Distinct neural processes link speech planning and execution

Suseendrakumar Duraivel[1], Shervin Rahimpour[2,3], Katrina Barth[1], Chia-Han Chiang[1], Charles Wang[1], Stephen C. Harward[2,4], Shivanand P. Lad[2], Daniel P. Sexton[2], Allan H. Friedman[2], Saurabh R. Sinha[5], Gregory Hickok[6,7], Derek G. Southwell[1,2,4,8], Jonathan Viventi[1,2,4,8,#], Gregory Cogan[1,2,4,9,10,11, #]

[#]Corresponding authors

Author affiliations:

1 Department of Biomedical Engineering, Duke University, Durham, NC, USA

2 Department of Neurosurgery, Duke School of Medicine, Durham, NC, USA

3 Department of Neurosurgery, Clinical Neuroscience Center, University of Utah, Salt Lake City, UT, USA

4 Duke Comprehensive Epilepsy Center, Duke School of Medicine, Durham, NC, USA

5 Penn Epilepsy Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

6 Department of Cognitive Sciences, University of California, Irvine, CA, USA

7 Department of Language Science, University of California, Irvine, CA, USA

8 Department of Neurobiology, Duke School of Medicine, Durham, NC, USA

9 Department of Neurology, Duke School of Medicine, Durham, NC, USA

10 Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

11 Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

Correspondence to:

Gregory B. Cogan

Department of Neurology

Duke South, Blue Zone 4519, Durham, NC 27710

gregory.cogan@duke.edu

Jonathan Viventi

Biomedical Engineering

1149 CIEMAS, 101 Science Dr, Durham, NC 27705

j.viventi@duke.edu

Running title: Neural speech sequencing

Keywords: Speech, IEEG, syllable sequences, speech planning, speech production

Abstract

Speaking is the primary way that humans communicate. This communication is enabled by a production system that can plan and execute unique combinations of speech sounds. Although a distributed network of brain regions has been implicated in speaking, it is unclear how planning and execution of speech are coordinated to produce meaningful sounds. Leveraging the high spatio-temporal resolution of intracranial recordings at different spatial scales, we show distinct neural mechanisms that facilitate speech planning and execution. During planning, different levels of speech units are coded discretely at distinct prefrontal sites. These planned units are then dynamically integrated at various cortical levels to guide subsequent execution. During speech execution, speech motor regions generate continuous sequences that reflect both discrete speech sound units and their transitional properties between units. This rapid neural transition from discrete speech units to motor sequences links speech planning with execution and enables our effortless ability to speak.

Word Count: 150

Introduction

Speech production enables verbal communication through planning and motor execution of combinations of speech sounds[1–3]. This fundamental human capacity of speech occurs by turning mental ideas to words[4–6], which are then mapped to smaller speech units to generate motor sequences for execution[7,8]. This ability is acquired over years of learning to speak in which speech plans organize the vocal execution for continuous speech production. Despite this effortless everyday usage, the neural basis of how speech plans and execution are organized and linked is not well characterized.

From the psycholinguistic literature[9–11], speech plans are hypothesized to operate at two distinct levels: a framing level which corresponds to the syllable, and a content level which combines phonemes within syllable frames[12–16]. Behavioral evidence for these planning elements comes from analyses of speech errors such as slips of the tongue and phoneme exchange errors during everyday speech and in aphasia, suggesting neural processes for distinct stages in speech planning[9,17,18]. These error patterns reveal that different linguistic units – phonemes and syllables, – interact systematically during speech planning, suggesting neural substrates for distinct stages in this pre-execution process.

Planned utterances are subsequently fed into a motor execution system that: 1) generates commands to control the sequence of vocal tract gestures for articulation, and 2) updates the commands based on sensory feedback from the auditory and the somatosensory system[12–15,19,20]. Together, representations for syllables and phonemes combine during planning and execution to allow the speaker to effortlessly utter any sequence of speech sounds within the phonological rules of their proficient language[9–11].

Despite the importance of planning and execution for speech production, the neural processes that distinguish these hypothesized distinct stages remain largely unexplored *(**Fig. 1a, b**)*.

Previous work attempting to distinguish planning and execution of speech production have come from neuroimaging studies, which have identified distributed neuroanatomical networks of frontal, parietal, and temporal regions. Specifically, these regions include: inferior frontal gyrus (IFG), rostral and caudal middle frontal gyrus (rMFG & cMFG), pre-central and post-central gyrus (PrCG & PoCG; sensorimotor cortex), supra-marginal gyrus (SMG), inferior parietal cortex (IPC), insula, superior temporal gyrus (STG), and primary auditory cortex (PAC)[1,12,21–30]. Within this network, prefrontal regions have primarily been associated with planning, sensorimotor circuits (PrCG, PoCG, SMG, and IPC) associated with motor execution,[3,31] and superior temporal regions and the insula associated with monitoring based on sensory feedback (auditory and somatosensory)[3,32–34], as well as internal error correction[35,36]. Significant progress has been made to understand how speech is executed by sensorimotor circuits through coordinated activity to produce speech sounds[4,24,29], as well as how speech is monitored through neural suppression of self-generated speech sounds in the superior temporal gyrus (STG)[37–40]. For speech planning, however, neuroimaging studies have been limited by their lack of precise timing to distinguish neural processes associated with planning from neural processes associated with execution. This limitation makes it difficult to directly study the presence of distinct speech units such as syllables and phonemes, as distinct processing stages during planning.

One indirect approach to study the structure of speech planning has been to manipulate the aspects of syllables (order, complexity, and novelty of syllables), and examine the cortical regions activated using neuroimaging[41]. Notably, the studies that have manipulated the syllable structures found activation in the posterior inferior frontal gyrus (pars-opercularis)[10,11,41–43], supporting the idea that the left IFG could be involved in syllable planning for speech production. However, despite these insights, the precise temporal dynamics involved in the planning and execution across the speech network, particularly how syllable and phoneme processes unfold over time – remains unknown.

Our understanding of the temporal dynamics of speech planning and execution have mostly come indirectly from computational models of speech production, including the DIVA (Directions Into Velocities of Articulators), GODIVA (Gradient Order DIVA), and hierarchical state feedback control (HSFC) models have been developed based on the neuroimaging literature to integrate psycholinguistic (syllable-phoneme) and motor-control (planning-execution) theories that model the accurate planning, execution, and monitoring of sequences of speech sounds[14,16,27]. The DIVA model provides a neuroanatomically explicit framework for speech motor control through parallel feedforward and feedback systems, mapping abstract speech representations to articulatory movements with error

correction via auditory and somatosensory feedback[27]. GODIVA extends DIVA by adding mechanisms for sequencing syllables and phonemes in multi-syllabic utterances by incorporating sequential structure buffers and phonological content buffers[16]. Further, Hickok's HSFC model integrates these approaches within an hierarchical architecture where higher-level control codes speech segments at syllabic level while the lower-level execution processes update the contents at phonemic level[14]. Simulations using this hierarchical organization resulted in accurate generation of phoneme sequences, with a higher order planning state coding speech at the syllabic level and a lower order planning state at the phoneme level.[13,14] These computational studies suggest that during planning, the two distinct speech representations (syllables and phonemes) are composed both hierarchically and in serial order, with syllables planned prior to phoneme plans. While these models describe the computational and anatomical structure of speech production and incorporate both feedforward and feedback control mechanisms, they cannot empirically specify the precise timing and temporal evolution of neural activations across regions during natural speech production. As a result, direct empirical evidence for the temporal ordering and integration of syllable and phoneme representations during planning and execution is lacking, due to limitations in recording technology for resolving neural signals with both high spatial and temporal precision.

Intracranial recordings using electrocorticography (ECoG) and stereo-electroencephalography (SEEG) record neural activity directly from cortex with high temporal resolution (milliseconds) with precise anatomical localization[44–47]. This technology has been further advanced through the development of high-density micro-electrocorticographic (μECoG; 1.33 – 1.72 mm inter-electrode distance) arrays that can measure micro-scale cortical activity with millimeter scale spatial resolution[48–52]. Neural activations from ECoG and SEEG in the high-gamma band (HG: 70-150 Hz) have been shown to correlate with multi-unit firing rates, making this signal a good measure of local neural activity[53–57].

During speech production, studies have shown HG activations in the ventral sensorimotor cortex that encode articulatory properties during speech motor execution[19,20,58–63]. These findings have provided detailed insights into how specific brain regions control vocal tract movements and coordinate articulatory gestures. However, fundamental questions remain about how the brain enables the transition from speech planning to precise motor execution.

Surprisingly, there has been limited intracranial work examining speech planning, and the critical transition between planning and execution. One previous study used overt sentence completion to examine lexical access in the IFG prior to articulation[64], but this activity could be related to either speech perception or production processes, making its role in speech production unclear. Another study used an unstructured experimental setting to examine the timing of planning in relation to turn taking in

conversation[65] but it did not examine the content of planning or provide evidence for how speech plans enable motor execution. A more recent study demonstrated that speech motor sequencing relies on a distributed network centered on the middle precentral gyrus, showing sustained activity that is modulated by sequence complexity, yet little is discussed about the presence or absence of hierarchical organization of sublexical units (syllables and phonemes), or the temporal dynamics of the integration during the planning-to-execution transition[66]. These studies do highlight the promise of intracranial recordings for studying different stages of speech production via the measurement of spatially specific neural representations. The high temporal and spatial resolution of intracranial recordings also presents an opportunity to delineate the neural substrates associated with the transition from speech planning to execution. Understanding this transition could have important implications for treating neurological conditions that affect both higher order cognition and speech production, as well as in developing more advanced neural prosthetic technology.

