doi:10.1038/nature11911

Functional organization of human sensorimotor cortex for speech articulation

Kristofer E. Bouchard[1,2], Nima Mesgarani[1,2], Keith Johnson[3] & Edward F. Chang[1,2,4]

Speaking is one of the most complex actions that we perform, but nearly all of us learn to do it effortlessly. Production of fluent speech requires the precise, coordinated movement of multiple articulators (for example, the lips, jaw, tongue and larynx) over rapid time scales. Here we used high-resolution, multi-electrode cortical recordings during the production of consonant-vowel syllables to determine the organization of speech sensorimotor cortex in humans. We found speech-articulator representations that are arranged somatotopically on ventral pre- and post-central gyri, and that partially overlap at individual electrodes. These representations were coordinated temporally as sequences during syllable production. Spatial patterns of cortical activity showed an emergent, population-level representation, which was organized by phonetic features. Over tens of milliseconds, the spatial patterns transitioned between distinct representations for different consonants and vowels. These results reveal the dynamic organization of speech sensorimotor cortex during the generation of multi-articulator movements that underlies our ability to speak.

Speech communication critically depends on the ability to produce the large number of sounds that compose a given language[1,2]. The wide range of spoken sounds results from highly flexible configurations of the vocal tract, which filters sound produced at the larynx through movements of the lips, jaw and tongue that are coordinated precisely[3–5]. Each articulator has extensive degrees of freedom, making a large number of different speech movements possible. How humans exert such precise control despite the wide variety of movement possibilities is a central unanswered question[1,6,7].

The cortical control of articulation is mediated primarily by the ventral half of the lateral sensorimotor (Rolandic) cortex (ventral sensorimotor cortex, vSMC)[8–10], which provides corticobulbar projections to, and afferent innervation from, the face and vocal tract (Fig. 1a, b)[11,12]. The U-shaped vSMC is composed of the pre- and post-central gyri (Brodmann areas 1, 2, 3 and 6b), and the gyral area directly ventral to the termination of the central sulcus called the guenon (Brodmann area 43) (Fig. 1a, b)[13]. Using electrical stimulation, Foerster and Penfield described the somatotopic organization of face and mouth representations in human vSMC[14,15,16]. However, focal stimulation could not evoke meaningful utterances, implying that speech is not stored in discrete cortical areas. Instead, the production of phonemes and syllables is thought to arise from a coordinated motor pattern involving multiple articulator representations[1,3,4,5,9].

To understand the functional organization of vSMC in articulatory sensorimotor control, we recorded neural activity directly from the cortical surface in three human subjects implanted with high-density multi-electrode arrays as part of their preparation for epilepsy surgery (Fig. 1a). Intracranial cortical recordings were synchronized with microphone recordings as subjects read aloud consonant-vowel syllables (19 consonants followed by /a/, /u/ or /i/; Supplementary Fig. 1) that are commonly used in American English. This task was designed to sample across a range of phonetic features, including different constriction locations (place of articulation) and different constriction degrees or shapes (manner of articulation) for a given articulatory organ[17,18,19].

vSMC physiology during syllable production

We aligned cortical recordings to acoustic onsets of consonant-to-vowel transitions ($t = 0$) to provide a common reference point across consonant-vowel syllables (Fig. 1c–e). We focused on the high-gamma frequency component of local field potentials (85–175 Hz)[20,21,22], which correlates well with multi-unit firing rates[23]. For each electrode, we normalized the time-varying high-gamma amplitude to baseline statistics by transforming to *z*-scores.

During syllable articulation, approximately 30 active vSMC electrode sites were identified per subject (approximately 1,200 mm^2, change in *z*-score of greater than 2 for any syllable). Cortical activity from selected electrodes distributed along the vSMC dorsoventral axis is shown for /ba/, /da/ and /ga/ (Fig. 1c–e, same colouring as in Fig. 1a). The plosive consonants (/b/, /d/, /g/) are produced by transient occlusion of the vocal tract by the lips, front tongue and back tongue, respectively, whereas the vowel /a/ is produced by a low, back tongue position during phonation. Dorsally located electrodes (for example, Fig. 1c–e, electrodes 124 and 108; black) were active during production of /b/, which requires transient closure of the lips. In contrast, mid-positioned electrodes (for example, electrodes 129, 133 and105; grey) were active during production of /d/, which requires forward tongue protrusion against the alveolar ridge. A more ventral electrode (for example, electrode 104; red) was most active during production of /g/, which requires a posterior-oriented tongue elevation towards the soft palate. Other electrodes appear to be active during the vowel phase for /a/ (for example, electrodes 154, 136 and 119).

Cortical activity at different electrode subsets was superimposed to visualize spatiotemporal patterns across other phonetic contrasts. Consonants produced with different constriction locations of the tongue tip, (for example, /θ/ (dental), /s/ (alveolar), and /ʃ/ (post-alveolar)), showed specificity across different electrodes in central vSMC (Fig. 1f), although they were not as categorical as those shown for consonants involving different articulators in Fig. 1c–e. Consonants with similar tongue constriction locations, but different constriction degree or constriction shape, were generated by overlapping electrode sets exhibiting

[1]Department of Neurological Surgery and Department of Physiology, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, California 94143, USA. [2]Center for Integrative Neuroscience, 675 Nelson Rising Lane, University of California, San Francisco, California 94158, USA. [3]Department of Linguistics, University of California, Berkeley, 1203 Dwinelle Hall, Berkeley, California 94720, USA. [4]UCSF Epilepsy Center, University of California, San Francisco, 400 Parnassus Avenue, San Francisco, California 94143, USA.

different relative activity magnitudes (Fig. 1g, /l/ (lateral) versus /n/ (nasal stop) versus /d/ (oral stop)). Syllables with the same consonant followed by different vowels (Fig. 1h, /ja/, /ji/, /ju/) were found to have similar activity patterns before the consonant-vowel transition. During vowel phonation, a dorsal electrode is clearly active during /u/, but not / i/ or /a/ (Fig. 1h, /ju/; black arrow) whereas another electrode in the middle of vSMC had prolonged activity during /i/ and /u/ vowels compared to /a/ (Fig. 1h, /ji/ and /ju/; purple arrows). These contrasting examples show that important phonetic properties can be observed qualitatively from the rich repertoire of vSMC spatiotemporal patterns.

Spatial representation of articulators

To determine the spatial organization of speech-articulator representations, we examined how cortical activity at each electrode depended on the movement of a given articulator (using a general linear model). We assigned binary variables to four articulatory organs (lips, tongue, larynx and jaw) that are used in producing the consonant component of each consonant-vowel syllable (Supplementary Fig. 1). The spatial distribution of optimal weightings for these articulators (averaged over time and subjects) were plotted as a function of dorsoventral distance from the Sylvian fissure and anteroposterior distance from the central sulcus. We found representations for each articulator distributed across vSMC (Fig. 2a). For example, the lip representation was localized to the dorsal aspect of vSMC, whereas the tongue representation was distributed more broadly than the lip representation across the ventral aspect.

