Motor cortical dynamics during speech

Sergey D. Stavisky†, Francis R. Willett, Paymon Rezaii, Leigh R. Hochberg, Krishna V. Shenoy, Jaimie M. Henderson**

†sergey.stavisky@stanford.edu   *equal contribution

Speaking requires coordinating articulator muscles with exquisite timing and precision. Understanding how the sensorimotor system accomplishes this requires studying its neural underpinnings. Neural measurements are also critical for identifying the causes of speech disorders and building brain-computer interfaces (BCIs) to restore speech. Speech is a uniquely human behavior, which makes electrophysiological investigations challenging. Previous direct neural recordings during speech have come from electrocorticography (ECoG)[1] or single-unit (SUA) recordings from penetrating electrodes[2] during clinical epilepsy treatment. Such studies have begun to characterize motor cortical dynamics underlying speech[3], but not at the finer spatiotemporal scale uniquely afforded by high-density intracortical recordings often used in animal reaching studies[4].

We had the opportunity to study speech production at this detailed resolution by recording from multielectrode arrays previously chronically placed in human motor cortex as part of the BrainGate2 BCI clinical trial for people with paralysis[5,6]. The spiking rate of neurons in dorsal 'arm areas', where speech-related activity has not previously been reported, modulated during speaking. This finding challenges whether the conventional model of a 'motor homunculus' somatotopy[7] extends to the single-neuron scale. It is, however, consistent with known links between hand and speech networks that may reflect hand-mouth coordination and an evolutionary relationship between manual and articulatory gestures[8]. Two neural population dynamics features previously reported for arm movements were also present during speaking: a large initial condition-invariant signal[9], followed by rotatory dynamics[4,5]. This suggests that common neural dynamical motifs may underlie movement of arm and speech articulators. Lastly, spoken words and phonemes could be accurately decoded from single trials, demonstrating the potential utility of intracortical recordings for BCIs to restore speech.

Finding 1: Dorsal motor cortex neurons respond during speaking and orofacial movements.

Participant 'T5' performed a cued speaking task in which he heard a phoneme or word cue played from a computer speaker and spoke back that sound after hearing a go cue (**Fig. 1A**). We recorded SUA action potentials as well as potentially multiunit threshold-crossing spikes (TCs) and LFPs from electrodes in dorsal motor cortex (**1B**). This same dorsal neural population was previously shown to modulate during attempted arm and hand movements[6]. Here we found that these neurons' firing rates strongly modulated during speech (**1C**). Across the arrays, 13/22 neurons and 73/104 electrodes' TCs significantly responded during speaking at least one phoneme, with most responding to multiple phonemes (median = 4). This modulation was much greater when speaking than hearing sounds, and the same neural population was also active when T5 made orofacial movements of the mouth, lips, and tongue (**1D**). This suggests a broad encoding scheme related to motor cortical control of the speech articulators, consistent with previous ECoG findings in ventral motor cortex[3].