To elucidate the neural substrates for planning and execution in speech production, we performed intracranial recordings with 52 patients while they verbally articulated constructed pseudo-words. We show distinct neural activations for planning and execution that represent both syllables and phonemes. The neural activations demonstrate distinct spatial and temporal patterns that link speech planning to execution. During the transition from planning to execution, we showcase neural processes that transform speech sound units - syllables and phonemes - from plans to motor execution. In execution, we show evidence for motor sequencing through ordered patterns of phonemes as well as evidence for the neural processing of the transitions between phonemes. Together, these findings provide evidence for distinct neural patterns that enable the rapid transition from discrete speech units during planning to sequenced speech sounds during motor execution, supporting fluent speech production.

1 Results

We performed a speech task with 52 patients (n = 52, mean age = 33.1 years, 23 females, Supplementary table 1) in the epilepsy clinical monitoring unit, and with 3 patients intra-operatively (n = 3, mean age = 64.3 years, 0 females, Supplementary table 1). Patients in the monitoring unit were implanted with clinical intracranial electrodes (electrocorticography [ECoG: 2 patients] or stereo-electroencephalography [SEEG: 50 patients]) for the localization of epileptic tissue to guide surgical resection or neuromodulation targeting (**Fig. 1c**). Patients recorded intra-operatively were awake as part of their surgical procedure for tumor resection (1 patient) or deep brain stimulator (DBS) placement for the treatment of movement disorders (2 patients), and these neural recordings were obtained at micro-scale cortical level at high spatial density using μECoG electrode arrays[63].

Participants performed a controlled pseudo-word repetition task. Pseudo-words were presented auditorily through speakers connected to a presentation laptop running Psychtoolbox[67]. After the pseudo-word was presented, there was a short delay (1.2 – 1.6 s), followed by a prompt (using a visual "Speak" cue), during which the patients were instructed to repeat back the presented pseudo-word stimulus (**Fig. 1d**). Each pseudo-word was either composed with a monosyllabic (consonant-vowel-consonant: CVC) or a disyllabic (vowel-consonant-vowel: VCV) structure, that contained three phonemes (out of a set of 9 total available phonemes - **Fig. 1e,** Supplementary table 2). Each patient completed four task blocks that contained 52 unique tokens each. Due to the shorter availability of time, patients in the intra-operative setting underwent a similar but modified version of the pseudo-word repetition task in which they repeated the pseudo-words immediately after the auditory presentation. The intraoperative patients completed three to four repetitions of task blocks.

1.1 Behavioral evidence for syllables

On average, the patients who underwent delayed speech-repetition task, repeated the pseudo-words after 0.58 seconds (standard deviation: ± 0.33 s) from the visual "Speak" cue and with an average duration of 0.65 seconds (± 0.16 s). Mean response times (RT) and response durations (RD) were (mean ± SD): CVC: RT = 0.56 ± 0.35 s, RD = 0.58 ± 0.17 s; and VCV: RT = 0.6 ± 0.32 s, RD = 0.68 ± 0.14 s (**Fig. 1d**).

To quantify the effect of differences in syllable on behavioral performance, we performed linear mixed effects (LME) modeling on each trial, with a fixed effect modeling the syllable identity linked to each pseudo-word and a random effect modelling the patient identity. VCV structure utterances had longer durations compared to CVC structures [LME: $t_{(8696)} = 10.7$, $\beta = 0.098$, $p = 2.1e\text{-}26$, 95% CI = 0.08 to 0.1, **Fig. 1f,** Supplementary Fig. 1a], in keeping with their longer articulatory and acoustic duration (2 vs. 1 syllable). Interestingly, response times were significantly delayed for disyllabic VCV structures as compared to monosyllabic CVC structures [LME: $t_{(8698)} = 3.7$, $\beta = 0.045$, $p = 2.2\ e\text{-}04$, 95% CI = 0.02 to 0.07, **Fig. 1g,** Supplementary Fig. 1b], suggesting that speech motor planning reflects the composition of syllables. Further, this behavioral result also held when controlling for the response duration, supporting the interpretation that these effects reflect planning of syllables rather than the overall utterance length [LME: $t_{(9084)} = 2.75$, $\beta = 0.103$, $p = 0.006$, 95% CI = 0.03 to 0.18, Supplementary Fig. 1c].

1.2 Neural processes for speech planning and execution

To investigate the neural processes involved in speech production, we first extracted neural activations in the high-gamma band (HG: 70 – 150 Hz) that have previously been shown to be an index of local neural computation and a correlate of multi-unit firing[54,56,57]. Only correctly articulated trials with

response times greater than 50 ms (from “Speak”/go cue) were used for neural analysis. Out of 8106 preprocessed IEEG electrodes from 52 patients in the clinical monitoring unit, we observed significant increases in HG power in 3534 electrodes (2895 electrodes with grey matter contacts) during speech production as compared to a pre-stimulus baseline (one-sided permutation test with post-hoc FDR correction per patient, $p < 0.05$, **Fig. 2a,** Supplementary table 2). Speech neural activations were present in both the left (1701 electrodes) and right (1665 electrodes) hemisphere (**Fig. 2b**), confirming the bilateral nature of speech production networks[20].

We further selected electrodes within anatomical regions of interest (ROIs) that have been shown to be active for speech production[4,16,20,65,68]. The selected ROIs in both hemispheres included: inferior frontal gyrus (IFG: 92 significant electrodes), rostral middle frontal gyrus (rMFG: 137), caudal middle frontal gyrus (cMFG: 186), precentral and postcentral gyrus (PrCG & PoCG: 351), inferior parietal cortex (IPC: 268), insula (172), anterior superior temporal gyrus (aSTG: 91), posterior superior temporal gyrus (pSTG: 90), superior temporal sulcus (STS: 88), and primary auditory cortex (PAC: 43).

To characterize the timing of activation between ROIs, we estimated the peak onsets of HG activation for each electrode relative to the utterance onset (**Fig. 2c**). There were significant differences in peak activations between ROIs in both the left and right hemispheres (one-way ANOVA; left hemisphere: $F_{9,717} = 10.87$, $p = 6.1e\text{-}16$, right hemisphere: $F_{9,781} = 6.69$, $p = 2.8e\text{-}9$. **Fig. 2d**). Based on neuro-anatomical studies that investigated speech articulation[4,16,20,65,68], we divided these ROIs into three distinct networks based on previous literature: planning (IFG, rMFG, and cMFG), articulation (PrCG & PoCG, and IPC), and monitoring (aSTG, pSTG, STS, and PAC). As hypothesized, using a second level one-way ANOVA, we found that in the left-hemisphere, electrodes in the planning network were activated early, followed by the electrodes the articulation network, and then the monitoring network (one-way ANOVA: $F_{2,724} = 30.7$, $p = 1.6e\text{-}13$, **Fig. 2e**). Pair-wise post-hoc t-tests showed that the mean values were significantly different between all three groups, supporting distinct functional roles rather than simple anatomical groupings ($p < 0.05$, $T_{Planning} < T_{Articulation}$, $t_{(365)} = 2.938$, p = 0.0018; $T_{Articulation} < T_{Monitoring}$, $t_{(507)} = 9.25$, p = 3e-13, $T_{Planning} < T_{Monitoring}$, $t_{(394)} = 9.84$, p = 7.2e-21, **Fig. 2e**). Differences were also significant in the right hemisphere ($F_{2,788} = 8.86$, $p = 1.6e\text{-}4$, $T_{Articulation} < T_{Monitoring}$, $t_{(528)} = 7.53$, p = 2.2e-13, $T_{Planning} < T_{Monitoring}$, $t_{(385)} = 2.36$, $p = 0.0188$ **Fig. 2e**). However, the prefrontal-planning activations occurred later in time than the articulation regions ($T_{Planning} > T_{Articulation}$, $t_{(511)} = 2.3$, p = 0.0214), which is consistent with the role of the right hemisphere in error monitoring[69,70]. A similar organization existed when the neural data were aligned to the go-cue (Supplementary Fig. 2).

While these ROI-based network assignments were grounded in established neuroanatomical literature, such a-priori groupings could potentially introduce bias into the analysis. To address this potential

limitation and validate the network organization in a data-driven manner, we performed non-negative matrix factorization (NNMF) on the z-scored HG activation matrix to identify latent neural speech components without prior assumptions (Supplementary Fig. 3). The resulting five temporal components revealed distinct activation profiles that closely matched our hypothesized speech production stages (Supplementary Fig. 3c): Cluster 1 and 4 captured early and late auditory monitoring respectively, Cluster 2 showed articulation-related activity, Cluster 3 reflected planning processes, and Cluster 5 indicated sustained verbal working memory maintenance, as part of the sensory-motor transformation. Critically, on examining how these data-driven clusters mapped onto the anatomically defined ROIs and speech networks, we found strong correspondence between the unsupervised decomposition and our ROI-based groupings. Planning regions (IFG, rMFG, cMFG) showed maximal loading on Cluster 3, articulation regions (PrCG, PoCG, IPC) aligned predominantly with Cluster 2, and monitoring regions (aSTG, pSTG, STS, & PAC) correspond to Clusters 1 and 4. This convergence validates the network organization and demonstrates that the temporal progression reflects functional specialization rather than arbitrary anatomical clustering. Overall, these results clearly demonstrate the presence of three distinct speech networks in the left-hemisphere: 1) a planning network in the pre-frontal regions to initiate speech, 2) an articulation network that generates motor commands and executes planned speech, and 3) a monitoring network for auditory feedback tracking.