To determine topographic organization of articulators across subjects, we extracted the greatest 10% of weightings from individual articulator distributions (Fig. 2a) and used a clustering algorithm (*k*-nearest neighbour) to classify the surrounding cortex (Fig. 2b). We found an overall somatotopic dorsoventral arrangement of articulator representations laid out in the following sequence: larynx, lips, jaw, tongue and larynx (Fig. 2a, b and Supplementary Figs 2–5). An analysis of the fractional representation of all articulators at single electrodes showed a clear tuning preference for individual articulators at single electrodes and also demonstrated that single electrodes had functional representations of multiple articulators (Supplementary Fig. 6).

Timing of articulator representations

As the time course of articulator movements is on the scale of tens of milliseconds, previous approaches have been unable to resolve temporal properties associated with individual articulator representations. We examined the timing of correlations between cortical activity

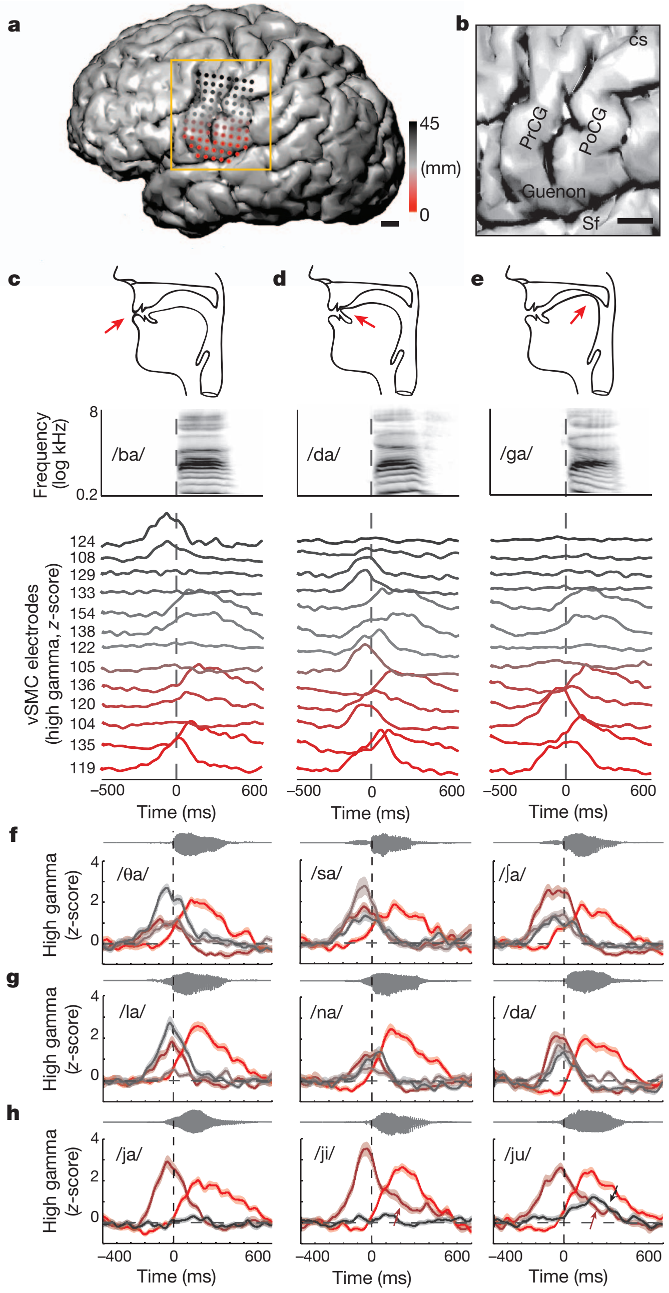

Figure 1 | vSMC physiology during syllable production. a, Magnetic resonance imaging (MRI) reconstruction of a single subject brain with vSMC electrodes (dots), coloured according to distance from the Sylvian fissure (black and red are the most dorsal and ventral positions, respectively). **b**, Expanded view of vSMC anatomy. cs, central sulcus; PoCG, post-central gyrus; PrCG, pre-central gyrus; Sf, Sylvian fissure. Scale bars, 1 cm. **c–e**, Top, vocal tract schematics for three consonants (/b/, /d/, /g/), produced by occlusion at the lips, tongue tip and tongue body, respectively (red arrow). Middle, spectrograms of spoken consonant-vowel syllables. Bottom, average cortical activity from a subset of electrodes (electrode number on far right, same colouring as in **a**). Vertical dashed line, acoustic onset of consonant-vowel transition. **f–h**, Cortical activity at selected electrodes for different phonetic contrasts (mean ± s.e.m.). Acoustic waveforms are displayed above. **f**, Fricatives (/θ/('th' of 'thin'), /s/, /ʃ/('sh' of 'shin')) with different constriction locations. **g**, Front tongue consonants (/l/, /n/, /d/) with different constriction degree or shapes. **h**, Single consonant (/j/ ('y' of 'yes')) with different vowels (/a/, /i/, /u/). Purple arrows correspond to a tongue electrode with prolonged activity for /i/ and /u/ vowels. Black arrow corresponds to an active lip electrode for /u/.

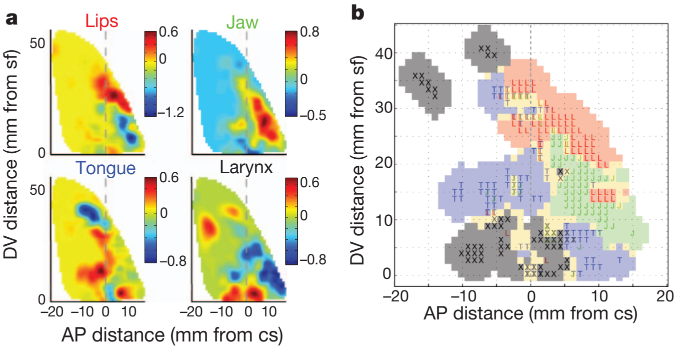

Figure 2 | Spatial representation of articulators. a, Localization of lips, jaw, tongue and larynx representations. Average magnitude of articulator weightings (colour scale) plotted as a function of anteroposterior (AP) distance from the central sulcus and dorsoventral (DV) distance from the Sylvian fissure ($n = 3$ subjects). **b**, Functional somatotopic organization of speech-articulator representations in vSMC. Lips (L, red); jaw (J, green); tongue (T, blue); larynx (X, black); mixed (yellow). Letters correspond to locations, based on direct measurement-derived regression weights; shaded rectangles correspond to regions classified by *k*-nearest neighbour.

and specific consonant articulators (using partial correlation analysis), and included two vowel articulatory features (back tongue and high tongue; Supplementary Fig. 1).

