Decoding Speech from Intracortical Multielectrode Arrays in Dorsal "Arm/Hand Areas" of Human Motor Cortex

Sergey D. Stavisky, *IEEE Member,* Paymon Rezaii, Francis R. Willett, Leigh R. Hochberg, Krishna V. Shenoy*, *IEEE Senior Member,* and Jaimie M. Henderson*

Abstract— **Neural prostheses are being developed to restore speech to people with neurological injury or disease. A key design consideration is where and how to access neural correlates of intended speech. Most prior work has examined cortical field potentials at a coarse resolution using electroencephalography (EEG) or medium resolution using electrocorticography (ECoG). The few studies of speech with single-neuron resolution recorded from ventral areas known to be part of the speech network. Here, we recorded from two 96-electrode arrays chronically implanted into the 'hand knob' area of motor cortex while a person with tetraplegia spoke. Despite being located in an area previously demonstrated to modulate during attempted arm movements, many electrodes' neuronal firing rates responded to speech production. In offline analyses, we could classify which of 9 phonemes (plus silence) was spoken with 81% single-trial accuracy using a combination of spike rate and local field potential (LFP) power. This suggests that high-fidelity speech prostheses may be possible using large-scale intracortical recordings in motor cortical areas involved in controlling speech articulators.**

I. Introduction

Regaining the ability to communicate is a critical need for people who cannot speak due to neurological injury (such as stroke) or disease (such as ALS). Brain-computer interfaces (BCIs) can potentially restore speech by converting neural signals related to what a person wants to say into synthesized speech or digital text [1]. A number of speech BCI strategies have been proposed, including decoding covert (silent) speech [2]–[4] or more abstract language signals [5]. Here, we explored decoding the neural correlates of attempted speech production [6]–[10], which has several potential advantages. First, a decoder training protocol in which the user attempts to speak a specific instructed sound may provide a naturalistic control scheme, even in mute patient populations; an analogous approach has allowed successful calibration of BCIs based on intended arm movements in people with paralysis [11], [12]. Second, decoding *attempted* speaking may be less likely to inadvertently communicate the user's private internal monologue. Last, studies with participants capable of overtly speaking (such as this study) provide neural data with accompanying ground truth acoustic data. This facilitates investigating the neural dynamics underlying speech production [13] and developing strategies for decoding this activity.

What kinds of neural signals are useful for a speech BCI, and from where can they be recorded? Most studies to date decoded EEG or ECoG signals, typically from face sensorimotor cortex, Broca's area, and superior temporal gyrus [1], [14]. In comparison, intracortical recordings during speech production remain relatively unexplored, although a handful of studies have examined activity in temporal gyrus [15]–[17], ventral precentral gyrus [7], [18], orbitofrontal cortex [16], and Broca's area [19]. Inspired by recent arm movement BCI research, where the highest-performing systems use signals from multielectrode arrays implanted in motor cortex [12], [20], we wondered whether speech-related neural activity could also be decoded using this sensor type. As a first step towards establishing feasibility, we recorded neural signals during overt speaking by a human participant in the BrainGate2 clinical trial. This person was previously implanted with arrays in dorsal areas of motor cortex known to be involved in arm and hand control [12]. To the best of our knowledge, action potential (spiking) signals in this area have not previously been examined during speaking, and there have not been reports of speech decoding from motor cortex using high channel count intracortical sensors. While our approach is limited by array placement that is probably suboptimal for detecting speech-related activity, it allows an initial assessment without incurring the additional risk of implanting arrays in what is traditionally thought of as eloquent cortex.

This work was supported by: ALS Association Milton Safenowitz Postdoctoral Fellowship; Stanford Institute for Neuro-Innovation and Translational Neuroscience; Larry and Pamela Garlick; Samuel and Betsy Reeves; NIHNIDCD R01DC014034; NIH-NINDS R01NS066311; NIH-NIDCD R01DC009899; Rehabilitation Research and Development Service, Department of Veterans Affairs (B6453R and N9288C); the Executive Committee on Research (ECOR) of Massachusetts General Hospital; and MGH-Deane Institute for Integrated Research on Atrial Fibrillation and Stroke. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the Department of Veterans Affairs, or the U.S. Government. CAUTION: Investigational Device. Limited by Federal Law to Investigational Use.

