**Title**

Human Sensorimotor Cortex Control of Directly Measured Vocal Tract Movements during Vowel Production

**Permalink**

https://escholarship.org/uc/item/5zg780gh

**Journal**

Journal of Neuroscience, 38(12)

**ISSN**

0270-6474

**Authors**

Conant, David F
Bouchard, Kristofer E
Leonard, Matthew K
et al.

**Publication Date**

2018-03-21

**DOI**

10.1523/jneurosci.2382-17.2018

Peer reviewed

Behavioral/Cognitive

# Human Sensorimotor Cortex Control of Directly Measured Vocal Tract Movements during Vowel Production

**David F. Conant,**[1,2,3,6] **Kristofer E. Bouchard,**[4,5] **Matthew K. Leonard,**[1,2,6] **and Edward F. Chang**[1,2,6]

[1]Department of Neurological Surgery, [2]Center for Integrative Neuroscience, [3]Graduate Program in Neuroscience, University of California–San Francisco, San Francisco, California 94143, [4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720, [5]Helen Wills Neuroscience Institute, University of California–Berkeley, Berkeley, California 94720, and [6]Weill Institute for Neurosciences, University of California-San Francisco, San Francisco, CA 94143

**During speech production, we make vocal tract movements with remarkable precision and speed. Our understanding of how the human brain achieves such proficient control is limited, in part due to the challenge of simultaneously acquiring high-resolution neural recordings and detailed vocal tract measurements. To overcome this challenge, we combined ultrasound and video monitoring of the supralaryngeal articulators (lips, jaw, and tongue) with electrocorticographic recordings from the cortical surface of 4 subjects (3 female, 1 male) to investigate how neural activity in the ventral sensory-motor cortex (vSMC) relates to measured articulator movement kinematics (position, speed, velocity, acceleration) during the production of English vowels. We found that high-gamma activity at many individual vSMC electrodes strongly encoded the kinematics of one or more articulators, but less so for vowel formants and vowel identity. Neural population decoding methods further revealed the structure of kinematic features that distinguish vowels. Encoding of articulator kinematics was sparsely distributed across time and primarily occurred during the time of vowel onset and offset. In contrast, encoding was low during the steady-state portion of the vowel, despite sustained neural activity at some electrodes. Significant representations were found for all kinematic parameters, but speed was the most robust. These findings enabled by direct vocal tract monitoring demonstrate novel insights into the representation of articulatory kinematic parameters encoded in the vSMC during speech production.**

*Key words:* electrocorticography; sensorimotor cortex; speech motor control; speech production; vowels

**Significance Statement**

**Speaking requires precise control and coordination of the vocal tract articulators (lips, jaw, and tongue). Despite the impressive proficiency with which humans move these articulators during speech production, our understanding of how the brain achieves such control is rudimentary, in part because the movements themselves are difficult to observe. By simultaneously measuring speech movements and the neural activity that gives rise to them, we demonstrate how neural activity in sensorimotor cortex produces complex, coordinated movements of the vocal tract.**

## Introduction

When we speak, we move the upper vocal tract articulators (lips, jaw, and tongue) to produce vocal tract constrictions of air flow in precise, rapid, and complex ways. These movements result in acoustic events that are highly distinguishable, maximizing communicative utility. In spoken languages, vowels are a major category of speech sounds. Despite their importance, it is unknown how cortical activity patterns control the vocal tract articulators to create vowels.

The ventral sensory-motor cortex (vSMC: precentral, postcentral, and subcentral gyri) is the primary cortical area controlling the speech articulators (Petersen et al., 1988; Lotze et al., 2000; Hesselmann et al., 2004; Brown et al., 2009; Kellis et al., 2010; Pei et al., 2011; Bouchard et al., 2013; Bouchard and Chang, 2014; Mugler et al., 2014; Kumar et al., 2016; Simonyan et al., 2016). Within vSMC, representations of vocal tract articulators are coarsely somatoto-

Received Aug. 22, 2017; revised Jan. 27, 2018; accepted Jan. 29, 2018.

Author contributions: D.F.C., K.E.B., and E.F.C. designed research; D.F.C. and K.E.B. performed research; D.F.C., K.E.B., and M.K.L. analyzed data; D.F.C., K.E.B., M.K.L., and E.F.C. wrote the paper.

This work was supported by National Institutes of Health Grant U01 NS098971 to E.F.C. and Grant F32-DC013486 to M.K.L. E.F.C. is a New York Stem Cell Foundation Robertson Investigator. This work was also supported by the New York Stem Cell Foundation, the Howard Hughes Medical Institute, the McKnight Foundation, the Shurl and Kay Curci Foundation, and the William K. Bowes Foundation. K.E.B. was supported by Laboratory Directed Research and Development funding from Berkeley laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy Contract DE-AC02-05CH11231.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Edward F. Chang, University of California–San Francisco, Department of Neurological Surgery, 505 Parnassus Avenue, M-779, San Francisco, CA 94143-0112. E-mail: edward.chang@ucsf.edu.

DOI:10.1523/JNEUROSCI.2382-17.2018

pically organized, with different neural populations in vSMC being associated with specific articulators (Crone et al., 1998; Brown et al., 2009; Bouchard et al., 2013; Mugler et al., 2014; Herff et al., 2015). However, our understanding of speech motor control in vSMC is incomplete, due to challenges in simultaneously acquiring neural and behavioral data with sufficient spatial and temporal resolution required to determine the precise correspondence between vSMC activity and the movement of the vocal tract articulators.

The precise movements (kinematics) of the articulators are challenging to measure because many of the vocal tract movements are internal to the mouth and throat, and are therefore difficult to monitor externally, especially in the context of neural recordings. As a result, previous studies have used the produced acoustics to infer which articulators are involved, based on extensive linguistic descriptions of speech movements for a given speech sound (Lotze et al., 2000; Crone et al., 2001; Brown et al., 2009; Fukuda et al., 2010; Kellis et al., 2010; Leuthardt et al., 2011; Pei et al., 2011; Grabski et al., 2012; Bouchard et al., 2013; Bouchard and Chang, 2014; Mugler et al., 2014; Herff et al., 2015). Although it is possible to describe the movements of each articulator according to phonetic labels derived from the acoustics, these behavioral descriptions cannot provide exact characterizations of the changing positions of the articulators over time. Moreover, there are many articulator configurations that can result in the same acoustics (Atal et al., 1978; Maeda, 1990; Johnson et al., 1993; Gracco and Löfqvist, 1994; Löfqvist and Gracco, 1999) and considerable across-speaker (Johnson et al., 1993) and across-trial (Perkell and Nelson, 1985) variability in movements that give rise to a particular speech sound. Thus, understanding how the brain produces complex sounds, such as vowels, requires determining how different kinematic parameters of articulatory movements are controlled in vSMC during speech production.

To understand how vSMC neural activity controls precise articulator movements, we have developed a system to simultaneously measure cortical activity using high-resolution electrocorticography (ECoG) while directly monitoring the lips and jaw with a camera, and the tongue with ultrasound. We previously detailed a technical description of the methods (Bouchard et al., 2016). Here, we examined how vSMC generates articulator kinematics, focusing on the production of American English vowels. We established that articulator kinematics are more strongly represented in vSMC compared with acoustics. We determined that specific kinematic parameters (position, speed, velocity, and acceleration) are all represented, although articulator speed is represented most strongly. Finally, we examined how distinct dynamics of neural activity are related to both movement (from rest to target position) and maintenance of articulators (at target position). By simultaneously measuring speech-related movements and the neural activity generating them, we demonstrate how neural activity in sensorimotor cortex produces complex, coordinated movements of the vocal tract.

## Materials and Methods

### Experimental design

*ECoG acquisition and signal processing*. Four human participants underwent chronic implantation of a high-density subdural ECoG array as part of the clinical treatment for epilepsy (3 female right hemisphere, 1 male left hemisphere). All subjects were implanted with 256 channel grids over peri-Sylvian cortex (1.17 mm diameter electrodes, 4 mm pitch, 60 × 60 mm coverage; Integra), referenced to a scalp electrode. The total number of vSMC electrodes for individual subjects ranged from 52 to 86 for a total of 270. Cortical-surface electrical potentials were recorded, and the voltage time series from each electrode was inspected for artifacts or excessive noise. Electrodes with excessive noise and time periods with artifacts were excluded from analysis, and the raw ECoG activity was rereferenced to the common average. For each channel, the time-varying analytic amplitude of the voltage signal in the high-gamma (HG) range (70–150 Hz) was extracted using the Hilbert transform, according to previously published procedures (Edwards et al., 2010). HG correlates well with multiunit firing (Ray and Maunsell, 2011) and has high spatial and temporal resolution (Muller et al., 2016). The HG signal was downsampled to 400 Hz for analysis and plotting purposes. HG power was *z*-scored relative to activity recorded during periods of silence during the same recording session. All analyses were limited to the vSMC, which was anatomically defined as the ventral portions of the precentral and postcentral gyri, as well as the subcentral gyrus.

*Task*. Participants listened to audio recordings of nine English vowels (*a*/*æ*/*ʌ*/*ɛ*/*ɝ*/*I*/*i*/*ʊ*/*u*) and were instructed to repeat each vowel. On each trial, to ensure they properly identified the vowel, they first heard it in an /h-V-d/ context (e.g., "hood"), and then they heard the vowel in isolation. After a 1–1.5 s delay, participants were presented with a visual cue to produce the isolated vowel. They were not explicitly instructed to hold the vowel for a specific amount of time. The median duration of production was 1.66 s (SD 0.35 s). For each participant, between 15 and 30 repetitions of each vowel were collected over the course of 3–6 recording sessions.

