, USA [11]School of Engineering and Carney Institute for Brain Sciences, Brown University, Providence, RI, USA; [12]VA RR&D Center for Neurorestoration and Neurotechnology, VA Providence Healthcare, Providence, RI; [13]Center for Neurotechnology and Neurorecovery, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA

Abstract:
Brain-computer interfaces (BCIs) can provide a rapid, intuitive way for people with paralysis to communicate by transforming the cortical activity associated with attempted speech into text. Despite recent advances, communication with BCIs has been restricted by requiring many weeks of training data, and by inadequate decoding accuracy. Here we report a speech BCI that decodes neural activity from 256 microelectrodes in the left precentral gyrus of a person with ALS and severe dysarthria. This system achieves daily word error rates as low as 1% (2.66% average; 9 times fewer errors than previous state-of-the-art speech BCIs) using a comprehensive 125,000-word vocabulary. On the first day of system use, following only 30 minutes of attempted speech training data, the BCI achieved 99.6% word accuracy with a 50 word vocabulary. On the second day of use, we increased the vocabulary size to 125,000 words and after an additional 1.4 hours of training data, the BCI achieved 90.2% word accuracy. At the beginning of subsequent days of use, the BCI reliably achieved 95% word accuracy, and adaptive online fine-tuning continuously improved this accuracy throughout the day. Our participant used the speech BCI in self-paced conversation for over 32 hours to communicate with friends, family, and colleagues (both in-person and over video chat). These results indicate that speech BCIs have reached a level of performance suitable to restore naturalistic communication to people living with severe dysarthria.

Introduction:

Communication is a top priority for the millions of people living with dysarthria from neurological disorders such as stroke and amyotrophic lateral sclerosis (ALS)[1]. As communication fails, people report increased rates of isolation, depression, and decreased quality of life[2,3]; losing communication often determines if a person will pursue or withdraw life-sustaining care in advanced ALS. Existing augmentative and assistive communication technologies such as eye trackers suffer from low information transfer rates and become increasingly less reliable and more onerous for patients as they lose voluntary motor control[4]. Brain-computer interfaces (BCIs) are a promising assistive technology to meet patients' fundamental need for fast and effortless communication by bypassing the damaged parts of the nervous system and directly decoding their intended speech from neural measurements (reviewed in [5]). Efforts to develop a speech neuroprosthesis are built on a large body of prior work, consisting mostly of offline (post hoc) speech decoding studies using data from able speakers undergoing electrophysiological monitoring for clinical purposes (e.g. [6–14], but see [15]). Several groups have now started closed-loop speech BCI studies specifically to restore lost speech using chronically implanted electrocorticography (ECoG)[16–19] and intracortical multielectrode arrays[20]. Two recent studies have established the state-of-the-art for 'brain-to-text' speech BCIs[18,20] by decoding the neural underpinnings of attempted speech into phonemes (the building blocks of words), which are then assembled into words and sentences using a language model and displayed on a computer screen. These studies achieved communication accuracies – as quantified using the word error rate (WER) metric – of 25.5%[18] and 23.8%[20]. However, as we wrote in our previous study: "it is important to note that it does not yet constitute a complete, clinically viable system… work remains to be done to reduce the time needed to train the decoder… 24% word error rate is probably not yet sufficiently low for everyday use."[20].

Here, we report an intracortical speech neuroprosthesis to meet the need for high accuracy communication (WER below 5%), using a comprehensive vocabulary (125,000 words), with low training data requirements. Our work builds upon prior results[20] with multiple innovations including: (1) doubling the number of electrodes chronically placed in the ventral precentral gyrus to 256; (2) improvements to the language model; (3) online decoder fine-tuning that enables consistently high accuracy decoding over hours of use; (4) a personalized text-to-speech module that reproduces the participant's original voice; and (5) demonstration of self-initiated personal communication with an open vocabulary. We report that these advances resulted in very

high accuracy brain-to-text communication in a person living with severe dysarthria due to ALS, beginning on the very first day of use.

Methods:

Study, participant, and implanted device

We recruited a left-handed male participant in his 40's (referred to as 'SP2' in this preprint rather than the actual trial participant designation, which the participant is familiar with, as per medRxiv policy) with amyotrophic lateral sclerosis (ALS) for the BrainGate2 pilot clinical trial (identifier: NCT00912041). SP2 retains limited orofacial movement with the capacity for vocalization, but is unable to produce intelligible speech (Audio 1). His eye and neck movements remain intact.

Our objective was to translate SP2's attempted speech by decoding his neural signals using four 64-electrode Utah arrays chronically implanted in the precentral gyrus, targeted to brain areas that contributed most to speech decoding from recent studies[16,18,20] using the Human Connectome Project's multi-modal MRI-derived cortical parcellation precisely mapped to SP2's brain[21] (Fig. S1, Section S1.02), and accounting for placement constraints from his brain's anatomy and vasculature (Fig. 2a).