SDS and FRW are with the Departments of Neurosurgery and Electrical Engineering, Stanford University, Stanford, CA. PR and JMH are with the Department of Neurosurgery, Stanford. JMH is also with the Stanford Neuroscience Institute. LRH is with Center for Neurorestoration and Neurotechnology, Rehabilitation R&D Service, VA Medical Center, Providence, RI; the School of Engineering and the Institute for Brain Science, Brown University, Providence, the Center for Neurotechnology and Neurorecovery, Department of Neurology, Massachusetts General Hospital, Boston, MA, and the Department of Neurology, Harvard Medical School, Boston. KVS is with the Departments of Electrical Engineering, Bioengineering and Neurobiology, the Neurosciences and Bio-X Programs, the Stanford Neuroscience Institute, and the Howard Hughes Medical Institute at Stanford. Financial interests: KVS: Neuralink Corp. (Consultant), CTRL-Labs Inc., MIND-X Inc., Inscopix Inc., and Heal Inc. (Scientific Advisory Boards). * KVS and JHM contributed equally to this work. (correspondence e-mail: sergey.stavisky@stanford.edu).

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

We found considerable speech-related firing rate changes and were able to decode spoken phonemes with high accuracy using both intracortical multiunit spikes and LFP.

II. Methods

A. Participant and Neural Sensors

Participant 'T5' gave informed consent and voluntarily enrolled in the BrainGate2 intracortical neural prosthesis clinical trial (http://www.clinicaltrials.gov/ct2/show/NCT00912041). This study was approved under an Investigational Device Exemption by the US Food and Drug Administration and by the Institutional Review Boards of Stanford University (20804) and Partners Healthcare / Massachusetts General Hospital (2011P001036). T5 is a right-handed man, 64 years old at the time of the study, diagnosed with C4 AIS-C spinal cord injury ten years prior to these research sessions. He is able to speak normally and converses naturally without hearing assistance, but has some trouble hearing from his left ear. Two 96-electrode Utah arrays (1.5 mm electrode length, Blackrock Microsystems) were surgically implanted in T5's left dorsal motor cortex (Fig. 1B) fourteen months prior to the present study. The arrays were targeted to arm-related motor cortical areas because two goals of the BrainGate2 trial are demonstrating BCI communication using point-and-click keyboards and restoration of reach and grasp function via control of a robotic arm or functional electrical stimulation.

B. Verbally Prompted Speaking Task

The participant performed a task in which he spoke ten different phonemes (vowels, consonants, and consonant-vowel syllables), one per trial, thirty times each. During each trial he was prompted as to which phoneme to speak by an audio recording of that phoneme being played to him through computer speakers positioned facing his right ear. Each trial consisted of two beeps to alert the participant that the trial was starting, followed by the phoneme prompt. Then, two clicks served as the 'go cue' that instructed the participant to speak the prompted phoneme. The experimental setup and the phases of each trial are diagrammed in Fig. 1A. As a baseline for comparison, there was an eleventh 'silence' condition in which the trial structure was identical, except that the speakers played a nearly-silent recording of the ambient background noise during recording of the phoneme prompts. The participant was instructed not to say anything in response to this silence prompt. He was also asked to avoid trying to move or speak during a trial, except when prompted to speak, and to keep his gaze fixated on the same object in the room in front of him.