*Articulator tracking*. We developed a system to record the movements of the main supralaryngeal articulators while participants performed the vowel production task (see Fig. 1*A*), the details of which have been described previously (Bouchard et al., 2016). Briefly, to capture the movement of the lips and jaw, a camera was placed in front of the participant's mouth. The participant's lips were painted blue, and red dots were painted on the tip of the nose and the chin to simplify the process of extracting the shape and position of these articulators. The camera captured video at 30 frames per second. To image the tongue, an ultrasound transducer was held firmly under the participant's chin with the plane of view capturing the midline of the tongue. The ultrasound recorded images at 30 frames per second, and the data were aligned to the lips/jaw video according to the peak of the cross-correlation of the audio signals from each video. Using hue thresholding, we extracted the lips and jaw automatically from these videos as binary masks (see Fig. 1*B*). From these binary masks, we extracted the locations of the four corners of the mouth (upper/lower lip, left/right corners) and the jaw. For the tongue, we used EdgeTrak to extract 100 points of the mid-sagittal contour, which were then downsampled to 3 points by taking the median *x* and *y* value for the front, middle, and back thirds of the contour (M. Li et al., 2005). Because video and ultrasound were collected in orthogonal spatial planes, *x* and *y* positions in the lips/jaw images reflect left/right and top/bottom, whereas *x* and *y* positions in the tongue images reflect front/back and top/bottom. To correct for differences in the relative position of the camera and ultrasound transducer with respect to the participant, we referenced each articulatory point to the neutral starting position at the beginning of each trial. From the measured position of each articulatory feature (X), we also derived movement parameters including velocity ($X'$), speed ($|X'|$), and acceleration ($X''$) of that articulator. We refer to these parameters collectively as the articulator kinematics. While the lips and jaw were both included in all analyses, we found that lip opening and jaw height were correlated for this vowel production task (cross-subject average correlation: $r = 0.73 \pm 0.12$). Therefore, to simplify visualizations we only show results for the lips.

### Statistical analysis

*Acoustic feature extraction*. Speech sounds were recorded using a Sennheiser microphone placed in front of the participant's mouth and recorded at 22 kHz (see Fig. 1*A*). The speech signal was transcribed offline using Praat (Boersma, 2001). For each vowel, we extracted the formants (F1–F4) using an inverse filter method (Watanabe, 2001; Ueda et al., 2007; Bouchard et al., 2016).

*Trial duration standardization*. To standardize the durations of the vowels across trials and participants, we linearly resampled each trial to be the median duration across vowels and subjects (1.66 s). Behavior and neural signals changed with rapid and stereotyped dynamics around onset and offset; resampling the entire trial would systematically change those dynamics based on vowel duration. Therefore, to preserve onset and offset dynamics, we only resampled data in the time window from 250 ms after the onset of the acoustics to 250 ms before the offset: corresponding to the steady-state hold. Trials with durations less than half or greater than twice the median were excluded from analysis (26 in total across all subjects). Final analyses used an average of 15.3 ± 5.7 trials per vowel per subject.

*Permutation tests*. To evaluate statistical significance in each analysis, we used permutation tests. A permutation distribution for a given model was constructed by randomly permuting the trial labels of the observed data, and then training and testing the model using this shuffled data. This process was repeated 500 times, and the performance of these shuffled models comprised the permutation distribution. A model was considered significant if its performance on test data was greater than the 99th percentile of its corresponding permutation distribution. For the correlations in Figure 2*D*, we tested whether $|r| >$ 99th percentile of $|r_{\text{null}}|$.

*Correlations with articulatory position*. To evaluate the relationship between vSMC HG activity and individual articulators, we correlated HG activity at individual electrodes with the measured trial-to-trial position for each articulatory feature. HG activity averaged over a 200 ms window centered at acoustic onset was correlated with the mean position of each articulator taken from a 200 ms window centered at the midpoint of the vowel. Electrodes were labeled according to whether they had significant correlations with 0, 1, or multiple articulator positions. Electrodes in 2D are examples of electrodes with significant correlations with only one articulator.

*Encoding of kinematics, formants, and vowel categories*. We compared the representation of articulator kinematics, vowel formants, and vowel category at each electrode using $L_1$-regularized linear regression (Lasso). These models predict HG activity at each time point 500 ms before acoustic onset to 500 ms after acoustic offset from a sparse linear combination the behavior as follows:

$$HG_e = \sum\nolimits_{i=1}^{n} \beta \mathrm{i} \quad (1)$$

Where $HG_e$ is the HG power at a given electrode, $\beta$ are the linear weights that describe the mapping, and $i$ is a vowel category ($n = 9$), vowel formant feature ($n = 10$), articulator kinematic feature ($n = 40$), or all feature sets jointly ($n = 59$). The formant features were F1–F4, as well as all pairwise ratios of F1–F4. The articulator kinematic parameters were position, speed, velocity, and acceleration for lip opening, lip width, jaw height, and the front, middle, and back tongue. Vowel identity was parameterized as nine binary vectors corresponding to the vowel being produced during vocalization. Formant, articulatory, and vowel identity features were lagged 100 ms relative to HG, corresponding to the causal direction of neural activity generating behavior. This lag was determined empirically by optimizing model performance over a range of lag values (−500 ms to 500 ms).

To train and test linear models, we used $L_1$-regularized linear regression in a leave-one-trial-out cross-validation procedure. We calculated the correlation between the observed and predicted HG values, averaged across cross-validations. Electrodes were included in visualizations and summary statistics only if their performance passed the permutation test described above for at least one of these models (i.e., formants, kinematics, vowel identity, or combined). To compare models with different numbers of parameters, we calculated the adjusted $R^2$ as follows:

$$1 - (1 - R2)(n - 1/n - p - 1) \quad (2)$$

Where $R^2$ is the unadjusted coefficient of determination of the model, $n$ is the number of observations on which the model was trained, and $p$ is the number of parameters.

*Organization of vowels in behavioral and neural spaces*. To examine the similarity of vowels in behavioral and neural representation spaces, we used multidimensional scaling (MDS). MDS provides a low-dimensional projection of the data that preserves the relative distances (similarities) between points in a higher-dimensional space. For each feature set (formants, articulator position, and neural), we extracted the median value for each vowel from a 200 ms window centered at the midpoint of the vowel (formants and articulator position) or the onset (neural), and then *z*-scored that value across vowels. We applied MDS to the distance matrix computed on these measurements for each feature set separately. To measure the differences in the organization of vowels between the formant, articulator, and neural spaces, we calculated the pairwise distances between the vowels in each space. We quantified the similarity between the neural and kinematic or formant spaces by calculating a bootstrapped correlation between the pairwise distances for each feature set. We performed agglomerative hierarchical clustering on the pairwise distances to visually organize the results.

*Encoding of kinematic parameters across time*. To assess the relative encoding of different kinematic parameters, we used the measured position of each articulator on each trial (X) to derive the velocity (X′), speed (|X′|), and acceleration (X″) of that articulator on that trial. To examine the encoding of these parameters independent of one another, we removed the shared variance between these parameters using semipartial correlation. For each time point, we first used linear regression to predict the values of one kinematic parameter, $y$, from a linear combination of the remaining three parameters, X as follows:

$$\hat{y} = \beta X \quad (3)$$

Where $\beta$ are the weights that describe the linear relationship and $\hat{y}$ is the model's prediction of that kinematic parameter. We then calculated the linearly independent component of the kinematic parameter, $y_{idp}$, by subtracting predicted parameter values from the observed as follows:

$$y_{idp} = y - \hat{y} \quad (4)$$

We then used $L_1$-regularized linear encoding models to predict HG activity from the kinematic parameters (position, speed, velocity, and acceleration) of the lips, jaw, and tongue. However, instead of including the entire trial time course in each model, we trained and tested models within 100 ms nonoverlapping windows that tiled the trial. Articulator kinematics were lagged 100 ms relative to HG to evaluate the causal nature of neural activity on behavior. Models were trained and tested independently for each time window. Performance was measured by the correlation between the observed and predicted HG values, averaged across cross-validations. Electrodes were included in visualizations and summary statistics only if their performance passed the permutation test described above for at least three contiguous time windows at any point in the trial.

*Decoding of kinematic parameters from vSMC HG*. To determine the degree to which variations in individual articulatory parameters could be predicted from vSMC population activity, we used linear decoding models. Similar to the encoding models above, we built $L_1$-regularized linear models to relate vSMC HG to articulator kinematics within 100 ms time windows that tiled the trial. However, instead of predicting the HG activity at a single electrode from a combination of all articulator parameters, we predicted the trial-to-trial variance of each articulator parameter from a combination of all vSMC electrodes. As with the encoding models, articulator kinematic features were lagged 100 ms relative to vSMC HG, and models were trained and tested independently for each time window using a leave-one-out cross-validation procedure. The resulting models thus express how well the vSMC population can predict each resulting articulator kinematic feature as a function of time within the vocalization. To simplify visualization, we averaged performance across subjects and similar articulators.