1.3 Neural syllabic code is present prior to articulation

We next sought to investigate the neural basis for speech representation associated with the three levels of processing (planning, articulation, and monitoring) at each stage. Based on prior computational work, we theorized that speech production involves the recruitment of syllables during the planning stage[4,11,14,42], followed by their execution during the produced utterance (articulation and monitoring)[14,27]. To assess the specific timing of neural activations that reflect differences in the syllable structure of the utterances, we examined differences in ROI-grouped HG activations between syllables (ΔHG; VCV and CVC pseudo-word trials, see Methods), at every time point around the response epoch (Supplementary Fig. 4). With phonemes controlled at each position within the syllable, these HG contrasts (ΔHG) can reveal activation patterns selective to the syllable alone. Results show that neural activations associated with the syllable structure showed temporal components with significant syllable coding (time clusters of significant differences in activation between VCV and CVC) for each speech network in the left hemisphere. We observed the earliest coding (ΔHG) in the prefrontal-planning network (cluster onset: -0.73s, peak onset: -0.16s), followed by the articulatory network (cluster onset: -0.22s, peak onset: 0.23s), and monitoring network (cluster onset: -0.16s, peak onset: 0.32s (one-way ANOVA, $F_{2,495} = 64.3$, $p = 1.5e\text{-}25$, **Fig. 3a**). Post-hoc t-tests showed that the differences showed pairwise significance between the three areas ($p < 0.05$), except for between the articulatory and the

monitoring areas, confirming the temporal progression from syllable planning to their motor execution ($T_{Planning} < T_{Articulation}$, $t_{(333)} = 3.3$, $p = 0.0011$; $T_{Articulation} < T_{Monitoring}$, $t_{(403)} = 1.3$, $p = 0.2$ (not significant), $T_{Planning} < T_{Monitoring}$, $t_{(352)} = 3.78$, $p = 1.7e\text{-}4$, **Fig. 3f**). To directly investigate this ROI division of syllable coding, we repeated the analysis on electrodes grouped within each ROI (**Fig. 3c,** except STS; $p < 0.05$, one-sided permutation test with cluster correction). The earliest time clusters for syllable contrasts were observed in left-hemispheric planning regions (cluster start time & mean peak onsets): 1) IFG: -0.74s & -0.3s, 2) rMFG: -0.68s & -0.05s, and 3) cMFG: -0.07s & -0.17s. Speech articulatory ROIs demonstrated relatively late timing: 1) PrCG & PoCG: -0.6 & 0.83s, and 2) IPC: -0.1s & 0.3s. Finally, the monitoring ROIs exhibited timing closer to voice-onset times: 1) Insula: -0.27s & 0.32s, 2) aSTG: -0.22s & 0.37s, 3) pSTG: -0.16s & 0.26s, and 4) PAC: -0.16s & 0.31s. We further performed a one-way ANOVA to assess the temporal properties of syllabic coding, which demonstrated significant differences in peak coding between individual ROIs (**Fig. 3d,** except STS, $F_{8,489} = 18.1$, $p = 8.9e\text{-}24$)

In the right hemisphere, significant differences in syllable neural codes were observed between the articulation and monitoring network (negative t-statistic on comparisons against planning network, $T_{Planning} < T_{Articulation}$, $t_{(406)} = -5.4$, $p = 1.1e\text{-}7$, Supplementary Fig. 5), further, supporting the right hemisphere's involvement in execution. Similar temporal progression is also observed when neural data was aligned with the go-cue onset, showing that the transition from planning to execution of syllables is also preserved at the commencement of planning with respect to the instructional cue (Supplementary Fig. 6). To address a potential confound between syllable class (mono-syllabic/CVC vs. di-syllabic/VCV) and initial phoneme category (C vs. V), we performed a control analysis by comparing the production of CV'CVC and CVC structures that differ in syllable number while sharing the same initial consonantal phoneme structure. We find that planning of monosyllabic vs. disyllabic utterances show similar planning and execution temporal activation patterns as in our main analysis (Supplementary Fig. 7). This analysis ruled out initial phoneme category influences, further supporting the existence of syllable level neural code in the left-hemispheric prefrontal regions. Our results show a clear temporal progression from planning to execution of neural activations associated with the syllable. Further, this neural activation in the planning stage prior to execution supports the idea that syllables function as units of representation for speech planning in left-hemispheric prefrontal-planning regions.

Having established the presence of the syllable during speech planning, we next sought to more closely investigate the transition from planning to execution for speech production. One possible process for this transition is that neural populations in prefrontal and sensorimotor brain regions could have distinct activation timing associated with syllable processing during the transition from planning to execution for speech production. These populations with differences in activation timing relative to the utterance

onset could form a spatial gradient from planning (prefrontal/anterior) to execution (sensorimotor/posterior) regions and could serve to link these two processes in space and time. This finding would be in keeping with prior intracranial studies[65], neuroimaging[11], and theoretical work[14,16].

Therefore, to more closely investigate the anatomical and temporal characteristics of the neural populations associated with syllable coding for planning and execution, we used high-channel count, high- density μECoG electrode arrays to record speech neural activations from two subjects. We used two versions of the μECoG arrays to record neural activity along the anterior-posterior orientations from both the planning and the articulation networks (**Fig. 4a**): subject S1, during an awake craniotomy, was acutely implanted with a 256-channel subdural array (12 x 22 matrix; center-to-center electrode distance: 1.72 mm), and subject S2, during an awake deep brain stimulation (DBS) procedure, was acutely implanted with a 128-channel subdural array (8 x 16 matrix; center-to-center electrode distance: 1.33 mm). The implantation procedure for S1 enabled simultaneous recordings from the prefrontal and sensorimotor regions with the array's longest side oriented parallel to the sylvian fissure. Similarly, the electrode array for S2 was primarily positioned over the prefrontal cortex, with the distal end stretching to sensorimotor cortex. Both subjects completed 4 blocks of the modified speech repetition task. We observed neural activations on individual electrodes with significant modulation in the HG band in both subjects (Supplementary Fig. 10a**,** S1: 201/256 significant electrodes, S2: 63/128 electrodes). Neural activations showed spatially varying temporal characteristics associated with both planning and execution. 18.9% of significant electrodes (Supplementary Fig. 10b**,** S1: 12/201, and S2: 38/63) reached their peak signal power prior to the onset of the utterance. These results indicate the presence of spatially distinct neural population sites that code for speech planning at a micro-scale level recorded with our high-resolution μECoG electrode arrays.

To examine the coding for syllables, we repeated our syllabic discrimination framework analysis on an electrode specific level, by contrasting HG power between disyllabic (VCV) and monosyllabic (CVC) pseudo-word trials. A one-sided permutation test followed by a second-level cluster correction ($p < 0.05$, 1000 shuffles) revealed electrodes that exhibited significant contrasts in activations for the two different syllables, providing further evidence for syllabic coding at the micro-scale (**Fig. 4b, c,** S1: 100/201, and S2: 37/63). Interestingly, these syllable contrasting electrodes were observed to be spatially organized, and they exhibited distinct temporal profiles identified by onset and peak syllabic coding. For both subjects, we observed spatially varying temporal coding of syllables, with earlier coding starting along the anterior portion of the array and later coding extending towards the posterior side (spatial linear model for S1 using F-statistic test against constant model; Onset coding: $F_{2,92} = 6.02$, $p = 0.016$, **Fig. 4d** and Peak coding: $F_{2,98} = 10.9$, $p = 1.3e\text{-}3$, **Fig. 4d**). Comparable prediction results were obtained for S2 where the timings of syllable codes demonstrated a significant spatial

anatomical gradient along the anterior-posterior orientation (Peak coding: $F_{2,28}$ = 5.68, *p = 0.024* and Onset coding ($F_{2,28}$ = 6.02, *p = 0.07,* **Fig. 4d**). Taken together, the presence of a spatial anatomical gradient suggests the existence of populations of temporally tuned neurons that activate in succession along the anterior posterior orientation. These populations could form the basis for an anatomical substrate for the transition from planning to execution for speech production.

1.4 Two-tiered hierarchical organization for speech production

The ROI-grouped analysis demonstrated how syllables are composed for planning and execution during speech production. However, as syllables are incrementally composed, the speech production system must incorporate appropriate phonemic codes into these structures before generating complete motor articulatory commands during execution[4,13,71]. In other words, the syllable (e.g. CVC) acts as a frame that is filled in with its constituent phonemes (e.g. CVC → /kæt/). We propose that this hierarchical organization is reflected in distinct activation for the syllable and phoneme units in both planning and execution for speech production. The syllables are composed prior to phonemes, which reflect their role as frames and their position at higher level in the speech compositional hierarchy. Conversely, phonemes are encoded later as they reflect a lower level of phonological unit for speech production to that are 'filled-in' to the syllabic frame[14](**Fig. 5a**).