Real-time acquisition and processing of neural data

A signal processing system (NeuroPort System, Blackrock Neurotech) was used to acquire signals from the 256 implanted electrodes and transmit them to a computer running custom software[22] (Section S1.5) for real-time signal processing (Section S1.4), decoding (Sections S2-3), and task control.

Speech task designs

The study consisted of 18 research sessions over the course of 11 weeks (Section S1.06; Table S2) and took place in the participant's home. SP2 engaged in two types of tasks: 1) an instructed-delay Copy Task (Videos 1-2 and Section S1.07), and 2) a self-paced Conversational Task (Video 3 and Section S1.08).

Decoding speech

We used neural activity collected during the speech tasks to train a recurrent neural network (RNN, Section S2) to predict the probability of each English phoneme being spoken. Day-specific input layers were used to correct for nonstationarities between neural data from each research session. Sequences of phoneme probabilities were converted to the most likely word sequence by a multi-stage language model (Section S3), as described in [20].

The RNN and language model ran in real time to convert neural activity during attempted speech into words that appear on a screen. Prior to each session, a new RNN was trained from scratch using all data from previous sessions (Section S2.02). Starting from session 12, we added an 'online training' capability[23], which used new neural data to fine-tune the RNN after each sentence (Section S2.03).

Evaluation

We used two metrics to analyze the speech decoding performance: phoneme error rate (PER) and word error rate (WER), consistent with previous speech decoding studies[16,18,20]. We evaluated our online speech decoding performance only on predetermined "evaluation blocks" (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. To calculate overall decoding performance during the Copy Task, we used all evaluation blocks from the final three sessions. For evaluating the self-paced conversational task, we used all blocks of data.

Statistical analyses

Results for each analysis are presented with 95% confidence intervals or as mean ± standard deviation. The evaluation metrics (phoneme error rate and word error rate) were chosen before the start of data collection.

Results:

Online decoding performance

In the very first research session, we asked participant SP2 to read prompted sentences, which were limited to a 50-word vocabulary[16], while we recorded his neural data. After collecting 213 sentences (30 minutes) of training data, we trained the RNN and switched to the BCI's closed-loop mode, where predicted words appeared on-screen as SP2 attempted to speak. In 50 evaluation sentences, SP2's attempted sentences were decoded with a word error rate (WER) of 0.44%. We replicated this high-accuracy result for 50-word decoding in the second research session, where all 50 of SP2's attempted sentences were decoded completely correct (0% WER; Fig. 1b).

In this second research session, we also expanded the vocabulary of the neuroprosthesis from 50 words to over 125,000 words, which encompasses the majority of the English language. We collected an additional 260 sentences of training data (1.9 hours), which contained a much larger vocabulary of words from conversational English[24]. After incorporating these data into the decoder, the BCI decoded SP2's attempted speech with a WER of 9.8% (Fig. 1b). Decoding performance continued to improve in subsequent research sessions as we collected more training data, optimized algorithm hyperparameters, added online decoder fine-tuning[23], and expanded the training dataset to include personal use data. We reduced the WER to 2.5% by session 15, and 1% by session 17. Average Copy Task decoding performance in the final 3 sessions had a 2.66% WER at SP2's self-paced speaking rate of 32.9 words per minute (Fig. S2).

Notably, the system achieved high accuracy at the start of new research sessions (2-5 days after the previous research session), maintaining an average WER of 4.8% over the first 50 sentences across six sessions (Fig. 1C). This "plug and play" utility is attributed to the increased stability due to a larger number of recording electrodes, and also to the decoder's continuous online fine-tuning[23].