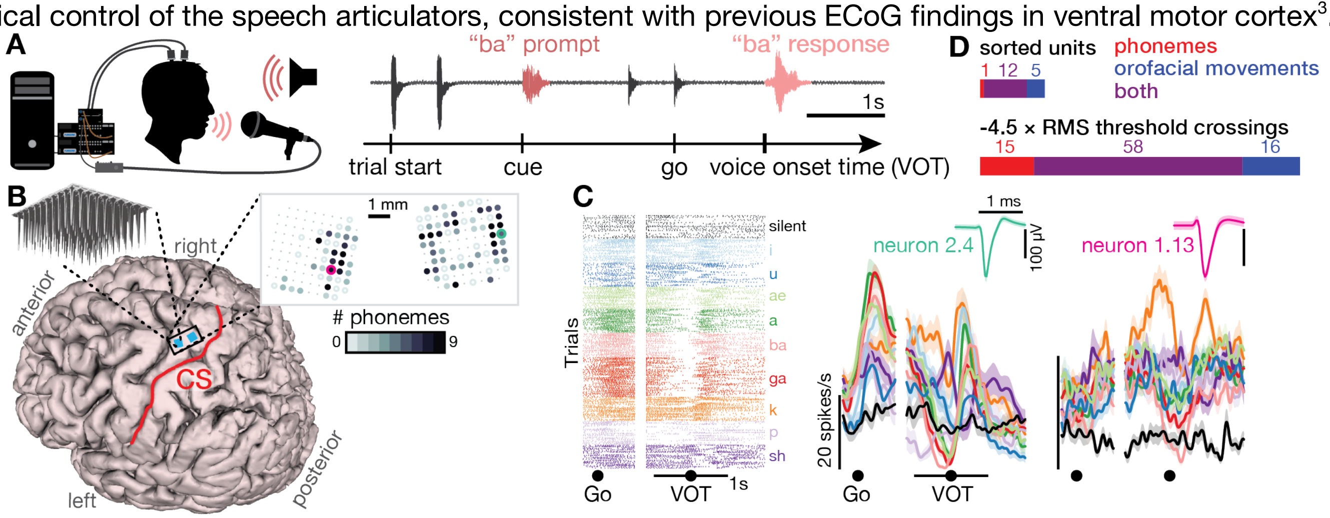

Fig. 1 A. Schematic of the cued speaking task. **B.** Blue squares mark the locations of the two 96-electrode arrays (1.5 mm length) in participant T5's 'hand knob' area of left motor cortex. The zoom-in inset shows electrode locations, with shading indicating the number of different phonemes for which significantly firing rate changes were recorded ($p < 0.05$, rank-sum vs. silence). **C.** Raster plot showing spike times of an example neuron across multiple trials of speaking nine different phonemes, or silence. Trial-averaged firing rates for this neuron and one other are shown to the right (mean ± s.e.). Insets show these neurons' action potential waveforms (mean ± s.d.). The electrodes where these neurons were recorded are circled in the panel B inset. **D.** How many neurons' (top) and electrodes' TCs (bottom) responded during only speaking phonemes (red), only non-speaking orofacial movements (blue), or both behaviors (purple).

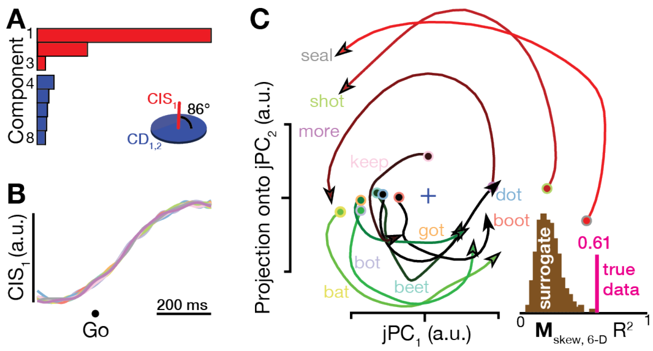

Fig. 2. A. Neural activity was decomposed into dPCA components like in [9]. Each bar shows the relative variance captured by each dPCA component, which consists of both condition-invariant variance (CI, red) and condition-dependent variance (CD, blue). These 8 dPCs captured 65% of the overall neural variance. The inset shows that the largest CIS component (CIS_1) was almost orthogonal to the largest CD components. **B.** Trial-averaged firing rates when initiating speaking different words (each trace) projected onto CIS_1.
C. Trial-averaged firing rates from 150 ms before to 100 ms after VOT were reduced to the top 6 PCs and then projected onto the first jPCA plane as in [4]. This plane captures 38% of the top 6 PCs' variance, and rotatory dynamics fit moment-by-moment neural state change with $R^2 = 0.81$ in this plane. Inset histogram shows significance testing of this result: the true data's R^2 in the full 6-PC subspace exceeded the R^2 from all 1000 surrogate datasets constructed using the population dynamics hypothesis testing method of [10] (specifically, the TNC tensor maximum entropy method).