We demonstrated this hierarchical organization by performing temporally resolved neural decoding of speech features (syllables and phonemes) in our speech production tasks from our in-unit patients. We performed a 10-fold, cross-validated decoding (SVD-LDA; see Methods) from HG neural activations (-0.5 s to 0.5 s around utterance onset) using speech significant electrodes across all ROIs pooled across subjects (2895 total grey matter electrodes across both hemispheres, 69 mean electrodes per subject). Significant decoding performance was observed for both syllables (2-way: CVC vs. VCV, average accuracy= 93%, chance = 50%, **Fig. 5b**) and phonemes (9-way: /a/, /æ/, /i/, /u/, /b/, /p/, /v/, /g/, and /k/, average accuracy = 38.3%, chance = 11.11%, **Fig. 5b**). Having established significant decoding for both syllables and phonemes, we sought to characterize the temporal evolution of syllables and phonemes. Using time-resolved decoding (200 ms window; 10 ms step size), we obtained stable decoding for both syllables and phonemes (temporal permutation cluster test, $p < 0.05$, one-sided permutation test against 250 shuffled decoding models). In the left-hemisphere, we observed sustained decoding for syllables and phonemes. Crucially, syllables (cluster onset: -1.3s) were detectable at earlier times compared to phonemes (cluster onset: -0.17s, $\text{Onset}_{\text{Syllable}} < \text{Onset}_{\text{Phoneme}}$, one-sided Mann-Whitney U test, *p = 3.14e-8*, **Fig. 5c**), providing evidence that syllables are coded prior to phonemes. We also demonstrated the same temporal order of syllable followed by phoneme within individual ROIs and observed similar stable decoding for both syllables and phonemes (Supplementary Fig. 11; except rMFG that exhibited coding only for syllables). For both planning and articulation, we

observed syllables represented earlier in time as compared to phonemes (**Fig. 5d, e**), suggesting that this representational hierarchy is preserved in time both within and across brain areas and speech stages.

Next, we sought to determine whether the temporal relationship between syllables and phonemes is dependent on planning, articulation, and monitoring. We observed a significant difference in onsets of the decoding ability (decoding onsets) for syllables and phonemes from planning to execution, with (as previously mentioned) syllables occurring earlier than phonemes (syllables: one-way ANOVA; onset $p = 1.7e\text{-}9$, peak $p = 3.5e\text{-}54$, **Fig. 5d, e**, phonemes: one-way ANOVA; onset $p = 2.5e\text{-}16$, peak $p = 3.6e\text{-}13$, **Fig. 5d, e**). Post-hoc Tukey's HSD test showed that the differences were also pairwise significant ($p < 0.05$) between the three networks, except for peak phoneme decoding between the articulatory and the monitoring networks where the difference did not reach significance ($p = 0.0785$). Interestingly, this finding suggests that the magnitude of the difference in timing between the syllable and the phoneme differs as a function of stages/networks.

To assess the change in relative timing of the neural response between syllables and phonemes, we then compared their activation timing within each ROI group. We reasoned that planning areas that code for distinct hierarchical speech units would show a greater temporal delay between syllable and phonemes as compared to execution stages which would show a more concurrent activation pattern. This result could reflect the difference between separately composing syllable and phoneme units during planning, integrating them together for execution, and finally having temporal overlap during feedback. Our results support this hypothesis: there was a maximum delay in the prefrontal-planning network (**Onset**: syllable mean = -0.43s, phoneme mean = -0.22s, difference = 0.25s, $p = 0.0385$, one-sided Mann Whitney U test; **Peak**: syllable mean = -0.0125s, phoneme mean = 0.3s, difference = 0.32s, $p = 3.11e\text{-}8$, **Fig. 5f, g**), followed by the articulation network (**Onset**: syllable mean = -0.22s, phoneme mean = -0.13s, difference = 0.13s, $p = 3.5e\text{-}5$; **Peak**: syllable mean = 0.24s, phoneme mean = 0.35s, difference = 0.11s, $p = 2.7e\text{-}8$, **Fig. 5f, g**), and then minimally within the monitoring network (**Onset**: syllable mean = 0.039s, phoneme mean = 0.055, difference = 0.017s, $p = 0.082$; **Peak**: syllable mean = 0.39s, phoneme mean = 0.37s, difference = -0.025s, $p = 0.9$, **Fig. 5f, g**). The mean differences were pairwise significant between all three networks (pairwise t-test, planning vs. articulation, **onset**: $t_{(798)} = -10.7$, $p = 5.5e\text{-}25$, **peak**: $t_{(798)} = 71$, $p = 0$; articulation vs. monitoring: **onset**: $t_{(798)} = 18.2$, $p = 1.9e\text{-}62$, **peak**: $t_{(798)} = 101$, $p = 0$; planning vs. monitoring: **onset**: $t_{(798)} = -8.04$, $p = 3.3e\text{-}15$, **peak**: $t_{(798)} = 117.8$, $p = 0$, **Fig. 5f, g**). Repeating the analysis with electrodes pooled from right-hemispheric ROIs, we observed significant decoding for syllables and phonemes, but we did not observe temporal differences in support of the syllable-phoneme hierarchy (Supplementary Fig. 12). Overall, these results highlight the hierarchical organization between the syllable and phoneme representation from planning to

execution. The differences in hierarchical representation at each speech stage support a transition from separate representations of syllables and phonemes for speech planning in the prefrontal cortex to execution coding in the pre/post central gyrus for articulation and in temporal cortex and insula for monitoring. This temporal transition of activation could serve as a substrate for the integration of different hierarchical levels of speech units for production from planning to execution (**Fig. 5h**).

1.5 Phonological units are sequentially represented only during execution

After planning, the speech production system must successfully articulate the correct ordering of phonological sequences (phonemes) in real-time for the correct execution of the utterance (**Fig. 6a**). To show neural evidence for this sequencing, we performed temporal decoding (9-way; chance 11.11%) to predict the positional phonological units within the pseudo-word utterance in time (e.g. /abæ/: /a/→/b/→/æ/, P1 = /a/, P2 = /b/, P3 = /æ/). A successful sequencing must involve significant decoding of all three phonemes, as well as the correct ordering of phonemes within the sequence. We observed stable decoding and the correct ordering of sequences (P1→P2→P3) in the ROIs specific to the articulation and monitoring networks. Surprisingly, this ordering was absent (absence of phoneme positions or incorrect ordering) in the planning network ROIs ($p < 0.05$, one-sided permutation test against 250 shuffled decoding models, **Fig. 6b**), indicating that phoneme codes retrieved during planning are not explicitly sequenced at this stage, rather, they are sequenced in real-time in the motor cortex. In the articulation ROIs, we observed pairwise significant differences between phoneme positions in peak decoding times (one-sided Mann-Whitney U test, $p < 0.05$, **Fig. 6c**). The monitoring ROIs exhibited similar positional phoneme coding in peak times (one-sided Mann-Whitney U test, $p < 0.05$, **Fig. 6c**), which demonstrate continuous monitoring from auditory feedback. Similar ordering of phonological sequencing was also observed in the articulation and the monitoring network in the right hemisphere (Supplementary Fig. 12). Taken together, the results show that for sequencing, only the articulation network generates positional phonological unit codes in their correct order. This is carried out during the execution of speech, along with the monitoring network which examines the phonological sequence of the utterance through feedback.

Lastly, we investigated the specific neural processes involved in speech sequencing during articulation. While we showed that activation of discrete ordered phonemes was present during execution, this activation could reflect discrete motor plans for each phoneme, or sequences of continuous motor activation that reflect the execution of the entire sequence. If the entire sequence is executed holistically, then we should see evidence for both discrete and continuous motor processes during articulation.

To demonstrate both discrete and continuous processing, we examined neural activation associated with the transitions between phonemes in the sequence. These transitions reflect the statistical properties of the distribution of phoneme co-occurrences from the word patterns specific to the native language (English in this case). We posited that the sequences that co-occur more frequently could demonstrate different neural motor transition properties as compared to transitions that co-occur less frequently. If supported, evidence for information in the neural signal about these transitions would suggest that the motor cortex executing utterances holistically, as the presence of information about transitions between phonemes would reflect evidence for continuous rather than discrete processing of the utterance[72,73](**Fig. 7a**).

We recorded micro-scale cortical activations in one subject (S3) using the 128-channel high-density µECoG electrode array over speech motor cortex (SMC) (**Fig. 7b**). We found significant HG activations associated with speech production (111/128 electrodes, see also patient S1 in our previous work[63]). To assess whether the transitions between phonemes were tracked in the SMC, we focused on trials with monosyllabic structures (CVC; to remove the confound of two syllables), by measuring the ability to decode the forward phonotactic transitions (P_{fwd1}: P(P2|P1) & P_{fwd2}: P(P3|P2)). The forward transitions were estimated by calculating the conditional probability of one phoneme given the previous, based on positional and co-occurrence counts. The resultant transition probabilities were quantized into three evenly distributed classes producing trials with low, medium, and high transitions scores (P_{fwd1}: low – 0.0024 to 0.034, medium – 0.034 to 0.1031, high – 0.1031 to 0.1791; P_{fwd2}: low – 0.0083 to 0.0314, medium – 0.0314 to 0.0543, high – 0.0543 to 0.0814).