Time courses of correlations were plotted for electrodes with highest values, sorted by onset latency (Fig. 3a). We found that jaw, high tongue and back tongue had very consistent timing across electrodes. Similar results were found for tongue, lips and larynx, but with more variable latencies. Timing relationships between articulator representations were staggered, reflecting a temporal organization during syllable production: lip and tongue correlations began well before sound onset (Fig. 3a, c, d); jaw and larynx correlations were aligned to the consonant-vowel transition (Fig. 3a, c, d); and high tongue and back tongue features showed high temporal specificity for the vowel phase, peaking near the acoustic mid-point of the vowels (approximately 250 ms, Fig. 3b–d). This sequence of articulator correlations was consistent across subjects (Fig. 3d, $P < 10^{-10}$, analysis of variance (ANOVA), $F = 40$, d.f. = 5, $n = 211$ electrodes from 3 subjects) and is in accordance with the timing of articulator movements shown in speech-kinematics studies[3,5,17,24]. We found no statistically significant onset-latency differences in those areas 10 mm anterior and posterior to the central sulcus) or across the guenon ($P > 0.4$, rank-sum test; $n = 71$ and $n = 67$, respectively; Supplementary Fig. 7). This is consistent with mixed sensory and motor orofacial responses throughout vSMC, which are also seen in stimulation experiments[14,25].

Phonetic organization of spatial patterns

The distributed organization of speech articulator representations (Fig. 2) led us to propose that coordination of the multiple articulators required for speech production would be associated with spatial patterns of cortical activity. We refer here to this population-derived pattern as the phonetic representation. To determine its organizational properties, we used principal component analysis to transform the observed cortical activity patterns into a 'cortical state-space' (approximately 60% of variance is explained by 9 spatial principal components for all subjects, Supplementary Figs 8 and 9)[26–30]. k-means clustering during the consonant phase (25 ms before the consonant-vowel transition, $t = -25$ ms) showed that the cortical state-space was organized into three clusters (quantified by silhouette analysis) corresponding to the major oral articulators: labial, coronal tongue, and dorsal tongue (Fig. 4a and Supplementary Fig. 10).

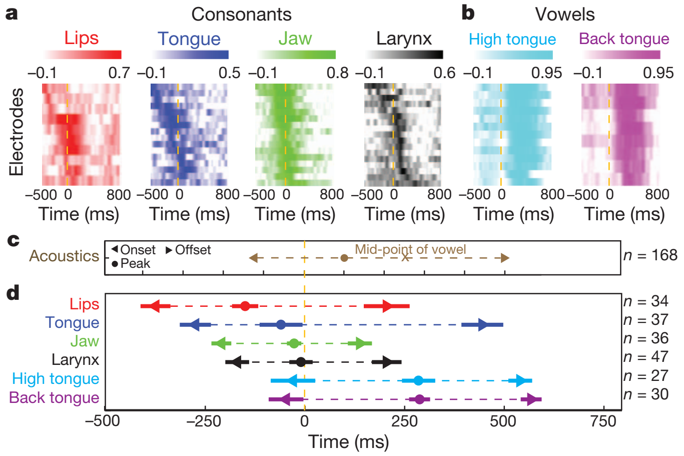

Figure 3 | Temporal representation of articulators. a, **b**, Timing of correlations between cortical activity and consonant (**a**) and vowel (**b**) articulator features. Colour maps display correlation coefficients (R) for a subset of electrodes. **c**, Acoustic landmarks. Onset (end of arrows, left), peak power (shown by a dot in each case) and offset (end of arrows, right) for consonant-vowel syllables (mean ± s.e.m., $n = 168$ syllables, all subjects). Error bars are smaller than the symbols. **d**, Temporal sequence and range of correlations. Symbols are as in **c**. Data are mean (symbols) ± s.e.m. (thick solid line) across electrodes from all subjects.

During the vowel phase (250 ms after the consonant-vowel transition, $t = 250$), we found clear separation of /a/, /i/ and /u/ vowel states (Fig. 4b). Similar clustering of consonants and vowels was found across subjects ($P < 10^{-10}$ for clustering of both consonants and vowels, Supplementary Fig. 11).

Theories of speech motor control and phonology have speculated that there is a hierarchical organization of phoneme representations, given the anatomical and functional dependencies of the vocal tract articulators during speech production[3,4,17,18,31]. To evaluate such organization in vSMC, we applied hierarchical clustering to the cortical state-space (Fig. 4c, d). For consonants, this analysis confirmed that the primary tier of organization was defined by the major oral articulator features: dorsal, labial or coronal (Fig. 4c). These major articulators were superordinate to the constriction location within each articulator. For example, the labial cluster could be subdivided

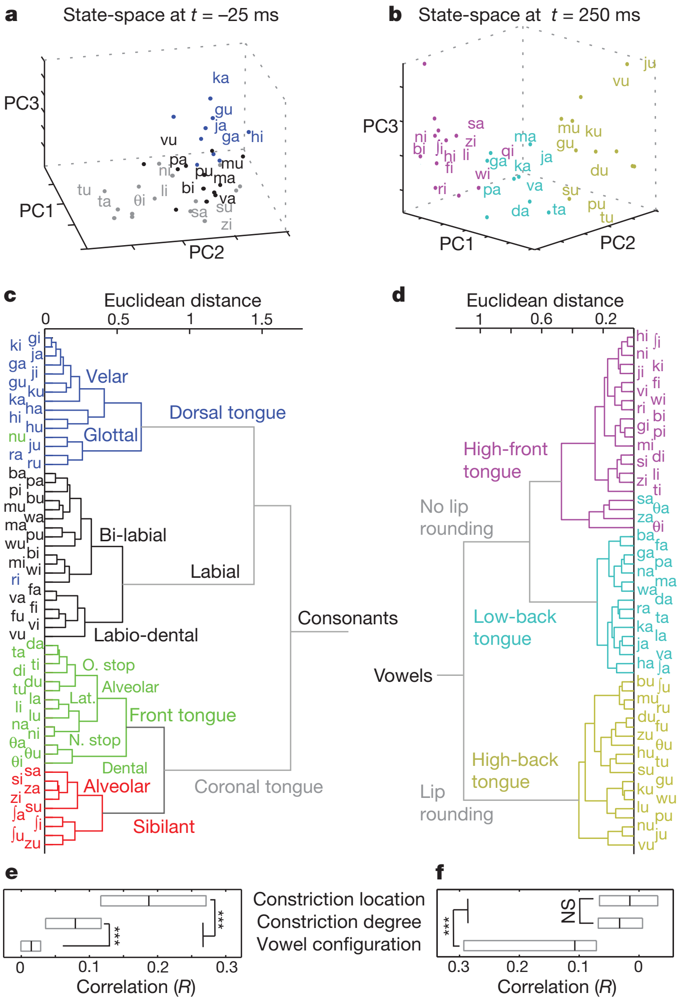

Figure 4 | Phonetic organization of spatial patterns. a, **b**, Scatterplots of consonant-vowel syllables in the first three principal components for consonants (25 ms before consonant-vowel transition) (**a**) and vowels (250 ms after consonant-vowel transition) (**b**). A subset of consonant-vowels are labelled with international phonetic alphabet (IPA) symbols, all others have dots. Colouring denotes k-means cluster membership. **c**, **d**, Hierarchical clustering of cortical state-space at consonant (25 ms before consonant-vowel transition) and vowel time points (250 ms after consonant-vowel transition). Individual syllables and dendrogram branches are colour-coded and labelled by known linguistic categories using the same colour scheme as in Fig 4 a, b, with new subdivisions of the coronal tongue into front tongue and sibilant (in green and red, respectively). Lat, lateral; N. stop, nasal stop; O. stop, oral stop. **e**, **f**, Correlations between cortical state-space and phonetic features. Black vertical lines, medians; grey boxes, 25th and 75th percentiles. ***$P < 10^{-10}$, WSRT; $n = 297$ for both consonants and vowels. NS, not significant.