*Description of vSMC HG dynamics*. To characterize the major physiological response types in HG dynamics, we used non-negative matrix factorization (NMF). NMF is a dimensionality reduction technique that extracts a predetermined number (i.e., rank, $k$) of bases ($B \in \mathbb{R}^{mxk}$) and weights ($W \in \mathbb{R}^{nxk}$) that linearly combine to reconstruct the non-negative data ($A \in \mathbb{R}^{mxn}$), such that $k < \min(n, \mathrm{m})$ under the constraint that both the bases and weights are strictly non-negative as follows:

$$A \approx BW^T;\ B,\ W \geq 0 \quad (5)$$

The solutions B and W are found by solving the (biconvex) constrained optimization problem as follows:

$$\hat{B}, \hat{W} = \min_{B,W} \frac{1}{2} \|A - BW^T\|_F^2; \; s.t. \; B, W \geq 0 \quad (6)$$

NMF is particularly useful for decomposing data into "parts" that have interpretable meanings (e.g., transient vs sustained response types) (Lee and Seung, 1999; Donoho and Stodden, 2004; Bouchard et al., 2016; Leonard et al., 2016). The HG activity for each vSMC electrode across all participants was averaged across trials, offset by the minimum value (such that all values were positive), and NMF was applied to the matrix of time courses × electrodes. To determine a parsimonious number of bases, we calculated the reconstruction error when projecting the data onto the NMF bases as follows:

$$err = \frac{1}{2} \|A - BW^T\|_F^2 \quad (7)$$

We then found the number of bases (i.e., rank $k$) beyond which reconstruction error only marginally reduced (i.e., the elbow of the curve): five bases were used. The first two bases resembled the transient and sustained activity observed in Figure 7*A*. Electrodes with sustained activity were defined as those that had weighting for basis 1 greater than for basis 2. The width ($HG_w$) of the HG activity for sustained electrodes was derived as follows:

$$HG_w = argmin\int(HG_{e,t} - \overline{HG_{e,t}}) - argmax\int(HG_{e,t} - \overline{HG_{e,t}}) \quad (8)$$

Where $HG_{e,t}$ is the HG activity at given sustained electrode for a given trial. This measure was calculated for each sustained vSMC electrode, for each trial. We assessed spatial organization by measuring the Euclidean distance between electrodes organized according to their maximum NMF weight (i.e., transient or sustained). We compared distributions of intraparameter distances and cross-parameter distances with randomized distributions derived by shuffling the labeling of the electrodes. If the HG dynamic variability across vSMC is spatially organized, the distribution of intraparameter and cross-parameter distances should differ from the distributions of the random distributions.

## Results

Participants produced nine English vowels in isolation (*a/æ/ʌ/ɛ/ɝ/I/i/ʊ/u*) (e.g., the vowels pronounced as in the following set of words: "calm," "cat," "send," "fun," "heard," "sit," "need," "should," "boot") while neural activity was recorded from vSMC and the movements of the supralaryngeal articulators were monitored. These vowels span both the acoustic and kinematic space of all American English vowels, and are a basic and essential component of all languages. We specifically studied vowels in isolation for several reasons. First, the associated movements of the speech articulators consist of a single displacement from rest, to the target position, and back to rest. This simplicity provides the opportunity to study isolated movements of the speech articulators free from the context of surrounding phonemes. The task was also

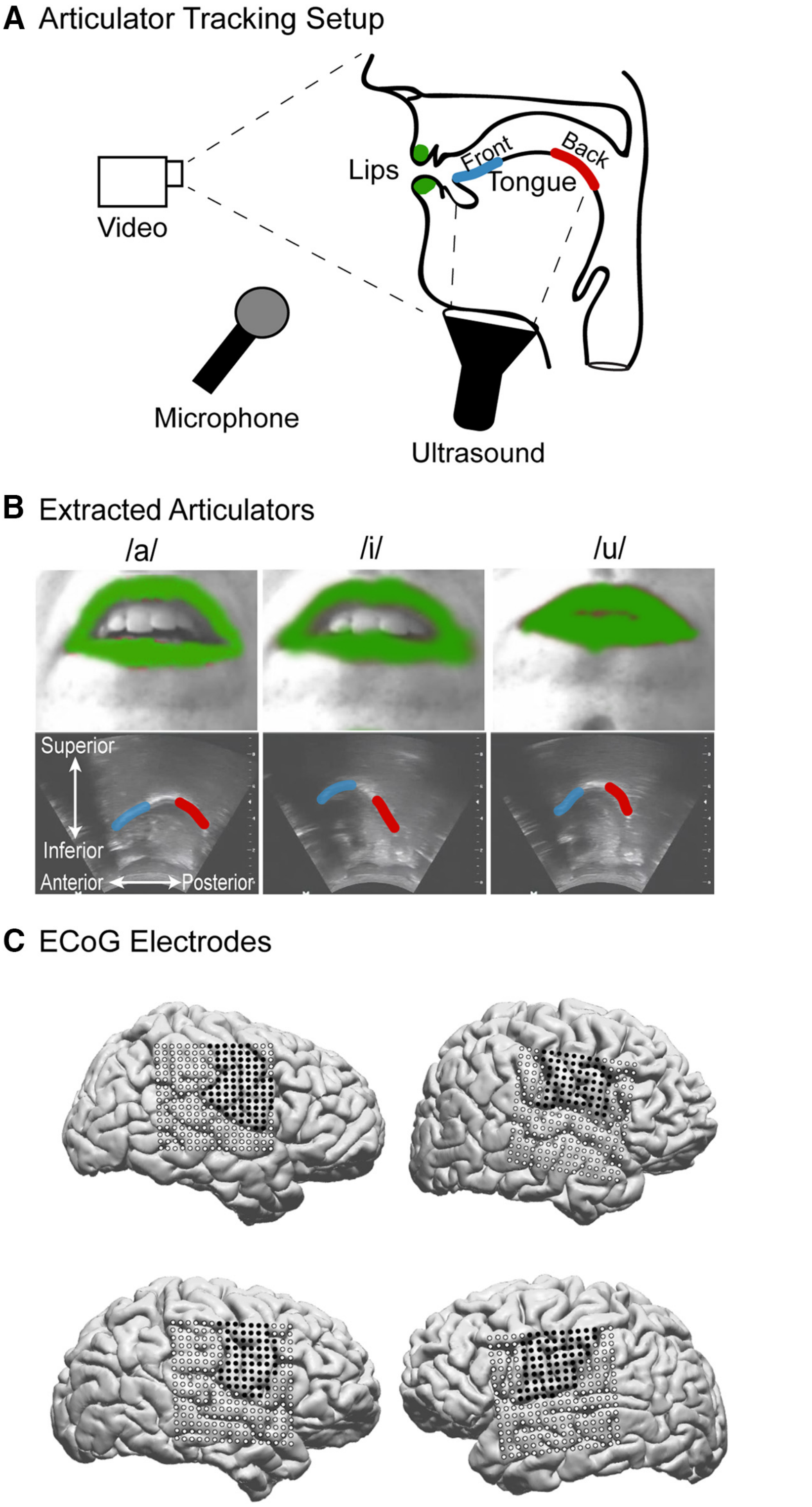

**Figure 1.** Experimental setup and articulator monitoring. ***A***, Schematic of the articulatory tracking system. A video camera placed in front of the subject recorded the movements of the lips while an ultrasound transducer under the jaw captured the tongue contour. ***B***, Example images of the the video (top) and ultrasound (bottom) imaging during production of the corner vowels /a/, /i/, and /u/. The lips and tongue contour were extracted from these images, and the resulting binary masks are shown in color on top of the raw images. ***C***, MRI reconstruction of the brains of the 4 subjects included in the study. Coregistered ECoG electrodes are plotted on the cortical surface. Dark points indicate electrodes over vSMC.