Our time resolved decoding again showed evidence of the correct order of phonological units (C: 5-way, V: 4-way) within the CVC trials (P1→P2→P3, **Fig. 7c**). Next, we repeated a similar decoding analysis to predict the forward transition scores (3-way) and computed these values in time. We found that the decoding for P_{fwd1} occurred earlier in time compared to P_{fwd2}. Within features, we observed significant temporal ordering of phonological units ($T_{P1} < T_{P2} < T_{P3}$; one-sided Mann Whitney U test, $p < 0.05$; $T_{P1} < T_{P2}$, p = 1.6e-08; $T_{P2} < T_{P3}$, p = 4.2e-07, **Fig. 7d**) and phonotactic transitions ($T_{Pfwd1} < T_{Pfwd2}$, p = 6.6e-08). Across features, the forward transitions were significantly embedded between the phonological unit positions ($T_{P1} < T_{Pfwd1} < T_{P2} < T_{Pfwd2} < T_{P3}$; pairwise one-sided Mann-Whitney U test, $p < 0.05$; $T_{P1} < T_{Pfwd1}$, p = 3.7e-08; $T_{Pfwd1} < T_{P2}$, p = 6.6e-08, $T_{P2} < T_{Pfwd2}$, p = 6.2e-08, $T_{Pfwd2} < T_{P3}$, p = 7e-07, **Fig. 7d**). These results indicate that the neural basis of speech sequential processing during articulation reflects both discrete activation of phonological units, as well as the continuous processing of the phonotactic transition between these units (**Fig. 7e**). This finding provides evidence that at least in part, the motor cortex is processing sequences holistically as a continuous combination of phonemes, rather than as a sequence of discrete phonemes.

2 Discussion

Speech planning involves preparation of speech sound units that will be executed as continuous sequences to produce utterances. The planning process requires that brain regions encode the structural characteristics of the intended utterances, such as syllable frames and phonemic content as well as the sequential organization needed for fluent execution. Existing neuroimaging studies of speech production have suggested the occurrence of this planning process in prefrontal regions, with representations of speech units maintained and organized before being transmitted to motor execution areas for sequence production.

In this work, we sought to characterize and distinguish the neural representations for speech planning and execution (**Fig. 1a,b**)[11,13]. We performed a speech repetition task with constructed pseudo-words to isolate and characterize planning and execution circuits involved in speech production (**Fig. 1d**). The repetition of pseudo-words involves speech planning that retrieves the sublexical contents and transforms this representation into a speech plan. Our use of pseudowords allows us to examine sub-lexical speech representations independent of lexical and semantic influences. In this context, syllables serve as organizational units that provide a structural framework while the phonemes serve as organizational content for speech planning and execution. Behaviorally, we found that utterances with two syllables demonstrated longer spoken durations as compared to utterances with one syllable, which is consistent with their syllabic structure. Further, we also found that these utterances were associated with longer response times, suggesting that syllable structure reflects part of the planning process prior to speech execution (**Fig. 1f, g**).

Through the examination of speech neural activations in the HG band, we characterized a broad network of neuro-anatomical ROIs in the frontal, parietal, and temporal regions that were active during speech production (**Fig. 2**). Our ability to directly record cortical activations at high temporal resolution allowed us to precisely track the timing of these activations and separate planning from execution processes (**Fig. 2b**). Specifically, we identified sequential neural activations (time to reach maximum HG power) in three networks of regions corresponding to their planning, motor articulation, and monitoring during speech production (**Fig. 2c**). These speech networks were further validated in a data-driven way using an unsupervised decomposition analysis that revealed five distinct temporal components without prior knowledge of anatomical organization (Supplementary Fig. 3). This convergence validates that our observed temporal progression reflects functional specialization rather than arbitrary anatomical clustering.

Planning activations were observed in pre-frontal regions that included the inferior frontal gyrus (IFG), and the rostral and caudal middle frontal gyrus (MFG). The unsupervised decomposition analysis

confirmed that these regions showed maximal loading on the planning component (Supplementary Fig. 3, cluster 3), providing data-driven support. In previous literature, the IFG and MFG were primarily implicated in different aspects of speech planning including conversation engagement[65], and retrieval of lexical information[64]. In this current study, using controlled stimuli in a single pseudo-word repetition task, we show that planning neural activations reflect the encoding of the speech components necessary for execution. This controlled setting allowed us to study the neural operations involved in planning sub-lexical speech features (syllables and phonemes) for execution and isolate them from higher order language processes such as lexical, semantic, and syntactic information. Neural activations associated with articulation were observed later in time, in the sensorimotor (precentral and postcentral gyrus) and inferior parietal cortices. This articulation class of regions was also isolated using an unsupervised decomposition (cluster 2 – Supplementary Fig. 3), and have previously been shown to code for phoneme articulatory sequences to control vocal tract gestures[14,20,58,74], as well as during phonological processing and internal speech representation[75,76]. These activations therefore likely reflect the motor commands necessary to execute phonological sequences. Neural activity in these motor articulatory regions have also been used to drive brain-computer interfaces; therefore, effective characterization of articulation processes is essential for accurate speech neuro-prostheses[60,63,77–81]. Finally, late activations associated with monitoring were localized to temporal regions that included primary auditory cortex, posterior and anterior superior temporal gyrus, as well as the insula. The relatively late activations are consistent with studies that examined the role of monitoring regions to generate feedback signals in order to reduce the sensitivity to self-generated sounds and to update motor commands by anticipating the sound and motor/somatosensory information produced during speech[68,82–84]. In addition to monitoring, neural activations in STG and STS have been hypothesized to code for auditory targets during planning which is in keeping with the early activations of STS prior to response onset[85,86]. These activations corresponded to two distinct temporal components: early (cluster 1 – Supplementary Fig. 3) and later (cluster 4 – Supplementary Fig. 3) auditory monitoring, suggesting that monitoring networks may have multiple complex dynamics, the full functional implications of which will be investigated in future work. We further acknowledge that the neural activations in articulation and monitoring regions likely reflect a combination of motor execution and sensory feedback processes, making it challenging to dissociate execution from monitoring. Overall, HG activations demonstrate a hierarchy of prefrontal, sensorimotor, and temporal regions that reflect the planning, articulation, and monitoring stages of speech (**Fig. 2e**).

We next analyzed different types of speech units for planning and execution. For the syllable, we estimated the time-epochs with significant syllable coding (Supplementary Fig. 4) as measured by contrasting HG responses between monosyllabic and disyllabic utterances. This allowed us to identify

time-points with significant contrasts indicative of syllabic coding. We show that syllable information is present in the planning network regions (IFG – pars opercularis, rMFG, cMFG) prior to articulation, demonstrating that syllables are coded as part of the representation for speech planning. To address potential confounds between syllable class (CVC vs. VCV) and initial phoneme category (C vs. V), we performed a control analysis comparing CVCVC and CVC syllabic structures, which differ in syllable number but share the same initial consonantal phoneme structure (Supplementary Fig. 7). By demonstrating a qualitatively similar time course as our main findings, these analyses ruled out the possible influence of initial phoneme category affecting the timing of the neural processes, supporting the presence of syllable level discrimination prior to articulation. Our study differed from previous neuro-imaging research on syllable sequencing as we explicitly showed that syllable coding is present in the planning stage prior to speech execution (**Fig. 3e-f)**[11,41,42]. This finding is in accordance with previous modeling studies, which predicted the origin of syllabic codes in pre-frontal regions[4,11,87]. Lastly, the syllabic coding in articulation and monitoring network regions occurred later in time across both hemispheres, indicating that the flow of syllabic information occurs across different speech stages.

We then used high-resolution μECoG recordings (1.3 – 1.7 mm inter-electrode distance) to independently quantify this flow of syllable information at a patient specific level. Simultaneous recordings from left-hemispheric prefrontal and sensorimotor in two patients revealed distinct population sites along the anterior regions that showed earlier syllabic coding during planning, followed by progressively later activations into posterior motor regions during execution (**Fig. 4**). This 'spatial gradient' provides a potential anatomical substrate for the transition from planning to execution through the sequential activation of neural populations with syllable information. Previous planning studies using ECoG identified similar spatial structures within IFG and MFG and revealed differing levels of speech selectivity[65,82]. Our results identified a similar network selective to syllable coding and their transition from planning to execution, providing novel evidence for the role of the syllable as a unit of representation in the neurobiology of speech production. Future work will focus on extracting functional connectivity between these sites and quantifying the relevance of syllable coding to these connections.

Speech production is not only dependent on syllabic structure, but it also must involve the encoding of phonemic content within the syllable for the successful execution of speech during motor articulation. Speech computational models, including the DIVA (Directions Into Velocities of Articulators), GODIVA (Gradient Order DIVA), and Hickok's hierarchical state feedback control (HSFC) models, have theorized the presence of a hierarchical organization between syllabic and phoneme activations which reflects the sequential execution of speech sounds (**Fig. 5a**)[13,14,16,27]. These models propose that higher-level planning states code speech at the syllabic level while lower-level states operate at the phoneme level. Further, each model proposed unique computational mechanisms supporting this

operation, with the DIVA model emphasizing parallel feedforward and feedback control systems[27], GODIVA extending this to include sequential structure buffers and phonological content buffers[16,42], and Hickok's HSFC model further integrating auditory targets with somatosensory targets through hierarchical state feedback control[14]. In our study, we specifically designed the stimulus to provide direct, data-driven neural evidence supporting key claims of these models in addition to revealing neural timing properties that were not explicitly studied. Specifically, we show the existence of phoneme information that follows syllabic coding with the help of neural decoding models by resolving the emergence of syllables and phonemes in time during speech planning and execution. At the network level and within individual ROIs, we find that the syllabic information is present prior to phoneme information for speech production. This temporal relationship indicates that speech neuro-anatomical areas operate at both the syllabic and phonemic levels (**Fig. 5b-g**). This nested pattern of phoneme representation within broader syllable representations broadly supports the hierarchical architectures of DIVA, GODIVA, and HSFC model.