into bi-labial and labiodental. Only at the lowest level of the hierarchy did we observe suggestions of organization according to constriction degree or shape, such as the sorting of nasal (/n/ syllables), oral stops (/d/, /t/) and lateral approximants (/l/). Similarly, during the vowel period, a primary distinction was based on the presence or absence of lip rounding (/u/ versus /a/ and /i/), and a secondary distinction was based on tongue posture (height, and front or back position) (Fig. 4d). Therefore, the major oral articulator features that organize consonant representations are similar to those for vowels.

Across an early time period (375 ms before, to 120 ms after, the consonant-vowel transition), we found that consonant features describing constriction location had a significantly greater correlation with the cortical state-space than constriction degree, which in turn was significantly more correlated than the upcoming vowel ($P < 10^{-10}$, Wilcoxon signed-rank test (WSRT), $n = 297$ from 3 subjects; see Supplementary Fig. 12 for phonetic feature sets). This analysis shows that constriction location accounts for more of the structure of spatial activity patterns than does constriction degree or shape. Similarly, across a later time period (125 ms to 620 ms after the consonant-vowel transition), we found that vowel features provided the greatest correlation (vowel configuration versus other feature sets, $P < 10^{-10}$, WSRT, $n = 297$ from 3 subjects).

Dynamics of phonetic representations

The dynamics of neural populations have provided insights into the structure and function of many neural circuits[6,26,27,29,32,33]. To determine the dynamics of phonetic representations, we investigated how state-space trajectories for consonants and vowels entered and departed target regions for phonetic clusters. Trajectories of individual consonant-vowel syllables were visualized by plotting their locations in the first two principal-component dimensions versus time (Fig. 5a, b; principal component 1 (PC1) and PC2 for one of the subjects).

We examined first how trajectories of different consonants transitioned to a single vowel, /u/ (Fig. 5a). The cortical state-space was initially unstructured, and then individual trajectories converged within phonetic clusters (for example, labial, front tongue, dorsal tongue and sibilant), and at the same time trajectories for different clusters diverged from one another. These convergent and divergent dynamics gradually increased the separability of different phonetic clusters (the mean difference of between-cluster and within-cluster distances). Later, as each consonant transitioned to /u/, trajectories converged to a compact target region for the vowel. Finally, trajectories diverged randomly, presumably as articulators returned to neutral position. Analogous dynamics were observed during the production of a single consonant cluster (for example, labials) transitioning to different vowels (/a/, /i/ and /u/) (Fig. 5b).

We quantified the internal dynamical properties of the cortical state-space by calculating cluster separability. The time course of cluster separability, averaged across subjects and consonant-vowel syllables (Fig. 5c) showed that separability peaked approximately 200 ms before the consonant-vowel transition for consonants (onset, approximately 300 ms before the consonant-vowel transition), and at 250 ms after the consonant-vowel transition for vowels (onset, approximately 50 ms after the consonant-vowel transition). We examined further the dynamics of correlations between the structure of the cortical state-space and phonetic features (averaged across subjects) (plotted in Fig. 5d). Across subjects, we found that cluster separability and the correlation between cortical state-space organization and phonetic features were tightly linked for both consonants and vowels in a time-dependent fashion (R^2 range $= 0.42$–0.98, $P < 10^{-10}$ in all cases). This shows that the dynamics of clustering in the cortical state-space is coupled strongly to the degree to which the cortical state reflects the phonetic structure of the vocalization.

Visualization of the dynamic structure of the cortical state-space during production of all consonant-vowel syllables (Fig. 5e) showed that, as the cortical state comes to reflect phonetic structure, different phonetic clusters diverge from one another, while the trajectories within the clusters converge. Furthermore, we observed correlates of the earlier articulatory specification for sibilants (/ʃ/, /z/, /s/). In addition, with all consonant-vowel syllables on the same axes, we observed that in comparison to vowels, consonants occupy a distinct region of cortical state-space, despite sharing the same articulators. The distribution of state-space distances was significantly greater in consonant-vowel comparisons than in consonant-consonant or vowel-vowel comparisons ($P < 10^{-10}$ for all comparisons, WSRT, $n = 4623$ in all cases, Supplementary Fig. 11). Finally, the consonant-to-vowel sequence reveals a periodic structure, which is sub-specified for consonant and vowel features.

Discussion

Our broad-coverage, high-resolution direct cortical recordings enabled us to examine the spatial and temporal profiles of speech articulator representations in human vSMC. Cortical representations are