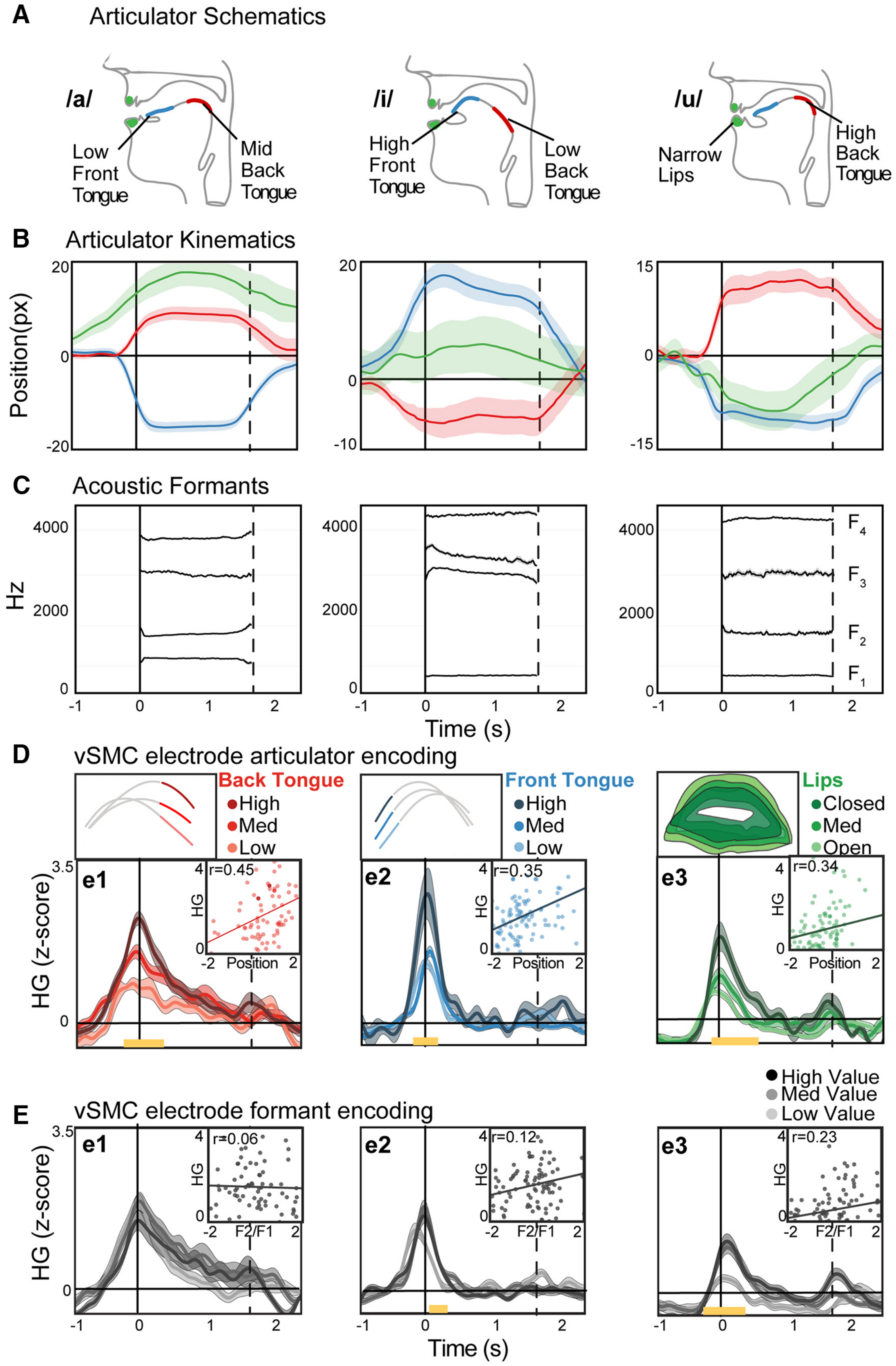

**Figure 2.** Articulatory and acoustic behavior correlates with single electrode vSMC neural activity. ***A***, Prototypical articulator positions for the corner vowels /a/, /i/, and /u/. ***B***, Average (±SEM) time courses of measured articulator displacements during production of the corner vowels. For illustration, the kinematic parameter of vertical position is shown; however, both vertical and horizontal measurements are used for subsequent analyses. ***C***, Average formant values (F1–F4) for the corner vowels during the same productions as in ***B***. (*Figure legend continues.*)

designed to minimize variability in the lower vocal tract (e.g., larynx), which we did not explicitly monitor. Using the recorded acoustics, we verified that subjects produced the vowels with little trial-to-trial variation in either pitch or intensity. Additionally, the movements occur at distinct epochs, allowing us to resolve the neural representation of the movement to the target, from the maintenance of that target, from the return to the resting configuration.

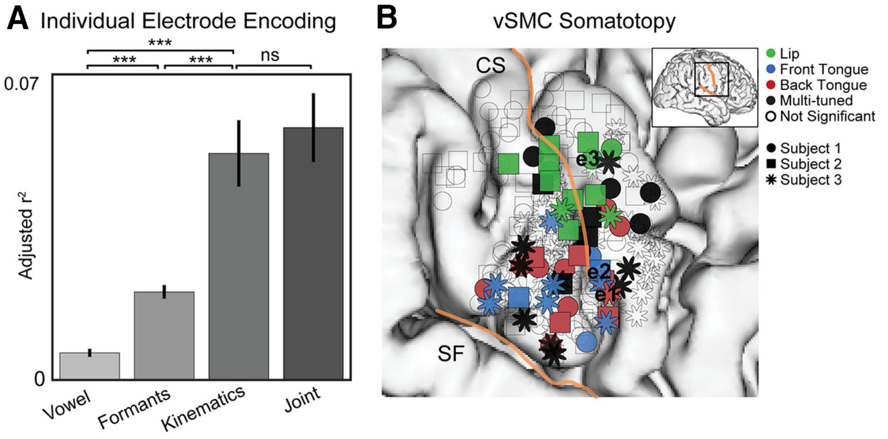

**Figure 3.** vSMC activity primarily encodes speech articulators. ***A***, Performance of encoding models predicting vSMC HG using vowel identify, acoustic formants, articulator kinematics, or all three. Articulator kinematics explain vSMC activity better than vowel identity and acoustic formants. ***$p < 0.01$ (Wilcoxon rank sum). Furthermore, the joint model does not explain more variance than the kinematic model alone, indicating that the vowel identity and acoustic formant models are likely driven by variance shared with the kinematics. ***B***, Electrodes over vSMC from 3 right hemisphere subjects were warped onto a common brain and color-coded according to articulator selectivity. Empty circles represent electrodes with no significant selectivity for any articulator. Black electrodes are selective for more than one articulator. Blue, red, and green electrodes are selective for front tongue, back tongue, or lips, respectively.

## Articulator tracking during vowel production

We simultaneously tracked the movements (Fig. 1*A*) of the major supralaryngeal articulators (i.e., lips, jaw, and tongue; Fig. 1*B*) while recording neural activity directly from the cortical surface (Fig. 1*C*; see Materials and Methods) (Bouchard et al., 2016). We first verified that, by extracting the positions of the articulators, we observed characteristic vocal tract configurations that reflect distinct vowels. For example, the vowel /a/ is characterized by lowering the front tongue, raising the back tongue, and opening the lips, whereas the vowels /i/ and /u/ have different configurations (Fig. 2*A*). The measured articulatory movements captured these characteristics and clearly discriminated vowel categories (Figs. 1*B*, 2*B*). We also used the produced acoustics as a behavioral measure of vowel discriminability. By extracting the formants from the acoustic signal, we observed distinct relative patterns of acoustic power for different vowels. For example, /a/ is characterized by high F1 and low F2, whereas /i/ and /u/ have different formant profiles (Fig. 2*C*).

## Representation of articulator kinematics in ventral sensory-motor cortex

These descriptions demonstrate that both articulator kinematics and acoustic formants provide rich descriptions of the same behavior. However, although kinematics and acoustics are causally related, their relationship is not 1:1 (Atal et al., 1978; Maeda, 1990; Johnson et al., 1993; Gracco and Löfqvist, 1994; Löfqvist and Gracco, 1999), nor are they perfectly correlated (in the present dataset, $\text{rho}_{\text{kin,acoust}} = 0.53 \pm 0.17$). For example, producing the vowel /uw/ ("h*oo*t") involves raising the back of the tongue toward the soft palate while rounding the lips. However, those movements can be compensatory. The vowel /u/ can be produced with less pronounced lip movements accompanied by greater tongue movements, or vice versa (Perkell et al., 1993). Therefore, we asked whether articulator kinematics or acoustic formants are the behavioral characterization of vowels represented in vSMC.

First, we quantified how well the positions of speech articulators or vowel formants explain the variance of HG at individual vSMC electrodes (i.e., encoding strength). We recorded cortical electrical potentials from a total of 270 electrodes from the surface of vSMC across 4 subjects (Fig. 1*C*). The HG activity at many vSMC electrodes was elevated above baseline during the speech movements and was significantly correlated with the trial-to-trial position of the speech articulators (Fig. 2*D*). We observed a clear relationship between articulator position and HG activity. For illustration, we identified representative electrodes where activity was most correlated with a single articulator. For example, the HG activity of electrode 1 at the time of vowel onset was significantly correlated with only the back tongue. Likewise, electrode 2 showed greater activity for higher front tongue positions. Electrode 3 was correlated with the opening of the lips. To examine whether HG activity at these electrodes was similarly correlated with the produced acoustics, we binned the activity by formant values (Fig. 2*E*). We observed weaker correlations with formants compared with articulator position, demonstrating more robust encoding of articulatory representations.

We were specifically interested in whether vSMC activity is best explained by articulator kinematics, vowel formants, or vowel identity. We used linear encoding models to predict neural activity from kinematic or acoustic features, or the vowel identity (see Materials and Methods). Across electrodes, we found that articulator kinematics provided significantly better model fits compared with vowel formants ($U = 5.3$, $p = 1.0\text{e-}8$; Wilcoxon rank sum) or vowel identity ($U = 8.7$, $p = 4.1\text{e-}18$; Wilcoxon rank sum) (Fig. 3*A*). We used nested models to examine how much additional neural variance is explained by predicting HG from both articulator kinematics and vowel formants. We found that the joint model explained no more variance than the articulator kinematics alone ($U = 0.4$, $p = 0.7$; Wilcoxon rank sum), suggesting that the performance of the formant models was likely driven by variance shared with the kinematics. Therefore, we find no evidence for encoding of vowel formants separate from their

←

(*Figure legend continued.*) ***D***, HG activity in three example vSMC electrodes. Each electrode was selected to be representative for the articulator shown in the top subplot (median configurations shown for back tongue, front tongue, and lips). Trials for each electrode are binned by the displacement of the articulator that best correlates with the HG values at acoustic onset (±100 ms). Yellow bars represent time points of significant difference between the bins (e1: $F_{(2,75)} < 5.1$, $p < 0.01$; e2: $F_{(2,102)} < 4.9$, $p < 0.01$; e3: $F_{(2,105)} < 4.8$, $p < 0.01$; ANOVA) "Scatter plots show the correlation between HG at acoustic onset and articulator position at the midpoint of the vowel." ***E***, HG activity in the same electrodes binned according to formant values (F1/F2 ratio).