The task was performed in three blocks, with each block consisting of ten sets of eleven trials (one trial of each condition presented in a random order). There was a brief pause between sets (~10 seconds) which ended when the participant indicated that he was ready to continue. A longer pause of several minutes between blocks encouraged the participant to rest, adjust his posture, drink water, etc. We excluded infrequent trials in which the participant misspoke

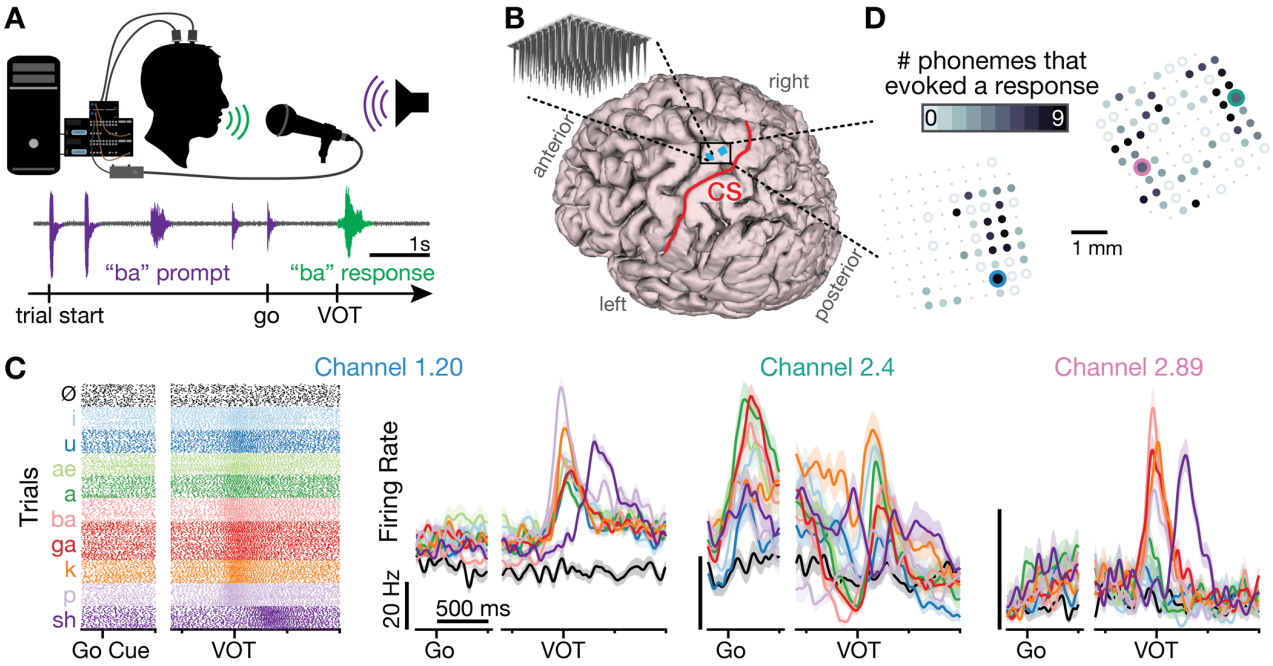

Figure 1. Spiking activity in dorsal motor cortex during speech production. (A) Schematic of the experimental setup (top) and task structure (bottom). The participant heard an auditory prompt consisting of two beeps followed by a phoneme. After a brief delay, a go cue consisting of two clicks instructed him to repeat back the phoneme. Neural data were recorded throughout the task using a NeuroPort system. Acoustic data were recorded and digitized alongside the neural signals. The acoustic amplitude during one example trial is shown above the trial timeline. VOT is voice onset time. (B) Imaging-derived 3D model of the participant's brain showing the locations of two 96-electrode arrays implanted in motor cortex. CS is central sulcus. (C) (Left) raster plot shows the times of neuronal action potentials (tick marks) for one example electrode during all trials. Each row denotes a trial, which are grouped by which phoneme the participant spoke. Ø is silence condition. Data are aligned to both the go cue and the VOT, which was on average 1210 ms after the go cue. (Right) traces show trial-averaged firing rates (mean ± 1 s.e.) for the same electrode as well as two other electrodes. (D) Map of electrode locations showing which electrodes' firing rates changed significantly ($p < 0.001$) during speaking, compared to the silence condition. Shading denotes how many of the nine different phonemes evoked a response on a given electrode; the marker corresponding to zero phonemes is hollow. The three example electrodes from panel C are marked with circles of the corresponding color. Electrodes that did not record spikes are shown as smaller dots.

or mistimed his response, or when there were unexpected sounds in his environment. The participant frequently misheard *da* as *ga* or *ba*; we labeled those trials based on his response, rather than the prompt, and excluded the remaining *da* trials from further analysis due to low trial count. The dataset consists of one experimental session (day) during which the number of accepted trials for each condition were: 30 each of silence, *i*, *a*, *k*, *p*, and *sh*; 28 ae; 31 *ba*; and 50 *ga*.