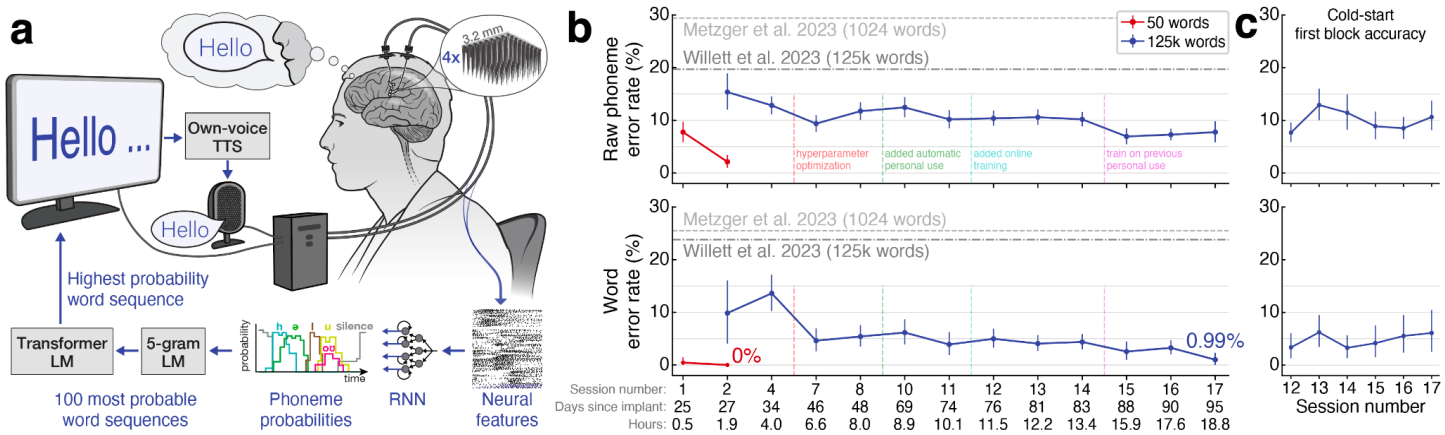

Figure 1. Real-time neural decoding of attempted speech. a, Diagram of the brain-to-text speech BCI system. Neural activity is measured from the left ventral precentral gyrus using four 64-electrode Utah arrays and processed into neural features (threshold crossings and spikeband power), temporally binned, and smoothed. An RNN (five-layer gated recurrent unit architecture) decodes recent neural activity into phoneme probabilities every 80 ms. A two-step language model first employs a large-vocabulary 5-gram model (125,000 words) to convert the phoneme probability sequence into the (up to) 100 most probable word sequences, then a transformer LLM (OPT 6.7B) refines this to the most likely sequence of spoken words. The decoded words are displayed in real-time, and at the end of a sentence, an own-voice text-to-speech algorithm vocalizes the decoded sentence in the participant's pre-ALS voice (Section S5). **b**, Online decoding performance from sessions 1-17. Raw (pre-language model) phoneme error rates (top) and word error rates (bottom) are shown for each session for two vocabulary sizes (50 versus 125,000 words). Vertical blue lines indicate 95% confidence intervals (CIs). Vertical dashed lines represent when associated decoder improvements were introduced. **c**, Raw phoneme error rates (top; average 10.06%, 95% CI: [8.97, 11.21]) and word error rates (bottom; average 4.84%, 95% CI: [3.58, 6.25]) for the first 50 sentences of sessions 13-18. Average WER was 5.0% (95% CI: 2.9%, 7.5%) for the first 20 sentences of each session.

Recording array implant locations and decoding contributions

This performance was enabled by four Utah arrays chronically implanted in the left precentral gyrus (Fig. 2a; Supplemental Fig. S1), targeting putative language-related area 55b, premotor areas dorsal 6v (d6v) and ventral 6v (v6v), and primary motor cortex (area 4) with action potential resolution (Fig. 2b). To identify each array's contribution to speech decoding, we trained decoders with data from one array at a time, or by omitting one array, and evaluated offline the raw (pre-language model) PERs (Fig. 2c). Consistent with our previous findings[20], the ventral 6v array provided the most accurate decoding. The dorsal 6v array's performance was notably worse, while the performance of the 55b and M1 arrays was only slightly worse than ventral 6v. Moreover, phoneme-specific error rates showed differences across arrays but no one array was essential for decoding specific phoneme groups (Fig. 2e). Finally, decoding performance as a function of the total number of electrodes utilized revealed an expected trend: an increase in channel count correlated with higher decoding accuracy, but the gains in performance showed diminishing returns as more electrodes were added (Fig. 2d).