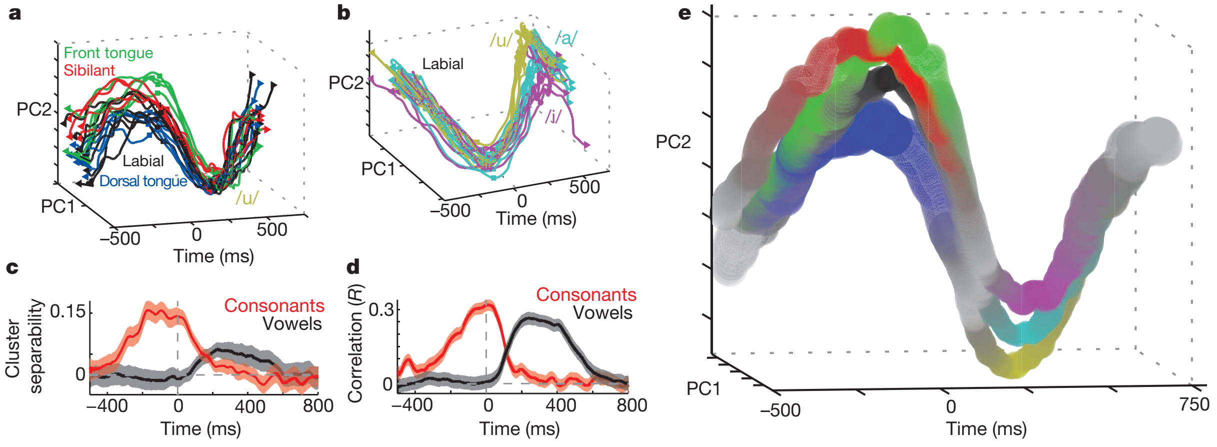

Figure 5 | Dynamics of phonetic representations. a, **b**, Cortical state-space trajectories. **a**, Consonants transitioning to the vowel /u/. Each line corresponds to a single consonant-vowel trajectory. For each line, the left triangle indicates $t = -500$ ms, the square indicates $t = -25$ ms, the circle indicates $t = 250$ ms, and the right triangle indicates $t = 750$ ms. **b**, Trajectories of the labial consonants transitioning to /a/, /i/ and /u/. **c**, **d**, Across-subject averages of cluster separability (**c**) and correlation between cortical state-space structure and phonetic features (**d**) for consonants and vowels (mean ± s.e.m.). **e**, Time-course of consonant-vowel syllable trajectories for one subject. Each colour corresponds to one of the consonant or vowel groups (colours are the same as in **a** and **b** above). The centre of each coloured tube is located at the centroid of the corresponding phonetic cluster. Tube diameter corresponds to cluster density and colour saturation represents the correlation between the structure of the cortical state-space and phonetic features.

somatotopically organized, with individual sites tuned for a preferred articulator and co-modulated by other articulators. The dorsoventral layout of articulator representations recapitulates the rostral-to-caudal layout of the vocal tract. However, we found an additional laryngeal representation located at the dorsal-most end of vSMC[8,10,34,35]. This dorsal laryngeal representation seems to be absent in non-human primates[11,36,37], suggesting a unique feature of human vSMC for the specialized control of speech. Pre- and post-central gyrus neural activity occurred before vocalization, which may reflect the integration of motor commands with proprioceptive information for rapid feedback control during speaking[9,38–43].

Just as focal stimulation is insufficient to evoke speech sounds, it is not any single articulator representation, but the coordination of multiple articulator representations across the vSMC network that generates speech. Analysis of spatial patterns of activity showed an emergent hierarchy of network states that organizes phonemes by articulatory features. This functional hierarchy of network states contrasts with the anatomical hierarchy often considered in motor control[44]. The cortical state-space organization probably reflects the coordinative patterns of articulatory motions during speech, and is notably similar to a theorized cross-linguistic hierarchy of phonetic features ('feature geometry')[3,18,31,45]. In particular, the findings support gestural theories of speech control[3] over alternative acoustic (a hierarchy organized primarily by constriction degree)[19] or vocal-tract geometry theories (no hierarchy of constriction location and degree)[18].

The vSMC population showed convergent and divergent dynamics during the production of different phonetic features. The dynamics of individual phonemes were superimposed on a slower oscillation that characterizes the transition between consonants and vowels. Although trajectories were found to originate or terminate in different regions, they consistently pass through the same (target) region of the state-space for shared phonetic features[46]. Consonants and vowels occupy distinct regions of the cortical state-space. Large state-space distances between consonant and vowel representations may explain why it is more common in speech errors to substitute consonants with one another, and vowels with vowels, but very rarely consonants with vowels or vowels with consonants (that is, in 'slips of the tongue')[47].

We have shown that a relatively small set of articulator representations can combine flexibly to create the large variety of speech sounds in American English. The major organizational features found here define phonologies of languages from across the world[31]. Consequently, these cortical organizational principles are likely to be conserved, with further specification for unique articulatory properties across different languages.

METHODS SUMMARY

Three subjects underwent surgical placement of subdural arrays as part of their clinical treatment for epilepsy (see Supplementary Table 1 for clinical details). Statistical tests were considered significant if the Bonferroni corrected rate of incorrectly rejecting the null hypothesis was less than 0.05.

Full Methods and any associated references are available in the online version of the paper.

Received 23 June 2012; accepted 15 January 2013.
Published online 20 February 2013.