Voice onset time (VOT) was manually annotated for each trial by visual and auditory inspection of the recorded audio data. To allow for a consistent application of VOT-aligned data analysis and visualization, a faux VOT was marked for each silence trial; this silence VOT was matched to the mean time between the go cue and VOT for all the spoken phonemes during the same block.

C. Neural Signal Processing

During the experiment session, wired cables were connected to the participant's cranial pedestal connectors. Voltages recorded by the NeuroPort system (Blackrock Microsystems) were analog band-pass filtered between 0.3 Hz and 7.5 kHz, and then digitized at 30 kHz. Seventeen of the electrodes were removed from further analysis because they were disconnected or exhibited large noise artifacts. The signals from each of the remaining 175 electrodes were then split into spikes and LFP features. The spikes signal path consisted of re-referencing within each array using common average referencing followed by digitally high-pass filtering the voltage (400 Hz cutoff frequency, zero-phase by forward and reverse DF-II filtering). For each electrode, a spike was detected at every millisecond if the voltage dropped below an electrode-specific threshold. The resulting threshold crossing spike trains represent action potentials from one or several neurons in the immediate vicinity of the electrode; we did not perform spike sorting (i.e., attempt to attribute threshold crossing waveforms to putative single neurons) for this BCI-focused study, as per other recent iBCI studies [12]. For examining the spikes data (III.A, Fig. 1), thresholds were set at -4.5 times each electrode's rms voltage value. We excluded 70 electrodes that recorded fewer than one spike/second and five with cross-talk (all but one electrode in subsets with correlations > 0.5 with one another were discarded). For the classification analyses (III.B, Fig. 2), we relaxed the threshold to -3.5 × rms because we found this slightly increased classification accuracy due to inclusion of more multiunit spikes which contained useful information. Spike series were smoothed with a 25 ms s.d. Gaussian to generate the peristimulus time histograms shown in Fig. 1C.

Seven different LFP features were derived from the voltage signals. The local motor potential (LMP) was obtained by first re-referencing within each array, and then smoothing the LFP with a 50 ms boxcar filter. This low frequency signal is believed to strongly reflect local synaptic input [21], and we previously found that it is highly informative for decoding arm movement kinematics [22] and errors [23]. Six additional LFP features were derived by taking the power of each signal after band-pass filtering (3rd order Butterworth) 10 to 25 Hz, 25 to 40 Hz, 40 to 65 Hz, 65 to 125 Hz, 125 to 250 Hz, or 125 to 5,000 Hz. The latter two "high gamma" bands are believed to capture a considerable amount of action potential waveform power [24].

D. Classification of Spoken Phonemes from Neural Signals

We performed a classification analysis to estimate how much information about spoken phonemes was present in the recorded neural signals. Each trial's neural data were processed into a vector of neural features. The trials were divided into a training set used to fit the classifier parameters, and a (separate) testing set used to measure how well the classifier could use unfamiliar neural data to predict which phoneme (or silence) was being spoken. More specifically, neural features (e.g., spike rate or LFP power) were binned in ten non-overlapping 100 ms bins spanning from 500 ms before to 500 ms after VOT. Multiple features could be derived from each electrode, and all of the time bins' features were concatenated as a single feature vector. Thus, the classifier used in Fig. 2C operates on a 175 electrodes × 2 features per electrode × 10 time bins = 3500-dimensional feature vector. The class label associated with each trial's neural data vector was the phoneme being spoken.