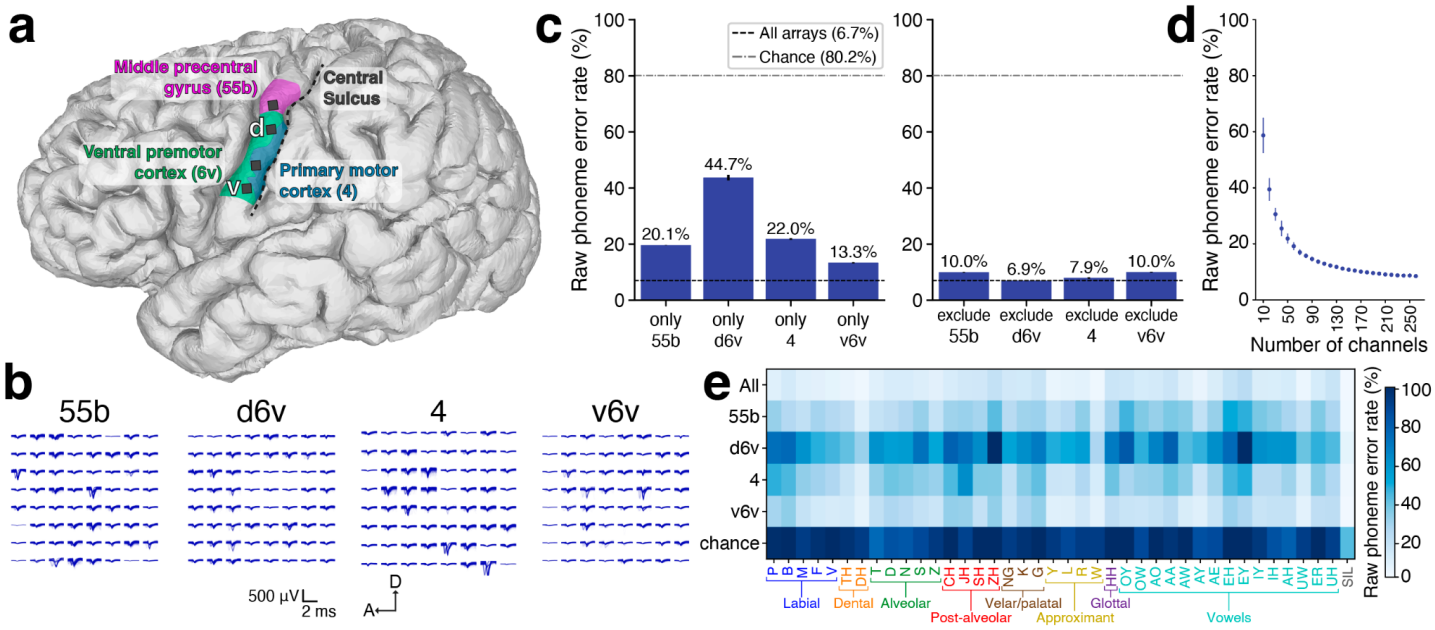

Figure 2. Array locations and role in speech decoding. a, Approximate microelectrode array locations, represented by black squares, superimposed on SP2's 3d brain reconstruction. Colored regions correspond to the Human Connectome Project's multi-modal atlas of cortical areas[21] precisely aligned to SP2's brain using the HCP's MRI protocol scans before implantation. **b**, Representative spike waveforms from a 60-second instructed delay speech task segment. The waveforms show a 1 ms period around the -4.5 RMS threshold crossings. **c**, Analysis of raw phoneme error rates derived from the RNN output. RNNs were trained on data from all sessions and evaluated on randomly-chosen held-out validation trials. *Left,* decoding contribution of each individual array (mean ± standard deviation from 5 RNN seeds). *Right,* performance if any single array were removed. The black dashed line represents decoding performance using all 4 arrays. Omitting the dorsal 6v array did not have a detrimental effect on the raw PER. The gray dashed line represents chance decoding performance (Section S4.01). **d**, An evaluation of decoding efficacy (raw phoneme error rate, mean ± standard deviation) when varying the number of electrodes used (randomly selected from all arrays; Section S4.01). **e**, Individual phoneme decoding accuracy for each array, compared to using all 4 arrays and chance.

Retrospective decoding analyses

Throughout the study, we refined our decoding pipeline several times, which significantly enhanced performance (Fig. 1b). This raises an intriguing question: *how good could performance have been on the first day of speech BCI use, had we used these more refined methods*? A retrospective decoding analysis shows that for a vocabulary of 50 words, we could achieve a 0% WER with just 165 training sentences. For a 125,000-word vocabulary, a WER as low as 8.3% could have been attained after training on 323 sentences (Fig. 3a).

To assess speech decoding stability, we tested (offline) pretrained decoders on data collected on subsequent days without additional fine-tuning. Results showed that fixed decoders maintained high accuracy up to 20 days post-training. Furthermore, decoders trained on larger amounts of data were more stable beyond 20 days (Fig. 3b).

In online evaluation blocks, most words (76.8%) were always decoded accurately, including 66.8% of words that the decoder had never previously encountered (i.e., they were not in the training dataset). This suggests the decoder generalizes well (Fig. 3c, inset). In cases where words were not decoded correctly, we found that the number of occurrences of a word in the decoder training dataset was predictive of the accuracy with which it was decoded (Fig. 3c).