Interestingly, our findings extend beyond existing models by revealing a temporal relationship that exhibited compression along the three anatomical networks. Specifically, the planning network regions showed a large temporal lag between syllable and phoneme activations within the planning network (250-350 ms lag), followed by a shorter lag for the articulation network (110-130 ms), and nearly concurrent temporal activation in the monitoring regions (17-25 ms). This difference in temporal lags suggests that 'temporal compression' may act as a basis for the integration of different speech representations (syllables and phonemes) during the transition from speech planning to execution (**Fig. 5d-e**). Overall, the hierarchical relationship we observe—with syllables coded earlier in planning regions and phonemes emerging later in execution regions—suggests that syllables provide a scaffold for organizing phonemic content. Without lexical information, these sub-lexical units maintain their organizational properties, indicating that the neural architecture for speech production operates effectively at multiple levels of processing.

While our findings are consistent with the hierarchical structure and temporal progression as identified by these models, they do not provide data-driven evidence for several of their core mechanistic claims. Our results did not demonstrate the neural implementation of feedback control and error correction loops as predicted by DIVA and HSFC, and the existence of competitive queuing mechanisms as proposed by GODIVA. Further, the decoding results do not reveal the neural correlates of latent neural features such as efference copy or predictive sensory targets, as proposed by Hickok's HSFC simulations[14]. Future research could identify these mechanistic predictions using experimental manipulations that specifically modulate model-predicted behaviors and by advancing the decoding work with cross-temporal analyses to identify neural architectures involved in this hierarchical

organization (e.g. maintenance vs. evolution of speech neural features)[88]. Overall, our findings provide essential timing information that can inform next-generation computational models to incorporate the real-time dynamics of sublexical features from planning to execution.

Speech articulation involves the precise and correct ordering of articulatory movements, and consequently the phonological units must be dynamically encoded sequentially during speech execution (**Fig. 6a**)[13]. We demonstrated the sequential generation of phonological units by accurately decoding phoneme sequences in time. These phonological sequence representations were found in the articulation and monitoring network in both hemispheres but were absent in the planning network (**Fig. 6b,** Supplementary Fig. 13). This absence of sequential phonological decoding could reflect several factors: the phonological code during planning may be compressed in time, encoded in a manner not accessible with our current decoding approach, or may reflect signal-to-noise limitations. However, our ability to robustly decode syllable identity during the same planning period suggests that relevant speech information is indeed present in the neural signal, but that specific temporal organization of phonemes may not be as explicitly represented during planning stages.

We observed significant speech sequencing in both the articulation and the monitoring network. The presence of sequencing in the articulation regions is in keeping with previous studies that examined phoneme articulatory networks in laryngeal motor cortex[58,74] (ventral) and vocalization properties in the mid pre-central gyrus[66,89] (dorsal), suggesting that phonological sequencing could be spatially distributed. Although our results focused on how the entire sensorimotor system is involved in generating phonological unit sequences, future studies could involve high-resolution electrodes to delineate dorsal-ventral circuitry to better explain the discrete-linguistic and continuous motor-control properties of speech. The spatio-temporal distinctive characteristics of sensorimotor systems can further inform the design of speech-BCI algorithms to simultaneously model the spatial gradient[62] and temporal sequencing[77–79,90] of speech elements for accurate speech neuro-prostheses. The presence of neural activations sensitive to sequencing in the monitoring network regions could be attributed to both perception of self-generated sounds[68] and comparison of sounds with the generation of auditory efferent copies for intended speech[14]. The separation of these two elements could be examined in future work by identifying causal connections between monitoring and articulation regions and using these connections to help explain how the current state of phonological unit generation is modulated by inputs from auditory feedback.

In this work, we also sought to characterize one of the mechanisms that can facilitate articulatory sequencing during speech production. Adult speakers have entrained sensorimotor systems through years of language practice to enable fluent articulation of linguistically familiar speech sounds[4,8,16,41,73].

This linguistic tuning can be further quantified by measuring the statistical properties between phonemes: phonotactics[73]. Although the sensory (auditory) encoding of phonotactic distributions have been examined during speech perception[72], the motor component of internalized phonotactics has not yet been established. In this study, the use of phonologically constructed pseudo-words allowed us to calculate the statistical transitionary properties between phonemes. Using a decoding approach from high-resolution recordings in sensorimotor cortex, we observed robust and stable neural activation patterns associated with phonotactic transitions. Further, these neural activations occurred temporally between neural activations associated with phoneme positions (**Fig. 7)**. These findings suggest that during articulation, motor activity is organized (at least in part) continuously and holistically rather than as a set of discrete phoneme units. It also provides evidence for neural tracking of speech motor activations through phonotactics, which can enable continuous and fluent execution of speech sounds learned from years of native language use. Future studies could combine high-resolution recordings and stimuli with more complex syllabic structures to further study how sensorimotor systems plan and execute speech.

While our study delivers high-resolution temporal maps of speech planning, execution, and monitoring networks, it remains inherently correlational and does not establish causal links between neural representations across regions. Further, the role of somatosensory feedback in actively monitoring the speech commands remain quite unexplored in this study paradigm. The identification of sustained component (Supplementary Fig. 3; cluster 5) and the dual monitoring components (cluster 1 and 4), suggest that these representations are anatomically distributed and require further experimental probing to identify their functional roles. Future work should employ explicit neural content coding and causality analyses (e.g., perturbation protocols, directed connectivity modeling) to characterize the directional flow of information between ROIs. Finally, incorporating phonologically manipulated words into our paradigm by varying their lexical properties (phonological neighborhood, phonotactic transitions, etc) will reveal how higher-order language cues shape sub-lexical planning and execution, thus bridging the gap between abstract phonological representations and lexical access, during speech planning[4].

Overall, this work highlights the key neural substrates underlying planning and execution of speech during production. The usage of intracranial and high-resolution recordings enabled us to identify both the spatio-temporal neural activations as well as their transitions, that occurred across and within speech networks. These patterns were specific to different levels of processing, with sound units processed for planning, and continuous sequences executed during articulation. This work could have important implications for the development of therapeutic interventions for patients with speech and language disorders, where the ability to plan and implement speech coordinated activity is disrupted.

3 Methods

This study was performed at the Duke University Health System with approval from the Institutional Review Board under the protocol IDs:

Pro00065476: Network dynamics of human cortical processing.

Pro00072892: Studying human cognition and neurological disorders using μECoG electrodes.

3.1 Participants

Intracranial recordings with standard clinical intracranial recording systems were obtained from 52 patients with pharmaco-resistant epilepsy at Duke University Medical Center (DUMC) (mean age = 33.1 years, 23 Female, Supplementary table 1). Patients were implanted with clinical intracranial electro-encephalography (IEEG) systems as part of their work up for the neurosurgical treatment of pharmacologically resistant epilepsy. Recordings using high-density micro-scale cortical recordings were obtained in an intraoperative setting from 3 adult patients (mean age = 64.3 years, 0 Female), during awake neuro-surgical procedures. Patients were awoken as part of their surgical procedure and the decision to have them awake was entirely clinically motivated. Two of the three patients (S1, and S3) underwent the placement of a deep brain stimulator (DBS) for the treatment of movement disorders. The other patient (S2) underwent surgical resection for tumor removal. All patients were native English speakers and written informed consent was obtained in accordance with the Institutional Review Boards at the DUMC.

3.2 Electrode implantation, localization and neural recordings

Cortical recordings were obtained using electrodes from either one of the three recording systems: 1) macro-electrocorticographic grids (macro-ECoG: 2 participants), 2) stereo-electroencephalographic probes (SEEG: 50 participants), and 3) liquid crystal polymer thin-film (LCP-TF) micro-electrocorticographic arrays (μECoG: 3 participants). Macro-ECoG grid recordings were done using Ad-Tech electrode grids (48 or 64 electrodes; Ad-Tech Medical Instrument Corporation) with 10 mm electrode spacing and 2.3 mm exposed diameter, surgically implanted onto the cortex of the patients. SEEG placement was done via robotically guided surgery (ROSA Robotics) using either PMT (PMT Corporation), Ad-Tech, or Dixi (Dixi Medical) SEEG electrodes with 3.5 mm to 5 mm inter-electrode spacing (0.8 to 0.86 mm exposed diameter). μECoG recording included LCP-TF μECoG arrays (1.33 – 1.72 mm electrode spacing; 200 μm diameter), acutely implanted either through a burr hole (DBS surgery) or directly over the cortical surface (awake craniotomy).

Pre-operative anatomical 3T MRI were obtained for all patients, and post-operative CT images were obtained for both macro IEEG and DBS patients to identify electrode locations. Patients undergoing

DBS surgery underwent a routine intraoperative CT for assessment of DBS electrode placement with the subdural ECoG array in place. For the tumor resection patient, BrainLab coordinates[91,92] were obtained at the corner of the electrode array and this information was then mapped to three-dimensional reconstruction of the patient's brain obtained from their preoperative MRI. For each patient, electrode locations were manually marked using BioImage Suite, and were mapped on to the reconstructed T1 volume generated using Freesurfer[93,94]. Electrodes were then assigned to specific regions of interest (ROIs), based on the cortical parcellations obtained from connectivity based Brainnetome Atlas[95]. For the electrodes that were not centered at specific ROIs (e.g. depth electrodes from SEEG probes near/in white matter), a sphere with a 10 mm radius was constructed around the electrode's center, and the relative density of ROIs was determined through the calculation of the percentage proportion of ROI voxels within this sphere. Electrodes were then assigned to a specific ROI with the maximum density if the proportion exceeded 5% (otherwise they were labelled as white matter contacts). For visualization, electrode coordinates were also transformed to a population averaged MNI brain and color-coded for ROIs.