We used the off-the-shelf multiclass support vector machine (SVM) classifier implementation in MATLAB (Mathworks Inc.). Specifically, the model was fit using the multiclass model object (*fitcecoc*) using the SVM template (*templateSVM*) with key parameters Standardize = true, KernelFunction = 'linear', 'OutlierFraction' = 0.05. For feature comparisons, the test set consisted of one trial chosen at random from each label, while all the other trials were used as training data. This procedure was repeated 1,001 times to provide an estimate of the variability of the performance across difference samplings of test/train data. Once the best feature(s) were identified, we measured the classifier's performance in a leave-one-trial-out manner where each trial's phoneme was predicted using a classifier trained using every other trial. This analysis maximized the available training data, and its result is shown in Fig. 2B. The distribution of chance-level prediction accuracy was measured by randomly permuting the class labels of all training and test trials and repeating the train-and-test procedure 101 times.

III. Results

A. Neural Responses to Phoneme Production

We recorded neural activity in the hand knob area of motor cortex using two chronically-implanted 96-electrode arrays while a person with paralysis spoke individual phonemes in response to an audio prompt. Firing rates changed on many electrodes leading up to, and following, voice onset. Fig. 1C shows the response patterns of three example electrodes, which were chosen to exemplify the diversity of strong responses. Across both arrays, 71 out of 100 electrodes showed a significant change in firing rate depending on which phoneme, or silence, was spoken ($p < 0.001$, ANOVA comparing individual trials' firing rate from 500 ms before to 500 ms after VOT). Excluding silence, 65 electrodes significantly changed firing rates across spoken phonemes. Tuning to different phonemes was broad in the sense that the majority of electrodes which responded during *any* phoneme responded during *multiple* phonemes (Fig. 1D).

B. Phoneme Classification Performance

To help evaluate the feasibility of a speech BCI that decodes these intracortical signals, we tested how well we

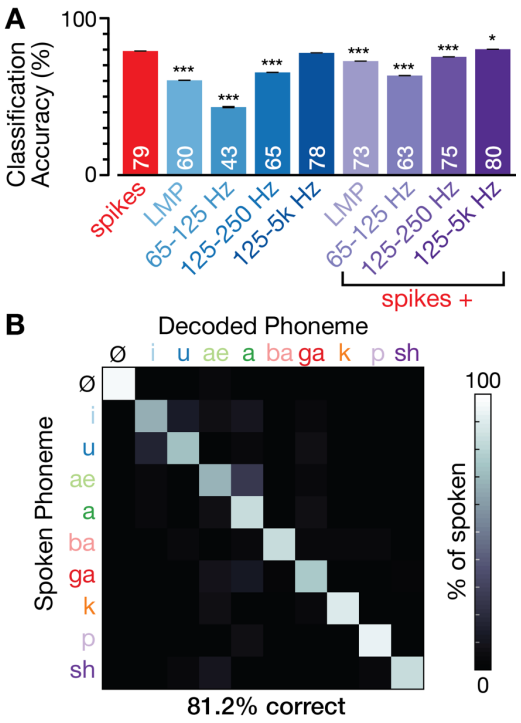

Figure 2. Predicting spoken speech from neural data. (A) Comparison of cross-validated classification accuracies when predicting the spoken phoneme using threshold crossing spikes (red), various LFP-derived features (blue), and a concatenation of spikes and LFP features (purple). Each bar shows mean ± 1 s.e. across 1,001 resampled train/test data folds. All features' accuracies were greater than chance ($p < 0.01$, shuffle test). Stars denote significant difference from the spikes-only classifier (rank-sum test, *$p<0.05$, ***$p<0.001$). (B) Confusion matrix for the combined spikes and 125-5k Hz LFP power decoder. Note that the overall performance differs slightly from panel A because these results are using leave-one-trial-out cross-validation to maximize available training data.

could predict which phoneme (or silence) was being spoken. Multi-class SVM classifiers were trained on different sets of neural features derived from each electrode; they were evaluated on held-out test data. Fig. 2A compares classification performance using various spike and LFP features alone and in combination. Lower-frequency power bands (10–25 Hz, 25–40 Hz, 40 –65 Hz) were also tested but were found to perform poorly (< 30% accuracy), and are not plotted. Threshold crossing spikes and very high frequency LFP power (125 – 5k Hz) yielded the highest accuracy amongst individual features, and a classifier combining both spikes and 125 – 5k power from each electrode resulted in a slight improvement over either feature alone. Fig. 2B shows the confusion matrix for classification using this hybrid decoder. Overall single-trial classification accuracy was 81.2%, compared to 11.1% mean chance performance across 101 label shuffles (range: 6.9% to 17.2%). The classification accuracy of each individual phoneme was higher than all of that phoneme's shuffled classification accuracies.