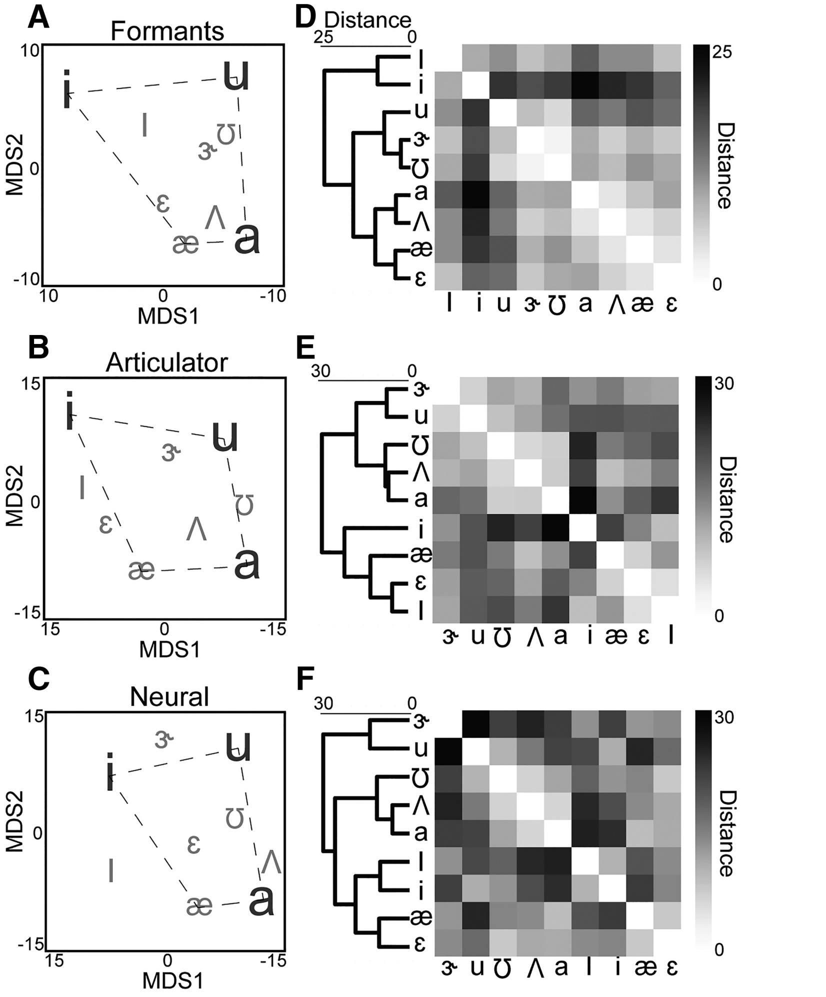

**Figure 4.** vSMC activity reflects articulator kinematic organization of vowels. ***A–C***, MDS representations of (***A***) acoustic formants, (***B***) articulator position, and (***C***) vSMC HG activity. Each letter indicates the position of the median production of the indicated vowel in MDS space across all subjects. The relative organization of the vowels is similar across spaces. For example, the low-back vowel /a/ is always near the mid-back vowel /Λ/, but far from the high-front vowel /i/. ***D–F***, Hierarchical clustering (left) and confusion matrices (right) derived from the pairwise distances between vowels in the MDS spaces of (***D***) acoustic formants, (***E***) articulator position, and (***F***) vSMC HG.

articulatory origin. Furthermore, these results demonstrate that the production of distinct vowels is grounded in direct control of articulator kinematics.

Across all vSMC electrodes, we found that 27% (74 of 270) were significantly correlated with movements of one or more articulators (correlations >99th percentile of permutation distributions). We observed a clear spatial organization to the articulator correlations, with lips/jaw more dorsal than the tongue (Fig. 3*B*), consistent with previously described somatotopy (Penfield and Roberts, 1959; Huang et al., 1988; Brown et al., 2009; Bouchard et al., 2013; Conant et al., 2014). Within the ventral region, we observed electrodes that more strongly reflected either the front or back of the tongue. Both front and back tongue electrodes were distributed throughout the broader tongue region. Finally, we observed 45 electrodes that had significant correlations with multiple articulators, which were distributed throughout vSMC. Together, these results extend our understanding of speech-motor cortex somatotopy by demonstrating that the dominant encoding scheme in these neural populations reflects the specific movements of the preferred articulators.

### Organization of vowels in vSMC population activity

To understand how vSMC encoding of articulator kinematics contributes to our ability to produce distinct vowels, we examined the organization of behavioral and neural activity in relation to all nine vowels. In addition to articulator kinematic representations at individual electrodes, population activity in vSMC may reflect the coordinated movements of the vocal tract that produce vowels. Furthermore, because vowel formants arise from the relative positions of multiple articulators, it may be the case that, while articulators are most strongly represented at single electrodes, population activity may reflect a different, emergent representation. We examined the organization of speech representations at the population level by comparing the relative distances of vowels in acoustic, articulatory, and neural space. We performed MDS on the vowel centroids (see Materials and Methods) measured by vowel formants, articulator position, or vSMC neural activity across all participants. In this analysis, vowel tokens that are near each other in MDS space have similar formant, kinematic, or HG values. Consistent with previous behavioral and linguistic descriptions of vowel production, the formant and kinematic MDS projections replicate the classic vowel space "trapezoid" (Fig. 4*A–C*) (Hillenbrand et al., 1995; Ladefoged and Johnson, 2011). The HG neural MDS projection also closely resembled the acoustic/kinematic organization of the vowels. For example, the vowel /a/ (as in h*a*ll) is near the vowel /Λ/ (as in h*u*t), but far from /i/ (as in h*ea*t).

To characterize the difference in organization of vowels across these feature spaces, we calculated the pairwise distances between the vowels in MDS space, visualized as confusion matrices (Fig. 4*D–F*, right). We additionally performed hierarchical clustering of the pairwise distances and organized the confusion matrices by the derived hierarchical organization. The pairwise distances and hierarchical clustering reaffirm the classic vowel organization but also highlight the specific differences between the feature spaces. For example, /i/ is distant from the other vowels in the formant space but closer in the articulator and neural spaces. We found that the organization of vowels in vSMC HG activity is significantly more correlated with the organization of vowels in the articulator space compared with the acoustic space ($U = 9.5$, $p = 1.3\text{e-}21$, Wilcoxon rank sum), although both representations were significantly correlated with vowel organization in the HG neural space (acoustic: $r = 0.56$, $p = 2.8\text{e-}4$, kinematic: $r = 0.73$, $p = 5.9\text{e-}5$).

### Encoding of articulator kinematic parameters

In the above analyses, we considered the joint encoding of multiple kinematic parameters for individual articulators. However, it is unknown whether kinematic encoding reflects particular aspects of the articulator movements. The movements of the articulators can be described according to a variety of different kinematic parameters (e.g., position, speed or velocity, acceleration). For each kinematic parameter, we used L1-regularized encoding models to explain vSMC HG from the moment-to-moment measurements of position, speed, velocity, and acceleration. Because all four kinematic parameters are correlated with one another, we removed shared variance between the parameters using semipartial correlations to better interpret their relative encoding performances.

We found electrodes that significantly encoded the trial-to-trial variability in position (Fig. 5*A*), speed (Fig. 5*B*), velocity (Fig. 5*C*), and acceleration (Fig. 5*D*). Speed was the most robustly encoded parameter at most vSMC electrodes, with significant encoding at more electrodes and a higher average correlation compared with the other parameters ($U$ = 1720 to 2735, $p$ = 3.3e-9 to 1.1e-14; Wilcoxon rank sum; Fig. 5*E*).

To understand the timing of kinematic parameter encoding throughout the production of vowels, we also examined models that predicted HG neural activity from the joint combination of all four parameters simultaneously. These models were evaluated over a sliding 100 ms window to characterize the kinematic parameter encoding during different phases of the trial (i.e., movement initiation, target position, steady-state maintenance, and movement back to the starting position). We observed a peak in encoding for most electrodes around the onset of the movement (91% of electrodes), with some electrodes also showing a peak around the offset (9%; Fig. 5*F*). There was no spatial organization associated with electrodes that specifically encoded particular parameters (intraparameter, $p$ = 0.31; cross-parameter, $p$ = 0.08; see Materials and Methods), nor was there a significant relationship between electrodes that encoded specific kinematic parameters and specific articulators ($\chi^2$ (9, $N$ = 155) = 9.26, $p$ = 0.4; $\chi^2$). Strikingly, encoding during the steady state was near zero for all kinematic parameters.