1. Levelt, W. J. M. *Speaking: From Intention to Articulation* (MIT Press, 1993).
2. Ladefoged, P. & Johnson, K. *A Course in Phonetics* (Wadsworth Publishing, 2010).
3. Browman, C. P. & Goldstein, L. Articulatory gestures as phonological units. *Haskins Laboratories Status Report on Speech Research* **99,** 69–101 (1989).
4. Fowler, C. A., Rubin, P. E., Remez, R. E. & Turvey, M. T. in *Language Production: Speech and Talk* Vol. 1 (ed. Butterworth, B.) 373–420 (Academic Press, 1980).
5. Gracco, V. L. & Lofqvist, A. Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *J. Neurosci.* **14,** 6585–6597 (1994).
6. Schöner, G. & Kelso, J. A. Dynamic pattern generation in behavioral and neural systems. *Science* **239,** 1513–1520 (1988).
7. Franklin, D. W. & Wolpert, D. M. Computational mechanisms of sensorimotor control. *Neuron* **72,** 425–442 (2011).
8. Brown, S. *et al.* The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain Cogn.* **70,** 31–41 (2009).
9. Guenther, F. H., Ghosh, S. S. & Tourville, J. A. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* **96,** 280–301 (2006).
10. Schulz, G. M., Varga, M., Jeffires, K., Ludlow, C. L. & Braun, A. R. Functional neuroanatomy of human vocalization: an H215O PET study. *Cereb. Cortex* **15,** 1835–1847 (2005).
11. Jürgens, U. Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26,** 235–258 (2002).
12. Kuypers, H. G. Corticobular connexions to the pons and lower brain-stem in man: an anatomical study. *Brain* **81,** 364–388 (1958).
13. Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* (Smith-Gordon, 1994).
14. Penfield, W. & Boldrey, E. Somatic motor and sensory representation in the cerebral cortex of man studied by electrical stimulation. *Brain* **60,** 389–443 (1937).
15. Foerster, O. The cerebral cortex in man. *Lancet* **221,** 309–312 (1931).
16. Penfield, W. & Roberts, R. *Speech and Brain: Mechanisms.* (Princeton, 1959).
17. Saltzman, E. & Munhall, K. A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* **1,** 333–382 (1989).
18. Clements, G. N. & Hume, E. in *The Handbook of Phonological Theory* (ed. Goldsmith, J. A.) 245–306 (Basil Blackwell, 1995).
19. Chomsky, N. & Halle, M. *The Sound Pattern of English* (MIT Press, 1991).
20. Mesgarani, N. & Chang, E. F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485,** 233–236 (2012).
21. Crone, N. E., Miglioretti, D. L., Gordon, B. & Lesser, R. P. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* **121,** 2301–2315 (1998).
22. Edwards, E. *et al.* Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* **50,** 291–301 (2010).
23. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9,** e1000610 (2011).
24. Kent, R. D. in *The Production of Speech* (ed. MacNeilage, P. F.) (Springer-Verlag, 1983).
25. McCarthy, G., Allison, T. & Spencer, D. D. Localization of the face area of human sensorimotor cortex by intracranial recording of somatosensory evoked potentials. *J. Neurosurg.* **79,** 874–884 (1993).
26. Afshar, A. *et al.* Single-trial neural correlates of arm movement preparation. *Neuron* **71,** 555–564 (2011).
27. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48,** 661–673 (2005).
28. Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63,** 544–557 (2009).
29. Ahrens, M. B. *et al.* Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485,** 471–477 (2012).
30. Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I. & Shenoy, K. V. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* **68,** 387–400 (2010).
31. McCarthy, J. J. Feature geometry and dependency: a review. *Phonetica* **45,** 84–108 (1988).
32. Briggman, K. L. & Kristan, W. B. Multifunctional pattern-generating circuits. *Annu. Rev. Neurosci.* **31,** 271–294 (2008).
33. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487,** 51–56 (2012).
34. Brown, S., Ngan, E. & Liotti, M. A larynx area in the human motor cortex. *Cereb. Cortex* **18,** 837–845 (2008).
35. Terumitsu, M., Fujii, Y., Suzuki, K., Kwee, I. L. & Nakada, T. Human primary motor cortex shows hemispheric specialization for speech. *Neuroreport* **17,** 1091–1095 (2006).
36. Hast, M. H., Fischer, J. M., Wetzel, A. B. & Thompson, V. E. Cortical motor representation of the laryngeal muscles in Macaca mulatta. *Brain Res.* **73,** 229–240 (1974).
37. Jürgens, U. On the elicitability of vocalization from the cortical larynx area. *Brain Res.* **81,** 564–566 (1974).
38. Pruszynski, J. A. *et al.* Primary motor cortex underlies multi-joint integration for fast feedback control. *Nature* **478,** 387–390 (2011).
39. Hatsopoulos, N. G. & Suminski, A. J. Sensing with the motor cortex. *Neuron* **72,** 477–487 (2011).
40. Tremblay, S., Shiller, D. M. & Ostry, D. J. Somatosensory basis of speech production. *Nature* **423,** 866–869 (2003).
41. Matyas, F. *et al.* Motor control by sensory cortex. *Science* **330,** 1240–1243 (2010).
42. Rathelot, J. A. & Strick, P. L. Muscle representation in the macaque motor cortex: an anatomical perspective. *Proc. Natl Acad. Sci. USA* **103,** 8257–8262 (2006).
43. Gracco, V. L. & Abbs, J. H. Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *J. Neurophysiol.* **54,** 418–432 (1985).
44. Sherrington, C. S. *The Integrative Action of the Nervous System* (Yale University Press, 1911).
45. Jakobson, R., Fant, G. & Halle, M. *Preliminaries to speech analysis: the distinctive features and their correlates* (MIT Press, 1969).
46. Keating, P. A. *The Window Model of Coarticulation: Articulatory Evidence* (Cambridge Univ. Press, 1990).
47. Dell, G. S., Juliano, C. & Govindjee, A. Structure and content in language production: a theory of frame constraints in phonological speech errors. *Cogn. Sci.* **17,** 149–195 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Ren for technical help with data collection and pre-processing, and M. Babiak for audio transcription. J. Houde, C. Niziolek, S. Lisberger, K. Chaisanguanthum, C. Cheung and I. Garner provided helpful comments on the manuscript. E.F.C. was funded by the US National Institutes of Health grants R00-NS065120, DP2-OD00862 and R01-DC012379, and the Ester A. and Joseph Klingenstein Foundation.

Author Contributions E.F.C. conceived and collected the data for this project. K.E.B. designed and implemented the analysis with assistance from E.F.C. N.M. assisted with preliminary analysis. K.E.B. and E.F.C. wrote the manuscript. K.J. provided phonetic consultation on experimental design and interpretation of results. E.F.C. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.F.C. (ChangEd@neurosurg.ucsf.edu).

METHODS

The experimental protocol was approved by the Human Research Protection Program at the University of California, San Francisco.

Subjects and experimental task. Three native-English-speaking human subjects underwent chronic implantation of a high-density, subdural electrocortigraphic (ECoG) array over the left hemisphere (two subjects) or right hemisphere (one subject) as part of their clinical treatment of epilepsy (see Supplementary Table 1 for clinical details)[48]. Subjects gave their written informed consent before the day of surgery. All subjects had self-reported normal hearing and underwent neuropsychological language testing (including the Boston naming and verbal fluency tests) and scored within the range considered normal. Each subject read aloud consonant-vowel syllables composed of 18 or 19 consonants (19 consonants for two subjects, 18 consonants for one subject), followed by one of three vowels. Each consonant-vowel syllable was produced between 15 and 100 times. Microphone recordings were synchronized with the multi-channel ECoG data.

Data acquisition and signal processing. Cortical local field potentials (LFPs) were recorded with ECoG arrays and a multi-channel amplifier connected optically to a digital signal processor (Tucker-Davis Technologies). The spoken syllables were recorded with a microphone, amplified digitally, and recorded simultaneously with the ECoG data. ECoG signals were acquired at 3,052 Hz.

The time series from each channel was inspected visually and quantitatively for artefacts or excessive noise (typically 60 Hz line noise). These channels were excluded from all subsequent analysis and the raw recorded ECoG signal of the remaining channels were then common-average referenced and used for spectro-temporal analysis. For each (useable) channel, the time-varying analytic amplitude was extracted from eight bandpass filters (Gaussian filters, logarithmically increasing centre frequencies (85–175 Hz) and semi-logarithmically increasing bandwidths) with the Hilbert transform. The high-gamma (high-γ) power was then calculated by averaging the analytic amplitude across these eight bands, and then this signal was down-sampled to 200 Hz. High-γ power was z-scored relative to the mean and standard deviation of baseline data for each channel. Throughout the Methods, high-γ power refers to this z-scored measure.

Acoustic analysis. The recorded speech signal was transcribed off-line using WaveSurfer (http://www.speech.kth.se/wavesurfer/). The onset of the consonant-to-vowel transition was used as the common temporal reference point for all subsequent analysis (see Supplementary Fig. 1). This was chosen because it permits alignment across all of the syllables and allows for a consistent discrimination of the consonantal and vocalic components. Post-hoc analysis of acoustic timing revealed the onset of the consonant-to-vowel transition to be highly reproducible across multiple renditions of the same syllable. As such, alignment at the consonant-to-vowel transition results in relatively small amounts of inter-syllable jitter in estimated times of acoustic onset, offset and peak power.