IV. Discussion

A. *Speech-Related Activity in Dorsal Motor Cortex*

A surprising finding of this study is substantial speech-related firing rate modulation in dorsal motor cortex. This area was not previously implicated in generating speech based on ECoG studies such as [13] and electrical stimulation mapping [25], [26]. One possible explanation for this difference is that dorsal motor cortical activity is not prominently evident in ECoG measurements due to spatial averaging [17], but can be detected in the spikes and LFP signals from penetrating electrodes. This activity could reflect functional connections between hand motor cortex and the language network [27], [28]. An alternative explanation is that the motor map of our participant is atypical due to his tetraplegia. This question will be difficult to resolve soon because there are currently few clinical reasons to record from within dorsal motor cortex of humans who are neither paralyzed, nor have intractable epilepsy (both of which can cause cortical remapping).

We have now reported that the same motor cortical population, in the same participant, is modulated by attempted arm movements [12], movement task outcome error [23], and speech production. A recent report by a different group recording in similar areas (also from a person with paralysis) during word reading attributed activity to the sensorimotor semantic grounding of the words [29]. We hypothesize that the activity we observed during phoneme production is related to cortical control of the speech articulators (e.g., tongue and jaw), but further experiments are needed to better understand the role of this activity.

B. *Implications for Speech BCIs*

The single-trial classification accuracies reported here compare well to previous state-of-the-art results [9], [10], which report <80% mean classification accuracies when decoding amongst four or five spoken sounds ([7], [9] also decoded amongst a much larger number of phonemes with 21- 36% accuracy, which we cannot make direct comparisons to). We attribute this to the high information content in the multiunit spiking activity recorded on many electrodes. We do not, however, wish to suggest that dorsal motor cortex is a preferred target for speech BCIs. Rather, we view these results — obtained from what are most likely suboptimal array locations — as an encouraging indication that speech decoding can be improved if large numbers of electrodes are implanted in cortical areas known to have strong speech-related activity [13], [26], [30]–[32]. These results argue against prior speculation, based on the broad distribution of speech representations observed using ECoG, that the limited coverage of Utah arrays makes them ill-suited for studying speech [33]. Rather, our observation of complex spatiotemporal tuning for multiple phonemes within each individual Utah array extends previous micro-ECoG reports of fine-scale spatial differences between phonemes [34].

Here we classified isolated phonemes offline. A practical speech BCI will need to operate with low latency in closed-loop [18], [34], either by classifying sounds within continuous speech [1], [35], or by directly decoding continuous variables such as formants [8], [18]. Decoding fluid speech will be more difficult due to coarticulation-related neural variability [8], but it will benefit from recent advances in automated speech recognition [1], [36]; language models can provide strong priors on what the observed neural activity means based on the recent history of decoded speech. Future work should also examine whether alternative representational basis sets, such as movements of the articulators [35] or formant frequencies [8], [18], correlate more closely with intracortical activity than phonemes.

ACKNOWLEDGMENT

We thank T5 and his caregivers for their dedicated contributions to this research, J. Ventura and N. Lam for administrative support, and Dr. Sydney Cash, a BrainGate2 consortium member, for initial scientific discussions.

REFERENCES

[1] C. Herff and T. Schultz, "Automatic speech recognition from neural signals: A focused review," *Front. Neurosci.*, vol. 10, no. 429, pp. 1–7, 2016.

[2] P. Suppes, Z. L. Lu, and B. Han, "Brain wave recognition of words.," Proc. Natl. Acad. Sci. U. S. A., vol. 94, no. 26, pp. 14965–9, 1997.

[3] S. Martin, P. Brunner, C. Holdgraf, H.-J. Heinze, N. E. Crone, J. Rieger, G. Schalk, R. T. Knight, and B. N. Pasley, "Decoding spectrotemporal features of overt and covert speech from the human cortex," Front. Neuroeng., vol. 7, no. May, pp. 1–15, May 2014.