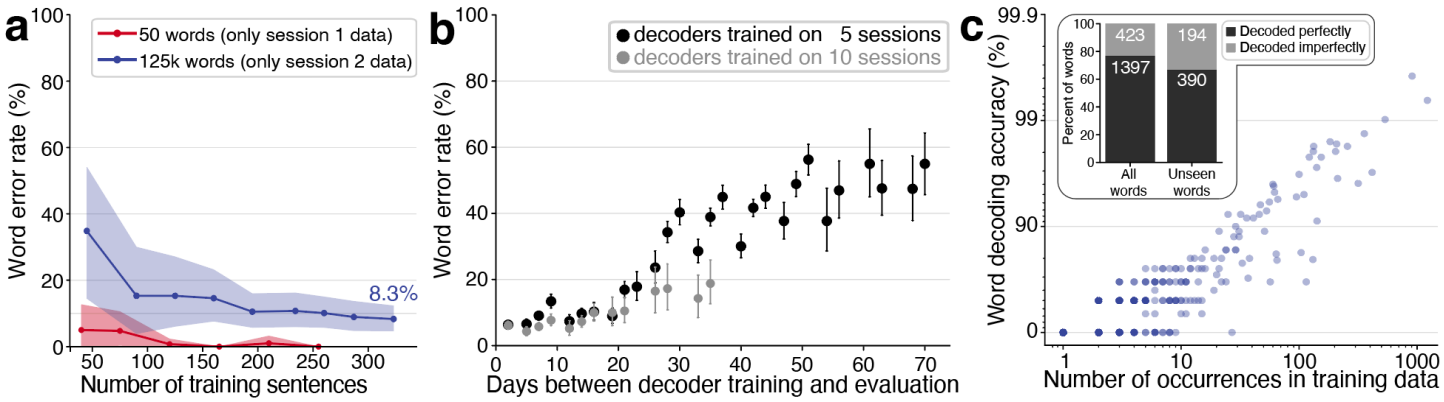

Figure 3. Offline decoding analyses indicate rapidly-calibrating, stable and generalizable decoding. a, Offline recreation of "day 1" performance for 50-word (red) and 125,000-word (blue) vocabularies with optimal decoding hyperparameters. Word error rate is plotted as a function of the number of training sentences. **b,** Decoding stability over time with no recalibration or model fine-tuning. Decoders were trained on data from 5 (black) or 10 (gray) sequential sessions, and then evaluated on all future evaluation blocks. Word error rate is plotted as a function of the number of days between the final day of data used to train each decoder and the date of the evaluation data. **c,** Characterizing how many training examples the system needs to learn to decode words that were initially decoded as incorrect. Word decoding accuracy (log scale) as a function of the number of occurrences of the word in the decoder's training data (log scale), for incorrect words only. After about ten occurrences of a "difficult" word, it could be correctly decoded most of the time. *Inset*, decoding accuracy for all words (*left*, 76.8%) and for words that had never been seen in the prior training data (*right*, 66.8%).

Conversational speech using the BCI

We developed a system for SP2 to have conversations via self-initiated speech. The BCI automatically detected when SP2 started or stopped speaking from neural activity, and decoded his attempted speech accordingly (Fig. 4d). Additionally, SP2 had the option to use an eye tracker for selecting actions (Fig. 4a) to i.) finalize and read aloud the sentence, ii.) indicate whether the sentence was decoded correctly or not, or iii.) spell out words letter-by-letter that were not correctly predicted by the decoder (e.g., because they were not in the vocabulary, such as certain proper nouns).

SP2's first use of the BCI for naturalistic communication with his family is exemplified in Fig. 4b (Table S3 provides additional transcripts). In subsequent sessions, SP2 utilized the neuroprosthesis for personal use (e.g., Video 3), communicating a total of 1189 sentences. For the majority of these sentences (925; 77.8%) we were able to confirm SP2's intended speech through directly asking SP2, contextual analysis, and examining the RNN-derived phoneme probability patterns. Self-initiated sentences for which we knew the ground-truth were decoded with a WER of 3.7% (Fig. 4c). For one session where we validated the ground truth of every sentence (43 sentences, 873 words) with SP2, the WER was 2.5%. Using the speech BCI, SP2 was empowered to tell 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 S3).