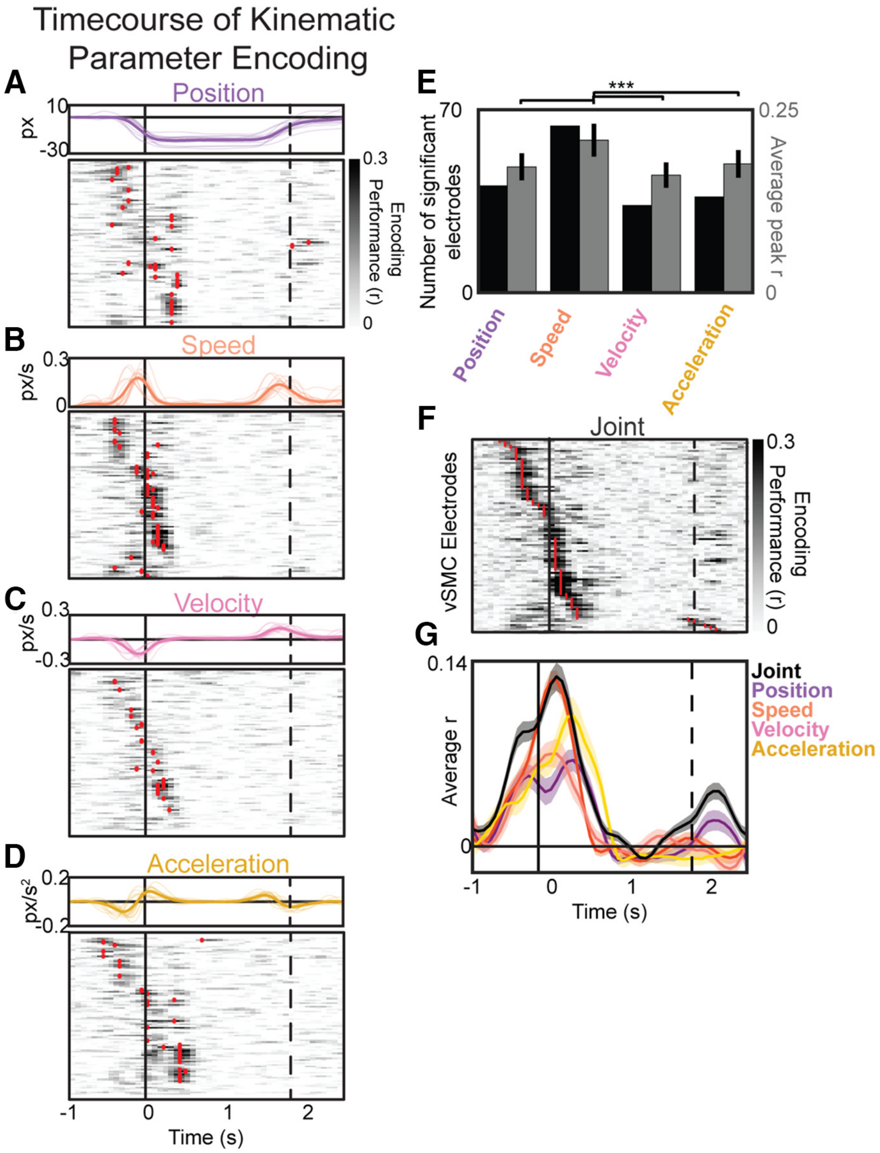

**Figure 5.** Representations of position, speed, velocity, and acceleration kinematic features over time. ***A–D***, Top, Example kinematic parameters of position (***A***), speed (***B***), velocity (***C***), and acceleration (***D***) for all utterances of /a/ for one subject. Thin lines indicate individual trials. Thick line indicates the across-trial average. Bottom, Performance of encoding models predicting vSMC HG from articulator position (***A***), speed (***B***), velocity (***C***), or acceleration (***D***). vSMC electrodes with significant performance are marked by red dashes at the time peak encoding performance. ***A–D***, Electrodes are plotted in order of their peak encoding times in the joint model (***F***). Vertical black lines indicate the onset (solid) and offset (dashed) of vowel acoustics. ***E***, Comparison of the number of significant electrodes (black) and average peak performance (gray) of position, speed, velocity, and acceleration encoding models. Speed is significantly encoded at more electrodes, with a higher average model performance. ***$p$ < 0.01 (Wilcoxon rank sum). ***F***, Performance of encoding models predicting vSMC HG from all articulator kinematics jointly. ***G***, Average performance across significant electrodes for the joint and independent parameter models.

To understand how these individual electrode kinematic representations relate to the population representations of articulator kinematics and dynamics, we used L1-regularized decoding models to predict the articulator kinematics from the population of vSMC HG electrodes. As with the encoding analyses, these models were constructed from a small (100 ms) sliding window of time, resulting in a description of how much of the trial-to-trial variability of the articulator position (Fig. 6*A*), speed (Fig. 6*B*), velocity (Fig. 6*C*), and acceleration (Fig. 6*D*) can be explained by vSMC HG activity. The time course of decoding strength was similar to the encoding models, with peaks around the onset and offset and near-zero values while the vowel was being held. Across kinematic parameters, articulator speed was the best-predicted parameter ($U$ = 2.6 to 3.3, $p$ = 8.1e-3 to 9.0e-4; Wilcoxon rank sum).

Together, these results demonstrate a strikingly sparse representation of kinematic parameters across time, despite the fact that there continues to be trial-to-trial variability in both kinematic and neural features throughout vowel (Fig. 5). Only 56% of time points had significant encoding performance at any elec-

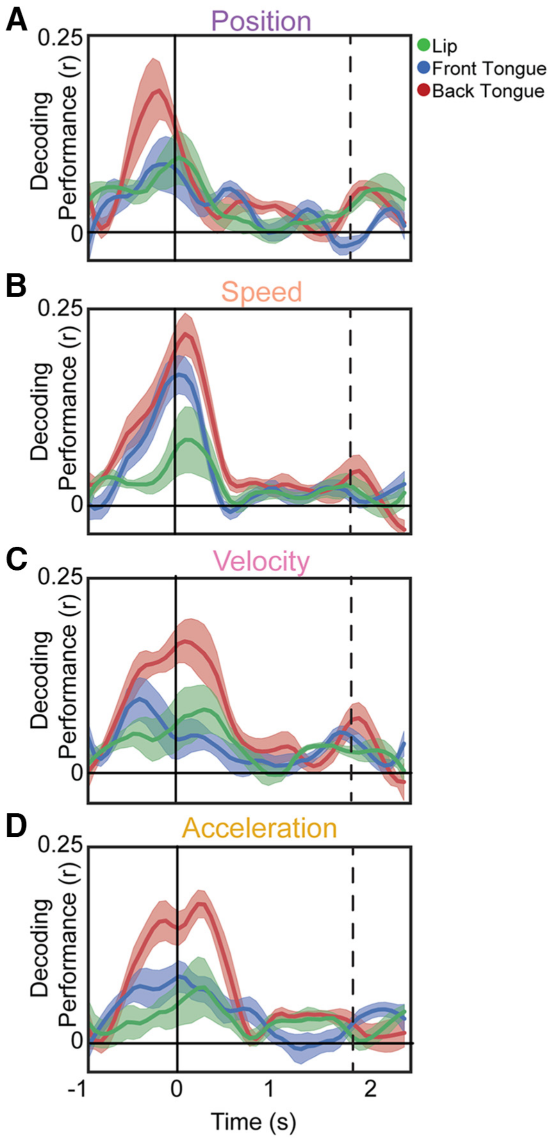

**Figure 6.** Time course of kinematic parameter decoding. ***A–D***, Performance of decoding models predicting articulator position (***A***), speed (***B***), velocity (***C***), or acceleration (***D***) from HG at all vSMC electrodes. Features are averaged across subjects and within articulators. Black lines indicate the onset and offset of vowel acoustics.

trode, and individual significant electrodes had an average of 15% (±1) significant time points. In particular, we did not observe any electrodes that exhibited significant kinematic parameter encoding during the steady state of the vowel.

### Onset versus steady-state HG activity and kinematic encoding

The temporal sparsity of neural representations described above is particularly notable given that many electrodes showed sustained HG activity during the steady-state portion of the vowel, independent of the particular articulatory movements that occurred (Fig. 7*A*). These electrodes contrast with other HG activity that is only transiently increased around the onset and/or offset of the vowel (Fig. 7*A*). To characterize these response types, we used NMF to derive basis functions that best describe vSMC HG temporal profiles across all electrodes (Hamilton et al., 2017). Our motivation for using NMF was not to provide a complete description of HG dynamics, but rather to provide an unsupervised method of quantifying the transient and onset/offset responses across electrodes. We found that the first two bases (i.e., the most important bases) captured the sustained and onset/offset response types we observed qualitatively (Fig. 7*B*). Organizing all vSMC electrodes by the degree to which their activity is reconstructed by the first or second NMF bases (i.e., the NMF weights), we observed a continuum of HG dynamics: some electrodes showed sustained activity throughout the utterance, whereas others showed transient increases in activity only at onset and offset of the utterance (Fig. 7*C*). Some electrodes showed a combination of sustained and transient components. There was no apparent spatial organization (intraresponse type, $p = 0.13$; cross-response type, $p = 0.09$; see Materials and Methods) or relationship between response type and the articulators represented at each electrode ($\chi^2$ (3, $N = 155$) = 2.3, $p = 0.5$; $\chi^2$).

We separately considered electrodes that showed stronger weights for the sustained NMF basis (basis 1 in Fig. 7*B*). The average HG activity at these electrodes indeed showed sustained activity throughout the vowel; however, there was not a concomitant sustained encoding of kinematic parameters (Fig. 7*D*). This dissociation between activity and encoding was apparent even at the single-trial level (Fig. 8*A*). Thus, although some electrodes exhibit sustained HG activity throughout the production of the vowel, there is not a systematic relationship between the trial-to-trial variability of that activity and the kinematics of the articulators. Instead, encoding of kinematics at electrodes with sustained activity was prevalent only around the onset and offset of movement. We hypothesized that, although activity during the steady state does not relate to kinematic variability, it still reflects an important aspect of the task, namely, the duration of each utterance. Across sustained electrodes, we found that the duration of the HG time course was significantly correlated with the duration of the vocalization (Spearman's $\rho = 0.61$, $p = 2e\text{-}153$; Fig. 8*B*). Thus, at a minimum, the sustained activity was associated with vowel production.

## Discussion

We report a detailed description of how activity in speech-motor cortex controls the precise movements of the vocal tract articulators to produce vowels. By simultaneously measuring the movements of the articulators, recording the acoustic consequences of those movements, and recording the neural activity in vSMC, we are able to establish that the dominant representation in vSMC is articulator kinematics. The precise control of these movements allows speakers to create specific configurations of the mouth, which lead to distinct categories of sounds.

Without simultaneous measurements of the articulators, previous studies of the neural basis of speech production have relied on approximate, categorical phonemic-based descriptions of speech behavior (Crone et al., 2001; Fukuda et al., 2010; Kellis et al., 2010; Leuthardt et al., 2011; Pei et al., 2011; Bouchard et al., 2013; Bouchard and Chang, 2014; Mugler et al., 2014; Herff et al., 2015). Although the produced acoustics and categorical vowel descriptions reflect the ultimate (perceptual) outcome of vocal tract movements, the many-to-one relationship between kinematics and vowels (Atal et al., 1978; Maeda, 1990; Johnson et al., 1993; Gracco and Löfqvist, 1994; Löfqvist and Gracco, 1999) means that it was not possible to understand the precise nature of the neural representation in vSMC. Previous studies have implicated vSMC in articulator kinematic control in several ways. First, stimulation to sites in vSMC elicits involuntary activations of the orofacial muscles (Penfield and Boldrey, 1937; Huang et al., 1988). Second, neurons in these and other sensorimotor regions are often tuned to movement kinematics (Georgopoulos et al., 1986; Paninski et al., 2004; Arce et al., 2013). Finally, the spatiotemporal patterns of HG activity in vSMC are consistent with the engagement of the articulators (Bouchard et al., 2013). The present results confirm these interpretations by showing directly that kinematic descriptions of speech behavior are more closely related to neural activity compared with acoustic or categorical vowel descriptions. Further, we find no evidence that vSMC activity encodes either produced acoustics or vowel category distinct from their correlations with the articulator kinematics. Crucially, we observed that this encoding scheme exists both at single electrodes and across the spatially distributed set of electrodes. For spatially distributed activity patterns, we demonstrate the neural existence of the classic vowel "trapezoid," which has dominated linguistic descriptions of speech (Harshman et al., 1977; Alfonso and Baer, 1982; Hillenbrand et al., 1995).