[4] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features," J. Neural Eng., vol. 15, no. 1, p. 16002, Feb. 2017.

[5] A. M. Chan, E. Halgren, K. Marinkovic, and S. S. Cash, "Decoding word and category-specific spatiotemporal representations from MEG and EEG," Neuroimage, vol. 54, no. 4, pp. 3028–3039, 2011.

[6] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface.," J. Neural Eng., vol. 7, no. 5, p. 56007, Oct. 2010.

[7] J. S. Brumberg, E. J. Wright, D. S. Andreasen, F. H. Guenther, and P. R. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex," Front. Neurosci., vol. 5, pp. 1–12, 2011.

[8] K. E. Bouchard and E. F. Chang, "Control of Spoken Vowel Acoustics and the Influence of Phonetic Context in Human Speech Sensorimotor Cortex," J. Neurosci., vol. 34, no. 38, pp. 12662–12677, 2014.

[9] E. M. Mugler, J. L. Patton, R. D. Flint, Z. A. Wright, S. U. Schuele, J. Rosenow, J. J. Shih, D. J. Krusienski, and M. W. Slutzky, "Direct classification of all American English phonemes using signals from functional speech motor cortex," J. Neural Eng., vol. 11, no. 3, 2014.

[10] N. F. Ramsey, E. Salari, E. J. Aarnoutse, M. J. Vansteensel, M. G. Bleichner, and Z. V. Freudenburg, "Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids," Neuroimage, 2017.

[11] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia.," Nature, vol. 442, no. 7099, pp. 164–71, Jul. 2006.

[12] C. Pandarinath, P. Nuyujukian, C. H. Blabe, B. L. Sorice, J. Saab, F. R. Willett, L. R. Hochberg, K. V. Shenoy, and J. M. Henderson, "High performance communication by people with paralysis using an intracortical brain-computer interface," Elife, vol. 6, Feb. 2017.

[13] K. E. Bouchard, N. Mesgarani, K. Johnson, and E. F. Chang, "Functional organization of human sensorimotor cortex for speech articulation," Nature, vol. 495, no. 7441, pp. 327–332, 2013.

[14] S. Chakrabarti, H. M. Sandberg, J. S. Brumberg, and D. J. Krusienski, "Progress in speech decoding from the electrocorticogram," Biomed. Eng. Lett., vol. 5, no. 1, pp. 10–21, 2015.

[15] O. Creutzfeldt, G. A. Ojemann, and E. Lettich, "Neuronal activity in the human lateral temporal lobe: I. Responses to Speech," Exp. Brain Res., vol. 77, pp. 451–475, 1989.

[16] A. Tankus, I. Fried, and S. Shoham, "Structured neuronal encoding and decoding of human speech features," Nat. Commun., vol. 3, p. 1015, 2012.

[17] A. M. Chan, A. R. Dykstra, V. Jayaram, M. K. Leonard, K. E. Travis, B. Gygi, J. M. Baker, E. Eskandar, L. R. Hochberg, E. Halgren, and S. S. Cash, "Speech-Specific Tuning of Neurons in Human Superior Temporal Gyrus," Cereb. Cortex, vol. 24, no. 10, pp. 2679–2693, Oct. 2014.

[18] F. H. Guenther, J. S. Brumberg, E. Joseph Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, D. S. Andreasen, P. Ehirim, H. Mao, and P. R. Kennedy, "A wireless brain-machine interface for real-time speech synthesis," PLoS One, vol. 4, no. 12, 2009.

[19] N. T. Sahin, S. Pinker, S. S. Cash, D. Schomer, and E. Halgren, "Sequential processing of lexical, grammatical, and phonological information within Broca's area," Science, vol. 326, no. 5951, pp. 445–449, 2009.

[20] Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. C. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, "High-performance neuroprosthetic control by an individual with tetraplegia.," Lancet, vol. 381, no. 9866, pp. 557–64, Feb. 2013.