For temporal analysis of the consonant-vowel acoustic structure, each individual vocalization was first converted to a cochlear spectrogram by passing the sound-pressure waveform through a filter bank emulating the cochlear transfer function[49]. As the current analysis of cortical data leverages the cross-syllabic variability in (average) high-γ (see below), we reduced the data set of produced vocalizations to a single exemplar for each consonant-vowel syllable. For each unique consonant-vowel syllable, the cochlear spectrograms associated with each utterance of that consonant-vowel ($S_i(t,f)$) were analysed to find a single prototypical example (Pspct), defined as the syllable that had the minimum (min) spectrotemporal difference from every other syllable of that kind:

$$\text{Pspct} = \min_{S_i}\left(\sum_{j}\sum_{t,f}\left(S_j(t,f) - S_i(t,f)\right)^2\right) \qquad (1)$$

where, $S_i(t,f)$ is the spectrogram of the ith example of the syllable, corresponding to the power at time t and frequency f. The onset, peak and offset of acoustic power were extracted for each syllable prototype using a thresholding procedure.

Articulator state matrix and phonetic feature matrix. To describe the engagement of the articulators in the production of different consonant-vowel syllables, we drew from standard descriptions of the individual consonant and vowel sounds in the International Phonetic Alphabet (IPA)[50]. Each consonant-vowel syllable was associated with a binary vector describing the engagement of the speech articulators used to produce the consonant-vowel syllable. For the linear analyses presented in Figs 2 and 3, the articulator state vector (B_i) for each consonant-vowel syllable s_i was defined by six binary variables describing the four main articulator organs (lips, tongue, larynx and jaw) for consonant production and two vocalic tongue configurations (high tongue and back tongue) (Supplementary Fig. 1). Although more detailed descriptions are possible (for example, alveolar-dental), the linear methods used for these analyses require the articulator variables to be linearly independent (no feature can be described as an exact linear combination of the others), although the features may have correlations. An expanded phonetic feature matrix (nine consonant constriction location variables, six consonant constriction degree or shape variables, and four vowel tongue and lip configuration variables; derived from the IPA, Supplementary Fig. 12) was used in the non-parametric analysis of the cortical state-space (Figs 4 and 5).

Spatial organization derived from a general linear model. To examine the spatial organization with which high-γ was modulated by the engagement of the articulators, we determined how the activity of each electrode varied with consonant articulator variables using a general linear model (GLM). Here, at each moment in time (t), the GLM described the high-γ of each electrode as an optimally weighted sum of the articulators engaged during speech production. High-$\gamma(t)$ (H$\gamma(t)$) recorded on each electrode (e_i), during the production of syllable s_j, H$\gamma_{i,j}(t)$, was modelled as a linear weighted sum of the binary vector associated with the consonant component of s_j, ($B^c{}_j$):

$$\mathrm{H}\gamma_{i,j}(t) = \beta_i(t) \bullet B_j^C + \beta_{i0}(t) \qquad (2)$$

The coefficient vector $\beta_i(t)$ that resulted in the least-mean square difference between the levels of activity predicted by this model and the observed H$\gamma(t)$ across all syllables was found by linear regression. For each electrode e_i at time t, the associated 1 × 4 slope vector ($\beta_i(t)$) quantifies the degree to which the engagement of a given articulator modulated the cross-syllable variability in H$\gamma(t)$ at that electrode. Coefficients of determination (R^2) were calculated from the residuals of this regression. In the current context, R^2 can be interpreted as the amount of cross-syllabic variability in Hγ that can be explained by the optimally weighted linear combination of articulatory state variables.

The spatial organization of the speech articulators was examined using the assigned weight vectors ($\beta_i(t)$) from the GLM described above. First, the fit of the GLM at each electrode e_i was determined to be of interest if, on average, the associated P-value was less than 0.05 for any one of the four consonant articulator time windows (T_A; see below) determined from the partial-correlation analysis below. We defined this time window to be the average onset-to-offset time of statistically significant ($P < 0.05$) partial correlations for each individual articulator in each subject (see the section on partial correlation analysis below). This method identifies electrodes whose activity is predicted well by the GLM for any of the individual articulators, as well as for combinations of these articulators, for extended periods of time. As these time windows extend for many points, this is a relatively stringent criterion in comparison to a min-finding method or looking for single significant-crossings. In practice, the minimum P-values (across time) associated with the vast majority of these electrodes are several orders of magnitude less then 0.05. For the electrodes that were gauged to have statistically significant correlations in each subject, we averaged the weights for each articulator (A) in that articulators time window (T_A). Thus, each electrode of interest (e_i) is assigned four values, with each value corresponding to the weighting for that articulator, averaged across that articulator's time window:

$$W_i^A = \frac{1}{|T_A|}\sum_{t \in T_A} \beta_i(t) \qquad (3)$$

For the analysis of representational overlap at individual electrodes (Supplementary Fig. 6), each electrode was classified according to the dominant articulator weight in a winner-take-all manner. The fractional articulator weighting was calculated based on the positive weights at each electrode, and was plotted as the average percentage of summed positive weights.

For spatial analysis, the data for each subject were smoothed using a 2-mm uniform circular kernel. The maps presented and analysed in Supplementary Figs 2 and 3 correspond to these average weights for the lips, tongue, larynx and jaw. The maps presented and analysed in Fig. 2 correspond to these average weights for each articulator averaged across subjects. The spatial organization of vSMC is described by plotting the results of the GLM for an individual on the cortex of that individual. We used a Cartesian plane defined by the antero-posterior distance from the central sulcus (ordinate) and the dorsoventral distance from the Sylvian fissure (azimuth). This provides a consistent reference frame to describe the spatial organization of each subject's cortex and to combine data across subjects while preserving the individual differences.

Somatotopic map and *k*-nearest neighbours algorithm. To construct the summary somatotopic map of Fig. 2b, we first extracted the spatial location of the top 10% of weights for each articulator (averaged across subjects, data are shown in Fig. 2a). We then used a k-nearest neighbour algorithm to classify the surrounding cortical tissue based on the nearest $k = 4$ neighbours within a spatial extent of 3 mm of each spatial location; if no data points were present within 3 mm, the location is unclassified. Locations in which no clear majority (>50%) of the nearest neighbours belonged to a single articulator were classified as mixed. These values were chosen to convey, in summary form, the visual impression of the individual articulator maps, and to 'fill in' spatial gaps in our recordings.

The summary map changed smoothly, and as expected with changes in the threshold of weights for each articulator of individual articulator maps, k (number of neighbours), spatial extent and minimum number of points. Results are qualitatively insensitive to the details of this analysis, including the choice of 10% as a threshold, as changes in the clustering algorithm could be made to accommodate subtle differences in data inclusion (for visual comparison, we display the somatotopic maps derived from the same algorithm derived from the top 5%, top 10% and top 15% of weights in Supplementary Fig. 4).