IEEG recordings in the clinical monitoring unit (ECoG and SEEG) were analog filtered between 0.01 Hz and 1000Hz and digitized at 2,048 samples per second, using a Natus Quantum LTM amplifier (with Natus Neuroworks EEG software). Neural recordings from μECoG electrodes (intra-operative) were analog filtered between 0.1 Hz and 7500 Hz and digitized at 20,000 samples per second (with Intan RHX data acquisition software, v3.0; Intan Technologies, Inc).

3.3 Delayed speech repetition task

Patients performed a delayed pseudo-word repetition task in the clinical monitoring unit. In each trial, they listened and repeated one of 52 pseudo-words that were modelled on the phonology of American English. Pseudo-words were equally distributed between monosyllabic and di-syllabic structures (consonant-vowel-consonant - CVC: 26) and vowel-consonant-vowel (VCV: 26 tokens) and included a total of 9 phonemes (4 vowels: /a/, /æ/, /i/, /u/ and 5 consonants: /b/, /p/, /v/, /g/, /k/). The phonemes were arranged within the pseudo-words (CVC or VCV) by selecting biphones (CV or VC) with a positive likelihood of occurring in the English language corpus[96]. The third phoneme was selected with the constraint that the occurrence probability of the triphone sequence was zero, resulting in a constructed pseudo-word. For each pseudoword, forward biphoneme transition probabilities were calculated based on the occurrence probabilities of individual phonemes [P(C) or P(V)] and co-occurring biphones [P(CV) and P(VC)] as follows:

$P(V|C) = \frac{P(CV)}{P(C)}$, and $P(C|V) = \frac{P(VC)}{P(V)}$.

The resultant transitional probability scores across the pseudoword tokens were quantized into three classes: low, medium, and high transitions scores (P(V|C): low – 0.0024 to 0.034, medium – 0.034 to 0.1031, high – 0.1031 to 0.1791; P(C|V): low – 0.0083 to 0.0314, medium – 0.0314 to 0.0543, high – 0.0543 to 0.0814).

The constructed pseudo-words were auditorily presented and subsequently cued for the patients to repeat via a visual Go Cue ('speak'). Each trial lasted between 5.1 – 6.1 seconds, starting with a visual "Listen" cue followed by the auditory stimulus presentation (mean stimulus duration: 0.5 second). After a 1.5 s delay (150 ms jitter), the "speak" (Go) cue was visually presented and the patients were given 3 s (200 ms jitter) to respond, before the next trial began. Trials were presented in four consecutive blocks with 52 unique pseudo-word tokens shuffled in each block and the overall task time with 208 speech trials took less than 40 minutes in the monitoring unit. A subset of patients underwent a similar delayed repetition task where they repeated pseudo-words with di-syllabic structures (CVCVC). The patients in the intra-operative setting underwent a similar version of pseudo-word repetition task, but with the delay and "speak" cue removed for simplicity. Intraoperative patients completed 3 to 4 blocks of 52 trials (156 – 208 trials). The speech task was designed using Psychtoolbox scripts in MATLAB R2014a.

3.4 Audio recording and behavioral representation

The speech repetition task included both auditory and speaking components. The auditory stimulus wave files were presented using a laptop connected to a powered stereo-speaker (amazonbasics A100 series) through a USB DAC and audio amplifier (Fiio). The beginning of each trial and the subsequent stimulus onsets were marked by flashing white circles on the stimulus laptop screen which was detected by a photodiode (Thorlabs, Inc.) attached to an auxiliary BNC-to-mono TS cable connected to the Natus Recording system. The audio waveforms were recorded using a condenser microphone (Marantz Professional, MPM-1000) which was connected to pre-amplifier (Behringer) and digitized at 44100 Hz on the recording laptop. The auditory stimulus, the spoken pseudo-word onsets, and the offsets were either manually identified using the Audacity software package, with the built-in Mel-spectrogram visualization feature or automated through the Montreal Forced Aligner (Montreal Corpus Tools).

For behavior analysis, each trial was characterized by its response time (with respect to Go cue onset) and its response duration. For the neural analysis, the trials were categorically coded for their syllabic structure (CVC or VCV), the phonemes (9-way) at each position, and the corresponding forward transition category (3-way) for each biphoneme pair within its structure. Trials with no intelligible responses were removed from the analysis.

3.5 Behavioral representations of syllabic coding

A linear mixed effects (LME) model was used to predict the effects of syllable on response time and response duration. The models incorporated fixed effects for categorical syllabic IDs (one vs. two syllables) and random effects for patient and an intercept. Two sets of models were utilized separately to predict response time and response duration respectively, and in each case the estimate of syllable (beta coefficient for fixed effect) was computed. The significance of the beta estimates was computed using a permutation test by randomly shuffling syllables (1000 shuffles) and repeating the LME prediction for each shuffle.

3.6 Neural preprocessing and signal analysis

The neural preprocessing steps were performed offline. The IEEG recordings were filtered for line-noise with multi-taper band-stop notches centered at 60, 120, 180, and 240 Hz. The following categories of electrodes were removed from subsequent analyses: 1) electrodes that were outside the cortex (determined from localization), 2) electrodes that exhibited muscle artifacts during the utterance (as visualized by a wavelet scalogram), and 3) electrodes with recording power greater than 3 * log-RMS. From a total of 9679 electrodes, 1573 were excluded using the above steps. The remaining set of 8106 electrodes were subjected to common-average-referencing (CAR) and utilized for further processing.

3.6.1 Neural feature extraction

Analyses were performed to extract neural high-gamma signals (HG: 70 – 150 Hz) from each electrode. HG signals have previously been shown to characterize speech activations from cortical regions as well as demonstrating a high correlation with multi-unit firing[54,56,57]. The pre-processed neural signals from each IEEG electrode were band-pass filtered into 8 center frequencies (logarithmically arranged) between 70 and 150 Hz[97]. The envelope of each frequency band was extracted using the Hilbert transform and was averaged across the bands. The resultant envelope was then downsampled to 200 Hz using an anti-aliasing filter and z-scored with respect to a baseline period (-500 ms to 0 s before trial start). For each trial, we defined the response epoch as the time intervals from -1500 ms to 1500 ms with respect to the utterance onset.

3.6.2 Non-negative matrix factorization

To perform a data-driven quantification of neural sub-components underlying speech production, we performed non-negative matrix factorization (NNMF) on the neural HG power[98]. HG activations (***X***) were pooled across all speech-responsive electrodes and were averaged across the trials with respect to the utterance onset (from -1.5 seconds to +1.5 seconds). NNMF was used to decompose the electrode-time matrix into two non-negative matrices, $X \approx H \times W$, where ***H*** contains the electrode

weights (channels x components) and ***W*** captures the temporal profiles of each component (components x timepoints). Factorization was performed using an alternating least-squares algorithm with non-negativity constraints on both ***H*** and ***W***. The optimal number of components (k) was determined by evaluating the cumulative variance explained at each candidate k and selecting the factor that accounted for at least 90% of the data variance.

3.6.3 Neural representations of syllable coding

A non-parametric statistical test was performed to test whether HG power from each ROI could differentiate between syllables (CVC vs. VCV) at each time-point. For each electrode, the absolute difference in HG power (ΔHG) was computed between VCV trials and CVC trials. The resultant difference (across-syllables) was averaged across electrodes within a given ROI (within patient followed by across patient average) and was considered as the test statistic. For a control analysis, ΔHG was computed within CVC trials and within VCV trials, and the average within-syllable ΔHG was determined for each ROI. The null distribution was generated from 1000 iterations of the within-syllable statistic for each ROI, and the significance was determined by calculating the proportion of null-distribution (within-syllable) less than the test-statistic (across-syllable) ($p < 0.05$). The resultant p-value time-series were subjected to a second-level non-parametric cluster correction to determine consecutive sequence of time-points that exhibited significant syllable coding.

To rule out the confounding of initial phonemes (C vs. V) in the neural syllable code, a control analysis was performed in a subset of patients who underwent a modified version of speech repetition task, where they repeated pseudowords with di-syllabic structures (CVCVC). Similar to the above method, the absolute difference in HG power was computed between CVCVC trials and CVC trials as the test statistic and was subjected to second-level non-parametric cluster correction.

3.6.4 Linear classification model for speech decoding

A previously established linear-classification model with 10-fold cross-validation, was used to decode speech units from normalized HG[63]. First, we performed a singular value decomposition (SVD) to compress HG to low-dimensional features. We chose the number of kept components as the total number of components that explained 80% of the variance. We utilized a supervised linear-discriminant (LDA) model to predict speech units. The decoding performance was estimated by calculating the average prediction accuracy in the test set across all folds.

Multivariate decoding was performed by pooling HG epochs from electrodes across patients. To perform this pooling, trials with similar tokens across patients were concatenated together. In instances where there was a mismatch in the number of trials across patients for a particular token, a synthetic trial was generated using a Mix-Up technique through the linear combination of two randomly selected

trials for that token[99]. This Mix-Up approach was performed separately on the training and the testing set to avoid data leakage.