[21] G. T. Einevoll, C. Kayser, N. K. Logothetis, and S. Panzeri, "Modelling and analysis of local field potentials for studying the function of cortical circuits," Nat. Rev. Neurosci., vol. 14, no. 11, pp. 770–785, Oct. 2013.

[22] S. D. Stavisky, J. C. Kao, P. Nuyujukian, S. I. Ryu, and K. V Shenoy, "A high performing brain–machine interface driven by low-frequency local field potentials alone and together with spikes," J. Neural Eng., vol. 12, no. 3, p. 36009, 2015.

[23] N. Even-Chen, S. D. Stavisky, C. Pandarinath, P. Nuyujukian, C. H. Blabe, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, "Feasibility of Automatic Error Detect-and-undo system in Human Intracortical Brain-Computer Interfaces," *IEEE Trans. Biomed. Eng.*, 2017.

[24] S. Waldert, R. N. Lemon, and A. Kraskov, "Influence of spiking activity on cortical local field potentials," *J. Physiol.*, vol. 591, no. 21, pp. 5291–5303, Nov. 2013.

[25] J. D. Breshears, A. M. Molinaro, and E. F. Chang, "A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation," *J. Neurosurg.*, vol. 123, no. 2, pp. 340–349, Aug. 2015.

[26] H. Duffau, L. Capelle, D. Denvil, P. Gatignol, N. Sichez, M. Lopes, J. P. Sichez, and R. Van Effenterre, "The role of dominant premotor cortex in language: A study using intraoperative functional mapping in awake patients," *Neuroimage*, vol. 20, no. 4, pp. 1903–1914, 2003.

[27] I. G. Meister, B. Boroojerdi, H. Foltys, R. Sparing, W. Huber, and R. Töpper, "Motor cortex hand area and speech: Implications for the development of language," *Neuropsychologia*, vol. 41, no. 4, pp. 401–406, 2003.

[28] Vainio, M. Tiainen, K. Tiippana, and M. Vainio, "Shared processing of planning articulatory gestures and grasping," *Exp. Brain Res.*, vol. 232, no. 7, pp. 2359–2368, 2014.

[29] Y. Yang, M. W. Dickey, J. Fiez, B. Murphy, T. Mitchell, J. Collinger, E. Tyler-Kabara, M. Boninger, and W. Wang, "Sensorimotor experience and verb-category mapping in human sensory, motor and parietal neurons," *Cortex*, vol. 92, pp. 304–319, 2017.

[30] V. L. Towle, H. A. Yoon, M. Castelle, J. C. Edgar, N. M. Biassou, D. M. Frim, J. P. Spire, and M. H. Kohrman, "ECoG gamma activity during a language task: Differentiating expressive and receptive speech areas," *Brain*, vol. 131, no. 8, pp. 2013–2027, 2008.

[31] K. Simonyan, H. Ackermann, E. F. Chang, and J. D. Greenlee, "Symposium New Developments in Understanding the Complexity of Human Speech Production," *J. Neurosci.*, vol. 36, no. 45, pp. 11440–11448, 2016.

[32] C. Cheung, L. S. Hamiton, K. Johnson, and E. F. Chang, "The auditory representation of speech sounds in human motor cortex," *Elife*, vol. 5, p. e12577, 2016.

[33] K. E. Bouchard and E. F. Chang, "Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography.," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 6782–5.

[34] E. C. Leuthardt, C. Gaona, M. Sharma, N. Szrama, J. Roland, Z. Freudenberg, J. Solis, J. Breshears, and G. Schalk, "Using the electrocorticographic speech network to control a brain-computer interface in humans," *J. Neural Eng.*, vol. 8, no. 3, 2011.

[35] F. Lotte, J. S. Brumberg, P. Brunner, A. Gunduz, A. L. Ritaccio, C. Guan, and G. Schalk, "Electrocorticographic representations of segmental features in continuous speech," *Front. Hum. Neurosci.*, vol. 9, pp. 1–13, 2015.

[36] R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, and Y. Wu, "Exploring the Limits of Language Modelieng," *Int. Conf. Mach. Learn.*, 2016.