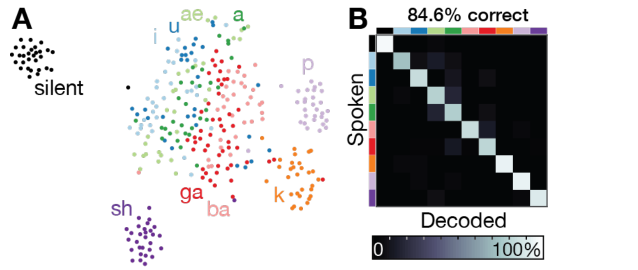

Fig. 3. A. t-SNE two-dimensional projections of single-trial neural feature vectors during speaking (all electrodes' spike counts and high frequency LFP power in ten 100 ms bins centered on VOT). **B.** Confusion matrix showing the accuracy of predicting which phoneme was spoken from single-trial neural activity (multiclass SVM, leave-one-trial-out cross-validated).

Finding 2: Motor cortical population dynamics exhibit low-dimensional structure during speech.
We tested whether two known key dynamical features of motor cortex population activity during arm reaching were also present during speaking. Prior monkey experiments showed that the neural state undergoes a rapid change during reach initiation which is dominated by a condition-invariant signal (CIS)[9]. This state change is believed to facilitate a shift from preparatory to movement neural dynamics. We observed similar dynamics during speech initiation: the CIS was the largest component of population activity (**2A**), its largest component (CIS_1) was virtually identical across speaking different words (**2B**), and the rise times of the CIS_1 predicted individual trials' speaking reaction times ($r = -0.14$, $p<0.05$, linear regression).

Previous monkey[4,9] and human[5] studies found that subsequent peri-movement population activity during arm and hand movements is characterized by orderly rotatory dynamics. We saw similar neural state rotations during speaking (**2C**). These results suggest that motor cortex may ubiquitously deploy these dynamical motifs – a large transient input that "kicks" the network into a different state from which activity lawfully evolves according to rotatory dynamics – across a variety of different behaviors (e.g., speech and reaching) to construct the desired muscle activity from an oscillatory basis set.

An important open question for these findings is to what extent these results were influenced by potential cortical remapping due to tetraplegia. Definitively resolving this ambiguity would require intracortical recording from this eloquent brain area in able-bodied people.

Finding 3: Spoken words and phonemes can be decoded from intracortical activity.
Examining the single-trial neural population activity during speaking revealed strong clustering by phoneme, with phonemes further grouping by phonetic similarities (**3A**). With an eye to evaluating the possible utility of these neural signals for speech BCIs, we quantified how well they could be used to classify which sound was being spoken on a single-trial basis. Prediction accuracy amongst nine phonemes, plus silence, was 84.6% (**3B**; shuffled chance accuracy was 10.1%). Prediction between ten short words, plus silence, was 83.5% accurate (chance was 9.1%).

These results compare favorably to previously published decoding accuracies using ECoG[1], despite our presumably suboptimal dorsal recording locations. This suggests a promising path forward for speech BCIs using intracortical electrophysiology, which can take advantage of an anticipated dramatic scale-up in the number of neurons that new, higher-density sensors will be able to record.

1. Mugler et al,. *J. Neural Eng.* (2014).
2. Chan et al., *Cereb. Cortex* (2014).
3. Chartier et al., *Neuron* (2018).
4. Churchland, Cunningham et al., *Nature* (2012).
5. Pandarinath et al., *Elife* (2015).
6. Pandarinath, Nuyujukian, et al,. *Elife* (2017).
7. Penfield & Boldrey, *Brain* (1937).
8. Gentilucci, Stefani & Innocenti, *Biolinguistics* (2012).
9. Kaufman et al*., eNeuro* (2016).
10. Elsayed & Cunningham, *Nat Neurosci* (2017).