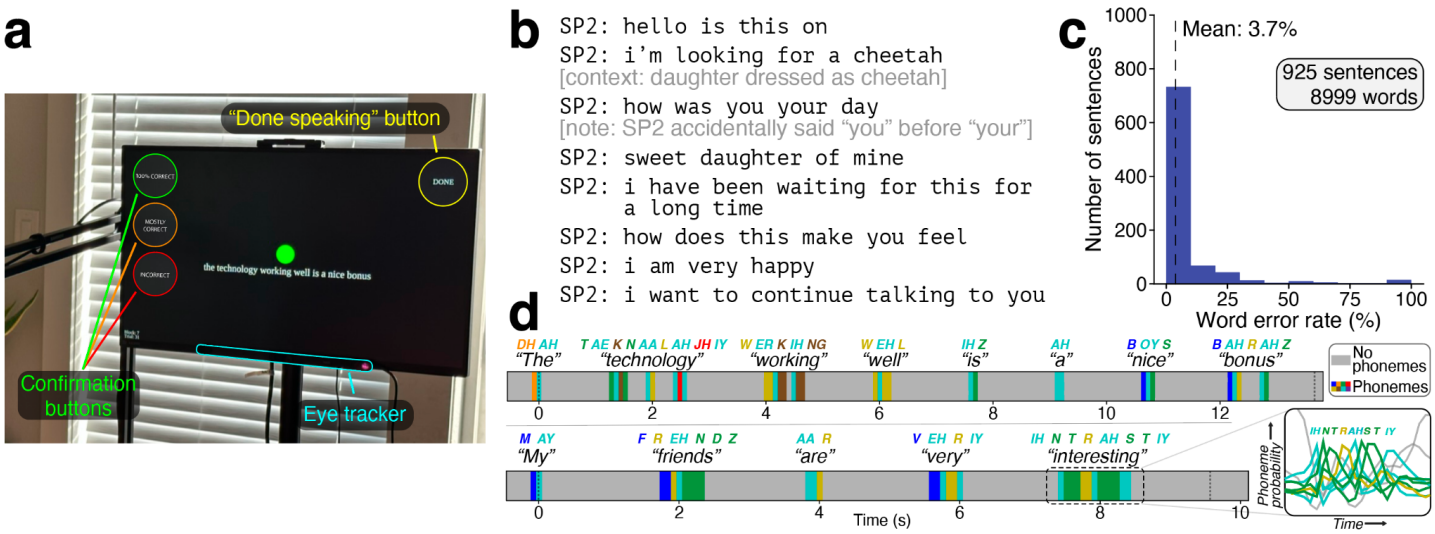

Figure 4. Decoding attempted speech during open conversations. a, Photograph of the participant's BCI interface during self-initiated speech. Sentence construction initiates when any phoneme's RNN output probability surpasses that of silence and concludes after 6 seconds of speech inactivity, or upon SP2's optional activation of an on-screen button via eye tracking. After the decoded sentence was finalized, SP2 used the on-screen confirmation buttons to indicate if the decoded sentence was correct. This photo has been cropped to not include the participant, as per medrXiv policy. **b**, Sample transcript of a conversation between SP2 and a family member, on the second day of use. **c**, Evaluating speech decoding accuracy in open conversations (n=925 sentences with known true labels). Average word error rate was 3.7% (95% CI: [3.3%, 4.3%]). **d**, Timeline of two example sentences showing the most probable phoneme at each time step, as indicated by RNN outputs. Gray intervals indicate the highest output probability is silence, while colored segments show the most probable phoneme. Phonemes are colored according to the phoneme category that they belong to (see Fig. 3e). Vertical dashed lines delineate the onset and termination of sentence construction. The decoded phonemes and words are annotated above each visualization. *Inset*, detailed view of selected phoneme probabilities as SP2 attempts to say the word "interesting".

Discussion:

Beginning on the first day of device use, a brain-to-text speech neuroprosthesis with 256 recording sites in the precentral gyrus accurately decoded intended speech in a man with severe dysarthria due to ALS. He communicated using a comprehensive 125,000 word vocabulary on the second day of use (and retrospective analysis indicated this could have been achieved on Day 1). Within 16 hours of use, the BCI correctly identified 97.3% of attempted words. To contextualize this 2.7% WER performance, the state-of-the-art for English automated speech recognition (e.g., smartphone dictation) has an approximate 5% WER[25] and able speakers have a 1-2% WER[26] when reading aloud. To our knowledge this is also the first study to report extensive open conversation via a large-vocabulary speech BCI, including decoding words never seen during training. We believe that the high decoding accuracy demonstrated in this study indicates that speech neuroprostheses have reached a level of performance suitable for rapidly and accurately restoring communication to people living with paralysis.

This study's participant used the brain-to-text speech BCI to converse with family, friends, healthcare professionals, and colleagues. His regular means of communication without a BCI involves either (1) having trained caregivers interpret his severely dysarthric speech, or (2) using a head-mouse with point-and-click selections on a computer screen. The (investigational) BrainGate Neural Interface System is now his preferred way to communicate with our research team, and he has requested the ability to use it on his own time to be able to more rapidly write and communicate as part of his occupation and family life. The own-voice text-to-speech at the end of each sentence is also a novel capability in a brain-to-text speech BCI; SP2 and his family reported being pleased that the system's voice resembled his own.

A clinically viable neuroprosthesis must not only be accurate, but should also minimize calibration time. This study demonstrated a large reduction in the quantity of training data required to achieve high accuracy decoding. In our previous study[20], the participant attempted to speak 260-480 sentences at the start of each day, after which up to ~30 minutes of computation time was required until the speech neuroprosthesis was ready for use. That previous study's reported closed-loop results were measured starting 113 days post-implant, and used more than 15 days of training data and 10,000 training sentences to achieve a WER of 23.8%, while a previous ECoG speech BCI required 17.7 hours of training data, collected over 13 days, to reach a WER of 25.5%[18]. This new neuroprosthesis provided over 99% accuracy on a limited set of 50 words[16]

after just 30 minutes of training data on the very first day of use. It also achieved over 95% accuracy on a large vocabulary after collecting 6.6 cumulative hours of training data (over 7 sessions), and offline analyses indicate that optimized methods could provide >91% accurate large-vocabulary communication on the first day of use. Rapid communication with an intracortical speech BCI builds on our previous demonstration of rapid point-and-click communication with first-time BCI users[27].