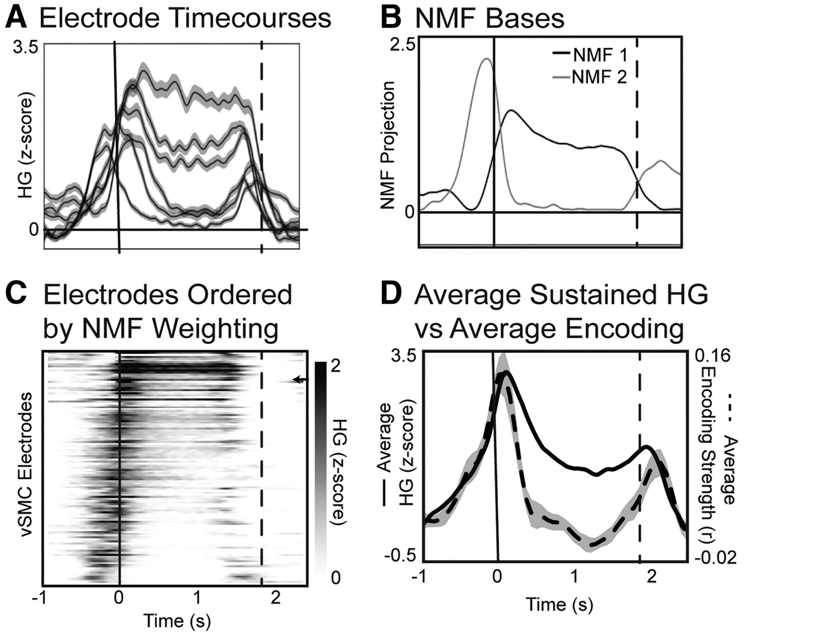

**Figure 7.** Relationship of vSMC HG dynamics and kinematic encoding. ***A***, HG activity at several example electrodes illustrating the diverse dynamics during the same behavior, especially during the time period when the vowel is being held. ***B***, The first two NMF bases extracted from HG dynamics across all electrodes. These bases recapitulate the key differences in dynamics seen in the example electrodes, and serve as an unbiased quantification of the HG dynamics seen across vSMC. ***C***, HG activity at all vSMC electrodes ordered by the ratio of NMF bases used to reconstruct their activity ($(NMF1 - NMF2)/(NMF1 + NMF2)$). Arrow indicates the example electrode used in Figure 8. ***D***, Average HG activity across all sustained (NMF1 > NMF2) electrodes (solid) plotted alongside the average encoding performance (dashed) across time. During the steady state of the vowel, there is elevated activity, but almost no significant encoding of articulator kinematics.

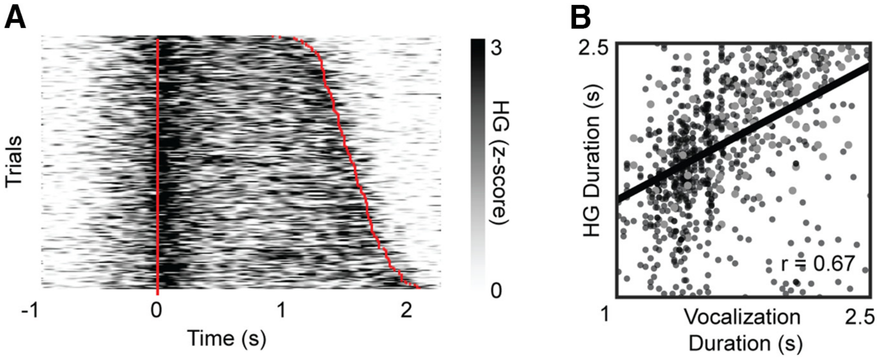

**Figure 8.** Sustained HG activity is related to vowel duration. ***A***, An example electrode that is characterized by sustained HG activity. Trials are ordered by vowel production duration. Red lines indicate the onset and offset of vowel acoustics. ***B***, The duration of the sustained HG activity at each trial is plotted against the trial duration for all sustained electrodes. Larger gray markers indicate observations from the example electrode in ***A***.

Furthermore, by characterizing the movements of the articulators according to a variety of kinematic parameters (position, speed, velocity, and acceleration), we demonstrated that neural activity encodes each of the examined parameters independent of one another. Previous studies examining arm movements using analogous parameters have also found significant encoding of these parameters (Georgopoulos et al., 1982, 1984; Ashe and Georgopoulos, 1994; Moran and Schwartz, 1999; Paninski et al., 2004). While we find electrodes that significantly encode each parameter examined, speed is by far the most robustly encoded parameter. Furthermore, the dominant kinematic parameter at individual electrodes was not significantly related to the articulator representation of those electrodes. The predominance of speed over other parameters is somewhat surprising; previous studies of the single-unit representation of kinematic parameters during arm reaching typically find that velocity and direction are the most commonly encoded parameter (Moran and Schwartz, 1999). Similar results were also observed in a recent ECoG study, which found that movement speed was predominately represented during arm reaching in humans (Hammer et al., 2016). The predominance of speed encoding was interpreted in the context of a model in which the summed activity of many velocity-tuned neurons with random directional tuning resembles speed tuning. Thus, it may be the case that individual vSMC neurons are actually representing mostly velocity, but the summed activity observed with ECoG electrodes reflects the magnitude of movement without direction (i.e., speed).

By studying vowels, we were able to examine the dynamics of kinematic encoding that are associated with movements to specific vocal tract articulators. We found that articulator kinematics were encoded around the time of movement onset and/or offset, but not while the vocal tract configuration was being held to maintain the vowel. Encoding of articulator kinematics only during movement onset and offset suggests that control of speech articulators is accomplished primarily through control of changes to the plant, rather than moment-to-moment maintenance of the vocal tract configuration. This is consistent with models of speech production that use changes to the plant as the primary mechanism by which sensorimotor cortex receives input from, and sends commands to, the vocal tract (Houde and Nagarajan, 2011; Tourville and Guenther, 2011). Furthermore, these dynamics have been observed in studies with analogous behavior from different body parts, including arm reaching. These studies have found individual neurons in motor cortex that exhibit transient firing patterns, where firing rates are high around movement onset and offset (Crammond and Kalaska, 1996; Shalit et al., 2012; Arce et al., 2013; Shadmehr, 2017).

We also found a subset of electrodes that exhibited sustained neural activity during the steady-state portion of the vowel, which was not correlated with any measured kinematic features. Instead, we found that the duration of the sustained activity correlated well with trial-by-trial vowel length. At a minimum, this suggests that this sustained activity covaries with whether the subject is vocalizing. One possibility is that sustained activity represents an articulatory parameter that has little variability in our task, such as respiration. However, a more intriguing possibility is that sustained activity may represent a nonspecific signal for holding the vocal tract configuration, which does not directly encode the articulatory kinematics, such as position. Such a signal combined with the onset/offset encoding of kinematics may provide sufficient information for encoding the observed behavior. Further studies using tasks with more variability in manner of articulation are necessary to resolve these possibilities.

It is important to emphasize that these analyses focus on the neural representation of the supralaryngeal articulators. While the movements of these articulators are critical to the production of vowels, the lower vocal tract (e.g., larynx, pharynx, and diaphragm) is also necessary to produce voiced sounds. It is likely that subregions of vSMC are involved in the control of the lower vocal tract (Brown et al., 2009; Bouchard et al., 2013; Conant et al., 2014), but the limitations of our vocal tract kinematic monitoring system and the lack of across-trial variability in pitch and intensity preclude a detailed examination of the representations of these articulators in the present experiment.

Further, we are not able to evaluate whether the activity we observed is due to feedforward signals originating in vSMC, or sensory feedback signals. Our models performed optimally at a neural-leading lag of ∼100 ms, implying that the representations we observed were driven more by feedforward activity. However, the relatively simple movements examined here exhibit temporal autocorrelation, which makes it difficult to dissociate feedforward activity from feedback. Examining speech tasks with faster, less stereotyped movements (e.g., naturally produced words or sentences) would make it possible to disentangle feedforward and feedback signals, and is an interesting and important future direction (Chang et al., 2013; Greenlee et al., 2013; Kingyon et al., 2015; Behroozmand et al., 2016; W. Li et al., 2016; Cao et al., 2017).

Finally, while we observed qualitatively similar results across patients regardless of the hemisphere in which the electrodes were implanted, 3 of the 4 participants had grid placements on the right (nondominant) hemisphere. It is presently unknown whether there are differences in the representation of articulator movements between left and right hemisphere, and the results presented here may not fully address the extent to which such differences exist.

We found that the representation of spoken vowels in vSMC is directly explained by the movements of speech articulators. The encoding of multiple kinematic parameters is present for the articulators, most prominently speed. Articulator kinematic encoding was primarily observed at the onset and offset of vowel production and not while the vowel was being held. Together, these findings provide insight into how neural activity in sensorimotor cortex results in the precise production of vowels. Future work will address how these encoding properties operate in the context of natural continuous speech.