Partial correlation analysis. To quantify the temporal structure with which single-electrode Hγ was correlated with the engagement of a single articulator, we used partial correlation analysis. Partial correlation analysis is a standard statistical tool that quantifies the degree of association between two random variables (here, $\mathrm{H}\gamma(t)$ and the engagement of a given articulator, A_i), and removes the effect of a set of other random variables (here, the other articulators, $A_j, j \neq i$). For a given electrode, the partial correlation coefficient between $\mathrm{H}\gamma(t)$ across syllables at time t and articulator A_i ($\rho(\mathrm{H}\gamma(t),A_i)$) is calculated as the correlation coefficient between the residuals $r(\mathrm{H}\gamma(t),A_j)$, $j \neq i$ resulting from de-correlating the $\mathrm{H}\gamma(t)$ and every other articulator A_j, $j \neq i$, and the residuals $r(A_i,A_j)$, $i \neq j$ resulting from de-correlating the articulators from one another:

$$\rho(\mathrm{H}\gamma(t),\, A_i) = \frac{\mathrm{cov}\left(r\left(\mathrm{H}\gamma(T),\, A_j\right),\, r\left(A_i,\, A_j\right)\right)}{\sigma_1 \times \sigma_2}, i \neq j \qquad (4)$$

Where σ_1 and σ_2 are the standard deviations of $r(\mathrm{H}\gamma(t),A_j)$ and $r(A_i,A_j)$, respectively. In the current context, the partial correlation coefficients quantify the degree to which the cross-syllabic variability in Hγ at a given moment in time was uniquely associated with the engagement of a given articulator during speech production. For each articulator, we analysed those electrodes whose peak partial correlation coefficient (ρ) exceeded the mean ± 2.5 σ of ρ values across electrodes and time ($>\mathrm{mean}(\rho(e_i,t)) + 2.5\sigma(\rho(e_i,t))$). In the text, we focus on the positive correlations (which we denote as R in the text), because there were typically a larger number of positive values than negative values (mean $\rho > 0$), and in general the temporal profiles are similar for negative values and for expositional simplicity. Results did not qualitatively change with changes in this threshold of approximately ± 0.2 σ. We extracted the onset, offset and peak times for each articulator for each electrode that crossed this threshold. The data presented in Fig. 3d are the mean ± s.e.m. of these timing variables across electrodes pooled across subjects. The average onset and offset for each of the four consonant articulators (lips, tongue, jaw and larynx) in each subject was used to define the articulator time window used in the spatial analysis described above.

Principal component analysis and cortical-state space. Principal components analysis (PCA) was carried out on the set of all vSMC electrodes for dimensionality reduction and orthogonalization. PCA was performed on the $n \times m^*t$ covariance matrix Z with rows corresponding to channels (of which there are n) and columns corresponding to concatenated $\mathrm{H}\gamma(t)$ (length t) for each consonant-vowel (of which there are m). Each electrode's time series was z-scored across syllables to normalize response variability across electrodes. The singular-value decomposition of Z was used to find the eigenvector matrix M and associated eigenvalues λ. The principal components (PCs) derived in this way serve as a spatial filter of the electrodes, with each electrode e_j receiving a weighting in PC_i equal to M_{ij}, where M is the matrix of eigenvectors. Across subjects, we observed that the eigenvalues (λ) exhibited a fast decay with a sharp inflection point at the ninth eigenvalue, followed by a much slower decay thereafter (Supplementary Fig. 8). We therefore used the first nine eigenvectors (PCs) as the cortical state-space for each subject.

The cortical state-space representation of syllable s_k at time t, $K(s_k,t)$, is defined as the projection of the vector of cortical activity associated with s_k at time t, $\mathrm{H}\gamma_k(t)$, onto M:

$$K(S_k,t) = M \bullet \mathrm{H}\gamma_k(t) \qquad (5)$$

We calculated the contribution of articulators to the cortical state-space ($\mathrm{PCw}_{\mathrm{ij}}$) by projecting each electrode's weight vector (β_j; derived from the GLM model above) into the first three dimensions of the cortical state-space (i = 1–3):

$$PC_{W_{ij}} = M_{ij} \bullet \beta_j \qquad (6)$$

Here, $\mathrm{PC}w_{ij}$ is a four-element vector of the projected articulator weights for electrode e_j into PC_i. In Supplementary Fig. 9, we plot $\log_{10}$ of the absolute value of $\mathrm{PC}w_{ij}$ across electrodes, which describes the distribution of magnitudes of the representations associated with the four articulators in a given PC.

Clustering analysis. The k-means and hierarchical clustering analyses were carried out on the cortical state-space representations of syllables, $K(s_k,t)$, based on the pair-wise Euclidean distances calculated between consonant-vowel syllable representations. Agglomerative hierarchical clustering used Ward's method. All analyses of the detailed binary phonetic feature matrix were carried out using both Hamming and Euclidean distances; results did not change between metrics qualitatively or statistically. We used silhouette analysis to validate the claim that there were three clusters at the consonant time. The silhouette of a cluster is a measure of how close (on average) the members of that cluster are to each other, relative to the next nearest cluster. For a particular data set, the average silhouette for a given number of clusters describes the parsimony of the number of clusters in the data. Hence, examining the silhouette across different numbers of clusters gives a quantitative way to determine the most parsimonious number of clusters[51]. Higher values correspond to more parsimonious clustering. On average across subjects, this analysis validated the claim that three clusters (average silhouette = 0.47) was a more parsimonious clustering scheme than either two (average silhouette = 0.45) or four clusters (average silhouette = 0.43).

Correlation of cortical state-space structure with phonetic structure. At each moment in time, we wanted to quantify the similarity of the structure of cortical state-space representations of phonemes and the structure predicted by different phonetic feature sets. To do this, we measured the linear correlation coefficient between vectors of unique pair-wise Euclidean distances between phonemes calculated in the cortical state-space ($\mathrm{DC}(t)$) and in the phonetic feature matrix (DP):

$$R(t) = \frac{\mathrm{cov}(\mathrm{DC}(t),\, \mathrm{DP})}{\sigma_{\mathrm{DC}(t)} \times \sigma_{\mathrm{DP}}} \qquad (7)$$

As described above, the phonetic feature matrix was composed of three distinct phonetic feature sets (consonant constriction location, consonant constriction degree or shape, vowel configuration). Distances were calculated independently in these three subsets and correlated with $\mathrm{DC}(t)$. Standard error measures of the correlation coefficients were calculated using a bootstrap procedure (1,000 iterations).

Cluster separability. Cluster separability is defined at any moment in time as the difference between the average of cross-cluster distances and the average of within-cluster distances. This quantifies the average difference of the distance between syllables in different clusters and the tightness of a given cluster. We quantified the variability in cluster separability estimation using a 1,000-iteration bootstrap procedure of the syllables used to calculate the metric.

Cluster density. We quantified the average cluster density by calculating the average inverse of all unique pair-wise distances between consonant-vowels in a given cortical state-space cluster. It is a density because the number of elements in a cluster does not change with time.

48. Mesgarani, N. & Chang, E. F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485,** 233–236 (2012).
49. Yang, X. *et al.* Auditory representations of acoustic signals. *IEEE Transactions Inf. Theor.* **38,** 824–839 (1992).
50. International Phonetic Association. *Handbook of the International Phonetic Association* (Cambridge Univ. Press, 1999).
51. Rousseeuw, P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20,** 53–65 (1987).