Previous studies have reported that intracortical devices often require recalibration due to signal nonstationarities[28–30]. Here, adopting the recent recognition that recent days' neural data can be used to calibrate an effective neural decoder for a new day[31,32], we demonstrated that a speech decoder could similarly be used to provide >95% accuracy at the start of each session. Future work is needed to establish whether the online fine-tuning we employed[23] can maintain performance indefinitely in the absence of ground-truth labels of intended speech.

We believe that a significant factor enabling the higher performance of this study relative to our prior intracortical speech BCI[20] was doubling the number of microelectrodes in speech motor cortex. Our finding that ~200 electrodes in these regions is sufficient for very high accuracy brain-to-text communication provides an important design parameter to guide ongoing efforts to build neural interface hardware that can reach patients at scale. Using an improved phoneme-to-sentences language model relative to our prior work[20] also improved performance (Fig. S5), and SP2's slow speaking rate (Fig. S2) may also have contributed.

In addition to recording from two arrays in the putative ventral portion of area 6v (speech motor cortex) as in [20], we also targeted one array each into two areas which, to our knowledge, have not previously been recorded from with multielectrode arrays: area 4 (primary motor cortex, which in humans is often in the sulcus[21] and thus largely not accessible with Utah arrays) and area 55b. We found that the strongest phoneme encoding was from the array in ventral 6v, which is consistent with our previous participant[20]. The array in area 4 also showed high phoneme encoding, as did the array in area 55b, which has recently been proposed as an important node in the wider speech production network[33]. We note that these brain area descriptions are estimations based on precisely aligning SP2's brain to a Human Connectome Project derived atlas using multi-modal MRI.

Limitations:

As with other recent clinical trial reports in the nascent field of implanted speech BCIs[16–18,20,34], this study involved a single participant. Future work with additional participants is needed to establish the across-individual distribution of performances for speech BCIs using similar methods. Whether similar results can be expected may depend on whether the signal-to-noise ratio of SP2's speech-related neural signals is typical. Nevertheless, these data, when combined with our previous speech BCI results with two 64-electrode arrays in area 6v[20], demonstrate both successful initial replication and subsequent methodological improvements of the intracortical speech BCI approach. It is also not yet known how the performance of the system may change over the long term, but previous studies decoding attempted arm and hand movements using Utah arrays[35] sustained high accuracy for multiple years after implantation[36–39].

The participants in both this study and [20] had dysarthria due to ALS. Further work will assess whether similar methods will work for other etiologies of dysarthria. Given that we recorded from ventral precentral gyrus, which is upstream of the neuronal injury incurred in many conditions, and that recent ECoG speech neuroprostheses were demonstrated in two individuals with brainstem stroke[16–18], we predict that this approach will also work in other conditions[35].

While the demonstrated brain-to-text capabilities can provide widely useful communication, they do not capture the full expressive richness of voice; the more difficult challenge of closed-loop brain-to-voice synthesis remains an active area of speech BCI research[18,34,40].

Audio 1 - Demonstration of SP2's unintelligible dysarthric speech. SP2 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.

Link to listen online: https://ucdavis.box.com/s/gegiqcl4jzqdnug6dwxjmnd5t7gams4h

Video 1 - Copy Task speech decoding with eye tracker control. This video shows the same speech decoding trials as in Audio 1 (session 10). Prompted sentences appear on the screen in front of SP2. When the red square turns green, SP2 attempts to say the prompted sentence aloud while the speech decoder predicts what he is saying in real time. In this video, SP2 is signaling the end of a sentence by using an eye tracker to hit an on-screen "done" button. 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.

Link to view online: https://ucdavis.box.com/s/0ono8rwx1evmp8ee27po8bs0u45bcvjn

Video 2 - Copy Task decoding with neural click control. This video shows another example of Copy Task speech decoding from a later session (session 17). Prompted sentences appear on the screen in front of SP2. When the red square turns green, SP2 attempts to say the prompted sentence aloud while the speech decoder predicts what he is saying in real time. In this video, SP2 is signaling the end of a sentence by attempting to squeeze his right fist, the neural correlates of which are decoded (Section S6). 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.