## References

Alfonso PJ, Baer T (1982) Dynamics of vowel articulation. Lang Speech 25: 151–173. CrossRef

Arce FI, Lee JC, Ross CF, Sessle BJ, Hatsopoulos NG (2013) Directional information from neuronal ensembles in the primate orofacial sensorimotor cortex. J Neurophysiol 110:1357–1369. CrossRef Medline

Ashe J, Georgopoulos AP (1994) Movement parameters and neural activity in motor cortex and area 5. Cereb Cortex 6:590–600. Medline

Atal BS, Chang JJ, Mathews MV, Tukey JW (1978) Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. J Acoust Soc Am 63:1535–1553. CrossRef Medline

Behroozmand R, Oya H, Nourski KV, Kawasaki H, Larson CR, Brugge JF, Howard MA 3rd, Greenlee JD (2016) Neural correlates of vocal production and motor control in human Heschl's gyrus. J Neurosci 36:2302–2315. CrossRef Medline

Boersma P (2001) Praat, a system for doing phonetics by computer. Glot Int 5:341–345.

Bouchard KE, Chang EF (2014) Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. J Neurosci 34:12662–12677. CrossRef Medline

Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. Nature 495:327–332. CrossRef Medline

Bouchard KE, Conant DF, Anumanchipalli GK, Dichter B, Chaisanguanthum KS, Johnson K, Chang EF (2016) High-resolution, non-invasive imaging of upper vocal tract articulators compatible with human brain recordings. PLoS One 11:1–30. CrossRef Medline

Brown S, Laird AR, Pfordresher PQ, Thelen SM, Turkeltaub P, Liotti M (2009) The somatotopy of speech: phonation and articulation in the human motor cortex. Brain Cogn 70:31–41. CrossRef Medline

Cao L, Thut G, Gross J (2017) The role of brain oscillations in predicting self-generated sounds. Neuroimage 147:895–903. CrossRef Medline

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci U S A 110:2653–2658. CrossRef Medline

Conant D, Bouchard KE, Chang EF (2014) Speech map in the human ventral sensory-motor cortex. Curr Opin Neurobiol 24:63–67. CrossRef Medline

Crammond DJ, Kalaska JF (1996) Differential relation of discharge in primary motor cortex and premotor cortex to movements versus actively maintained postures during a reaching task. Exp Brain Res 108:45–61. Medline

Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis: II. Event-related synchronization in the gamma band. Brain 121: 2301–2315. CrossRef Medline

Crone NE, Hao L, Hart J Jr, Boatman D, Lesser RP, Irizarry R, Gordon B (2001) Electrocorticographic gamma activity during word production in spoken and sign language. Neurology 57:2045–2053. CrossRef Medline

Donoho D, Stodden V (2004) When does non-negative matrix factorization give a correct decomposition into parts? Proc Adv Neural Inf Process Syst 16:1141–1148.

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50:291–301. CrossRef Medline

Fukuda M, Rothermel R, Juhász C, Nishida M, Sood S, Asano E (2010) Cortical gamma-oscillations modulated by listening and overt repetition of phonemes. Neuroimage 49:2735–2745. CrossRef Medline

Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J Neurosci 2:1527–1537. Medline

Georgopoulos AP, Caminiti R, Kalaska JF (1984) Static spatial effects in motor cortex and area 5: quantitative relations in a two-dimensional space. Exp Brain Res 54:446–454. Medline

Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. Science 233:1416–1419. CrossRef Medline

Grabski K, Lamalle L, Vilain C, Schwartz JL, Vallée N, Tropres I, Baciu M, Le Bas JF, Sato M (2012) Functional MRI assessment of orofacial articulators: neural correlates of lip, jaw, larynx, and tongue movements. Hum Brain Mapp 33:2306–2321. CrossRef Medline

Gracco VL, Löfqvist A (1994) Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. J Neurosci 14:6585–6597. Medline

Greenlee JD, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H, Kawasaki H, Howard MA 3rd (2013) Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. PLoS One 8:e60783. CrossRef Medline

Hamilton LS, Edwards E, Chang EF (2017) Parallel streams define the temporal dynamics of speech processing across human auditory cortex. bioRxiv. CrossRef

Hammer J, Pistohl T, Fischer J, Kršek P, Tomášek M, Marusič P, Schulze-Bonhage A, Aertsen A, Ball T (2016) Predominance of movement speed over direction in neuronal population signals of motor cortex: intracranial EEG data and A simple explanatory model. Cereb Cortex 26:2863–2881. CrossRef Medline

Harshman R, Ladefoged P, Goldstein L (1977) Factor analysis of tongue shapes. J Acoust Soc Am 62:693–713. CrossRef Medline

Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G, Schultz T (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. Front Neurosci 9:217. CrossRef Medline

Hesselmann V, Sorger B, Lasek K, Guntinas-Lichius O, Krug B, Sturm V, Goebel R, Lackner K (2004) Discriminating the cortical representation sites of tongue and up movement by functional MRI. Brain Topogr 16: 159–167. Medline

Hillenbrand J, Getty LA, Clark MJ, Wheeler K (1995) Acoustic characteristics of American English vowels. J Acoust Soc Am 97:3099–3111. CrossRef Medline

Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front Hum Neurosci 5:82. CrossRef Medline

Huang CS, Sirisko MA, Hiraba H, Murray GM, Sessle BJ (1988) Organization of the primate face motor cortex as revealed by intracortical microstimulation and electrophysiological identification of afferent inputs and corticobulbar projections. J Neurophysiol 59:796–818. CrossRef Medline

Johnson K, Ladefoged P, Lindau M (1993) Individual differences in vowel production. J Acoust Soc Am 94:701–714. CrossRef Medline

Kellis S, Miller K, Thomson K, Brown R, House P, Greger B (2010) Decoding spoken words using local field potentials recorded from the cortical surface. J Neural Eng 7:56007–56016. CrossRef Medline

Kingyon J, Behroozmand R, Kelley R, Oya H, Kawasaki H, Narayanan NS, Greenlee JD (2015) High-gamma band fronto-temporal coherence as a measure of functional connectivity in speech motor control. Neuroscience 305:15–25. CrossRef Medline

Kumar V, Croxson PL, Simonyan K (2016) Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. J Neurosci 36:4170–4181. CrossRef Medline

Ladefoged P, Johnson K (2011) A course in phonetics. Boston: Cengage Learning.

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791. CrossRef Medline

Leonard MK, Cai R, Babiak MC, Ren A, Chang EF (2016) The peri-Sylvian cortical network underlying single word repetition revealed by electrocortical stimulation and direct neural recordings. Brain Lang. Advance online publication. Retrieved Jul. 19, 2016. doi: 10.1016/j.bandl.2016.06.001. CrossRef Medline

Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J, Schalk G (2011) Using the electrocorticographic speech network to control a brain-computer interface in humans. J Neural Eng 8:036004. CrossRef Medline

Li M, Kambhamettu C, Stone M (2005) Automatic contour tracking in ultrasound images. Clin Linguist Phon 19:545–554. Medline

Li W, Chen Z, Yan N, Jones JA, Guo Z, Huang X (2016) Temporal lobe epilepsy alters auditory-motor integration for voice control. Sci Rep 6:1–13. CrossRef Medline

Löfqvist A, Gracco VL (1999) Interarticulator programming in VCV sequences: lip and tongue movements. J Acoust Soc Am 105:1864–1876. CrossRef Medline

Lotze M, Seggewies G, Erb M, Grodd W, Birbaumer N (2000) The representation of articulation in the primary sensorimotor cortex. Neuroreport 11:2985–2989. CrossRef Medline

Maeda S (1990) Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: Speech production and speech modeling, pp 131–149. Rotterdam, The Netherlands: Springer.

Moran DW, Schwartz AB (1999) Motor cortical representation of speed and direction during reaching. J Neurophysiol 82:2676–2692. CrossRef Medline

Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, Shih JJ, Krusienski DJ, Slutzky MW (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng 11:35015. CrossRef Medline

Muller L, Hamilton LS, Edwards E, Bouchard KE, Chang EF (2016) Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. J Neural Eng 13:056013. CrossRef Medline

Paninski L, Fellows MR, Hatsopoulos NG, Donoghue JP (2004) Spatiotemporal tuning of motor cortical neurons for hand position and velocity. J Neurophysiol 91:515–532. CrossRef Medline

Pei X, Barbour DL, Leuthardt EC, Schalk G (2011) Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J Neural Eng 8:046028. CrossRef Medline

Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain 60:389–443. CrossRef

Penfield W, Roberts L (1959) Speech and brain mechanisms. Princeton, NJ: Princeton UP.

Perkell JS, Nelson WL (1985) Variability in production of the vowels /i/ and /a/. Acoust Soc Am 77:1889–1895. CrossRef Medline

Perkell JS, Matthies ML, Svirsky MA, Jordan MI (1993) Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: a pilot "motor equivalence" study. J Acoust Soc Am 93:2948–2961. CrossRef Medline

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988) Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331:585–589. CrossRef Medline

Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol 9:e1000610. CrossRef Medline

Shadmehr R (2017) Distinct neural circuits for control of movement vs holding still. J Neurophysiol 117:1431–1460. CrossRef Medline

Shalit U, Zinger N, Joshua M, Prut Y (2012) Descending systems translate transient cortical commands into a sustained muscle activation signal. Cereb Cortex 22:1904–1914. CrossRef Medline

Simonyan K, Ackermann H, Chang EF, Greenlee JD (2016) New developments in understanding the complexity of human speech production. J Neurosci 36:11440–11448. CrossRef Medline

Tourville JA, Guenther FH (2011) The DIVA model: a neural theory of speech acquisition and production. Lang Cogn Process 26:952–981. CrossRef Medline

Ueda Y, Hamakawa T, Sakata T, Hario S, Watanabe A (2007) A real-time formant tracker based on the inverse filter control method. Acoust Sci Technol 28:271–274. CrossRef

Watanabe A (2001) Formant estimation method using inverse-filter control. IEEE Trans Speech Audio Process 9:317–326. CrossRef