Link to view online: https://ucdavis.box.com/s/afeqbmljk81rt4yscr2uh71ksn563g6y

Video 3 - Self-initiated conversational speech decoding. SP2 is using the speech decoder to engage in freeform conversation with those around him. The video is muted while conversation partners are speaking for privacy reasons. The BCI reliably detects when SP2 begins attempting to speak, and shows the decoded words on-screen in real time. SP2 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 BCI 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, SP2 uses the eye tracker to confirm whether the decoded sentence was correct or not. Correctly decoded sentences are used to fine-tune the neural decoder online.

Link to view online: https://ucdavis.box.com/s/79nyhal9q6x7kc4toq63jkcfezdtditq

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.
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. 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.
5. Luo S, Rabbani Q, Crone NE. Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication. Neurotherapeutics 2022;19(1):263–73.
6. 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
7. 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.
8. 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.
9. 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.
10. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. Nature 2019;568(7753):493–8.
11. 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.
12. 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.

13. 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.
14. 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.
15. Guenther FH, Brumberg JS, Wright EJ, et al. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. PLOS ONE 2009;4(12):e8218.
16. 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.
17. 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.
18. Metzger SL, Littlejohn KT, Silva AB, et al. A high-performance neuroprosthesis for speech decoding and avatar control. Nature 2023;620(7976):1037–46.
19. 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.
20. Willett FR, Kunz EM, Fan C, et al. A high-performance speech neuroprosthesis. Nature 2023;620(7976):1031–6.
21. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. Nature 2016;536(7615):171–8.
22. Ali YH, Bodkin K, Rigotti-Thompson M, et al. BRAND: A platform for closed-loop experiments with deep network models [Internet]. 2023 [cited 2023 Dec 11];2023.08.08.552473. Available from: https://www.biorxiv.org/content/10.1101/2023.08.08.552473v1
23. 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 [Internet]. 2023 [cited 2023 Dec 11];Available from: http://arxiv.org/abs/2311.03611
24. 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

25. 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
26. 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
27. Brandman DM, Hosman T, Saab J, et al. Rapid calibration of an intracortical brain–computer interface for people with tetraplegia. J Neural Eng 2018;15(2):026007.
28. Perge JA, Homer ML, Malik WQ, et al. Intra-day signal instabilities affect decoding performance in an intracortical neural interface system. J Neural Eng 2013;10(3):036004.
29. Jarosiewicz B, Sarma AA, Bacher D, et al. Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. Sci Transl Med 2015;7(313):313ra179-313ra179.
30. Downey JE, Schwed N, Chase SM, Schwartz AB, Collinger JL. Intracortical recording stability in human brain-computer interface users. J Neural Eng 2018;15(4):046016.
31. Sussillo D, Stavisky SD, Kao JC, Ryu SI, Shenoy KV. Making brain–machine interfaces robust to future neural variability. Nat Commun 2016;7(1):13749.
32. Hosman T, Pun TK, Kapitonava A, Simeral JD, Hochberg LR. Months-long High-performance Fixed LSTM Decoder for Cursor Control in Human Intracortical Brain-computer Interfaces [Internet]. In: 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER). Baltimore, MD, USA: IEEE; 2023 [cited 2023 Dec 11]. p. 1–5.Available from: https://ieeexplore.ieee.org/document/10123740/
33. 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.
34. Angrick M, Luo S, Rabbani Q, et al. Online speech synthesis using a chronically implanted brain-computer interface in an individual with ALS. medRxiv 2023;2023.06.30.23291352.
35. Rubin DB, Ajiboye AB, Barefoot L, et al. Interim Safety Profile From the Feasibility Study of the BrainGate Neural Interface System. Neurology 2023;100(11):e1177–92.
36. Willett FR, Avansino DT, Hochberg LR, Henderson JM, Shenoy KV. High-performance brain-to-text communication via handwriting. Nature 2021;593(7858):249–54.

37. Wodlinger B, Downey JE, Tyler-Kabara EC, Schwartz AB, Boninger ML, Collinger JL. Ten-dimensional anthropomorphic arm control in a human brain−machine interface: difficulties, solutions, and limitations. J Neural Eng 2015;12(1):016011.
38. Bacher D, Jarosiewicz B, Masse NY, et al. Neural Point-and-Click Communication by a Person With Incomplete Locked-In Syndrome. Neurorehabil Neural Repair 2015;29(5):462–71.
39. Flesher SN, Downey JE, Weiss JM, et al. A brain-computer interface that evokes tactile sensations improves robotic arm control. Science 2021;372(6544):831–6.
40. Wairagkar M, Hochberg LR, Brandman DM, Stavisky SD. Synthesizing Speech by Decoding Intracortical Neural Activity from Dorsal Motor Cortex [Internet]. In: 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER). Baltimore, MD, USA: IEEE; 2023 [cited 2023 Dec 11]. p. 1–4.Available from: https://ieeexplore.ieee.org/document/10123880/