To infer temporal characteristics of speech features, decoding was performed in the temporal domain by training and testing at each time-segment (200 ms window; 10 ms step size) within the epoch[88]. The decoding accuracy was calculated at each time-segment to characterize temporally evolving representations of speech features in the response epoch. Separate temporal models were evaluated per ROI to decode syllables (2-way, chance: 50%), phoneme units (9-way, chance: 11.11%) in each position (P1, P2, and P3), and phoneme transition probabilities (3-way, chance: 33.33%) for each biphone (P_{fwd1}, and P_{fwd2}) within the pseudo-word. The resultant accuracy (Acc_{Test}), as a percentage, was normalized (Acc_{Norm}) with respect to theoretical chance (Acc_{Chance}) for visualization purposes as follows:

$$Acc_{norm} = \frac{Acc_{Test} - Acc_{Chance}}{100 - Acc_{Chance}}$$

3.6.5 Statistical analysis

Multiple statistical tests were conducted to validate IEEG recordings for speech representations. To identify electrodes with significant speech-neural activations, signal power in the HG band was compared between the response (-500 ms to 500 ms) and the baseline period using a 1000-iteration one-sided permutation test (corrected for false discovery rate).

To quantify the neural morphology of HG responses for each ROI, the above non-parametric test was extended to determine the significance at each time point (against baseline), with HG power averaged across electrodes within the ROI. In addition, the HG onset times, and peak times were calculated for each electrode. The onset times were estimated by performing a change-point analysis which determines the time-point at which the mean of the HG time-series changes most significantly (findchangepts algorithm in MATLAB). Peak times were determined by either finding the time-point where HG activations achieved the maximum value or by computing the weighted mean of HG activations over time. The distribution of HG onset and peak times were used to quantify the evoked temporal characteristics of a speech responsive electrode. Significant differences in onset or peak times were evaluated using pairwise Mann-Whitney U test, or student t-tests.

For speech decoding, the prediction accuracies were validated against the null distribution obtained from random shuffling of the predictor labels (250 iteration). ROI characteristic decoding onsets were obtained using a similar change-point analysis. The resultant time-series was subjected to second-level non-parametric temporal cluster correction.

Data availability

The data from IEEG recordings that support neural analysis will be archived via the Data Archive for The Brain Initiative (DABI) under the restricted access upon publication. The access can be obtained upon written requests from the corresponding authors (gregory.cogan@duke.edu, j.viventi@duke.edu).

Code availability

The MATLAB files to perform neural analysis will be updated upon publication through GitHub at: https://github.com/coganlab/IEEG_Pipelines.

References

1. Guenther, F. H., Hampson, M. & Johnson, D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev.* **105**, 611 (1998).
2. Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science (80-.).* **279**, 1213–1216 (1998).
3. Brandi, M.-L., Wohlschläger, A., Sorg, C. & Hermsdörfer, J. The Neural Correlates of Planning and Executing Actual Tool Use. *J. Neurosci.* **34**, 13183 LP – 13194 (2014).
4. Indefrey, P. & Levelt, W. J. M. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).
5. Levelt, W. J. M., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–38 (1999).
6. Roelofs, A. The WEAVER model of word-form encoding in speech production. *Cognition* **64**, 249–284 (1997).
7. Fiez, J. A., Balota, D. A., Raichle, M. E. & Petersen, S. E. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron* **24**, 205–218 (1999).
8. Mechelli, A., Friston, K. J. & Price, C. J. The effects of presentation rate during word and pseudoword reading: a comparison of PET and fMRI. *J. Cogn. Neurosci.* **12**, 145–156 (2000).
9. Fromkin, V. A. The non-anomalous nature of anomalous utterances. *Language (Baltim).* **47**, 27–52 (1971).
10. Shattuck-Hufnagel, S. The role of word structure in segmental serial ordering. *Cognition* **42**, 213–259 (1992).
11. Rong, F., Isenberg, A. L., Sun, E. & Hickok, G. The neuroanatomy of speech sequencing at the syllable level. *PLoS One* **13**, e0196381 (2018).
12. Geschwind, N. The organization of language and the brain. *Science* **170**, 940–944 (1970).
13. Hickok, G. The architecture of speech production and the role of the phoneme in speech processing. *Lang. Cogn. Neurosci.* **29**, 2–20 (2014).
14. Hickok, G. Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* **13**, 135–145 (2012).
15. Houde, J. F. & Nagarajan, S. S. Speech production as state feedback control. *Front. Hum. Neurosci.* **5**, 82 (2011).

16. Bohland, J. W., Bullock, D. & Guenther, F. H. Neural representations and mechanisms for the performance of simple speech sequences. *J. Cogn. Neurosci.* **22**, 1504–1529 (2010).

17. Baars, B. J., Motley, M. T. & MacKay, D. G. Output editing for lexical status in artificially elicited slips of the tongue. *J. Verbal Learning Verbal Behav.* **14**, 382–391 (1975).

18. Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* **104**, 801 (1997).

19. Cogan, G. B. *et al.* Manipulating stored phonological input during verbal working memory. *Nat. Neurosci.* **20**, 279–286 (2017).

20. Cogan, G. B. *et al.* Sensory-motor transformations for speech occur bilaterally. *Nature* **507**, 94–8 (2014).

21. Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R. & Fox, P. T. Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp.* **25**, 105–117 (2005).

22. Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J. & Fedorenko, E. Functional characterization of the human speech articulation network. *Cereb. Cortex* **28**, 1816–1830 (2018).

23. Broca, P. Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bull. Mem. la Soc. Anat. Paris* **6**, 330–357 (1861).

24. Tourville, J. A., Nieto-Castañón, A., Heyne, M. & Guenther, F. H. Functional parcellation of the speech production cortex. *J. Speech, Lang. Hear. Res.* **62**, 3055–3070 (2019).

25. Wise, R. J. S., Greene, J., Büchel, C. & Scott, S. K. Brain regions involved in articulation. *Lancet* **353**, 1057–1061 (1999).

26. Turkeltaub, P. E., Eden, G. F., Jones, K. M. & Zeffiro, T. A. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* **16**, 765–780 (2002).

27. Tourville, J. A. & Guenther, F. H. The DIVA model: A neural theory of speech acquisition and production. *Lang. Cogn. Process.* **26**, 952–981 (2011).

28. Indefrey, P. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* **2**, 255 (2011).

29. Guenther, F. H. *Neural control of speech*. (Mit Press, 2016).

30. Eickhoff, S. B., Heim, S., Zilles, K. & Amunts, K. A systems perspective on the effective connectivity of overt speech production. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **367**, 2399–2421 (2009).

31. Berger, M., Agha, N. S. & Gail, A. Wireless recording from unrestrained monkeys reveals motor goal encoding beyond immediate reach in frontoparietal cortex. *Elife* **9**, 305334 (2020).

32. Gale, D. J., Flanagan, J. R. & Gallivan, J. P. Human somatosensory cortex is modulated during motor planning. *J. Neurosci.* **41**, 5909–5922 (2021).

33. Sober, S. J. & Sabes, P. N. Flexible strategies for sensory integration during motor planning. *Nat. Neurosci.* **8**, 490–497 (2005).

34. Crevecoeur, F., Cluff, T. & Scott, S. H. Computational approaches for goal-directed movement planning and execution. (2014).

35. Perkell, J. *et al.* Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Commun.* **22**, 227–250 (1997).

36. Perkell, J. S. Movement goals and feedback and feedforward control mechanisms in speech production. *J. Neurolinguistics* **25**, 382–407 (2012).

37. McGuire, P. K., Silbersweig, D. A. & Frith, C. D. Functional neuroanatomy of verbal self-monitoring. *Brain* **119**, 907–917 (1996).

38. Hirano, S. *et al.* Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport* **8**, 2379–2382 (1997).

39. Wheeldon, L. R. & Levelt, W. J. M. Monitoring the time course of phonological encoding. *J. Mem. Lang.* **34**, 311–334 (1995).

40. Levelt, W. J. M. Monitoring and self-repair in speech. *Cognition* **14**, 41–104 (1983).

41. Peeva, M. G. *et al.* Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage* **50**, 626–638 (2010).

42. Bohland, J. W. & Guenther, F. H. An fMRI investigation of syllable sequence production. *Neuroimage* **32**, 821–841 (2006).

43. Segawa, J. A., Tourville, J. A., Beal, D. S. & Guenther, F. H. The neural correlates of speech motor sequence learning. *J. Cogn. Neurosci.* **27**, 819–831 (2015).

44. Parvizi, J. & Kastner, S. Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* **21**, 474–483 (2018).

45. Lachaux, J. P., Rudrauf, D. & Kahane, P. Intracranial EEG and human brain mapping. *J. Physiol.* **97**, 613–628 (2003).

46. Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E. & Crone, N. E. High-frequency neural activity and human cognition: Past, present and possible future of intracranial EEG research. *Prog. Neurobiol.* **98**, 279–301 (2012).

47. Mullin, J. P. *et al.* Is SEEG safe? A systematic review and meta-analysis of stereo-electroencephalography–related complications. *Epilepsia* **57**, 386–401 (2016).

48. Chiang, C.-H. *et al.* Development of a neural interface for high-definition, long-term recording in rodents and nonhuman primates. *Sci. Transl. Med.* **12**, eaay4682 (2020).

49. Chiang, C.-H. *et al.* Flexible, high-resolution thin-film electrodes for human and animal neural research. *J. Neural Eng.* **18**, (2021).

50. Trumpis, M. *et al.* Sufficient sampling for kriging prediction of cortical potential in rat, monkey, and human $μ$ECoG. *Submiss.* (2020).

51. Sun, J. *et al.* Intraoperative microseizure detection using a high-density micro-electrocorticography electrode array. *Brain Commun.* **4**, fcac122 (2022).

52. Barth, K. J. *et al.* Flexible, high-resolution cortical arrays with large coverage capture microscale high-frequency oscillations in patients with epilepsy. *Epilepsia*.

53. Kellis, S. S., House, P. A., Thomson, K. E., Brown, R. & Greger, B. Human neocortical electrical activity recorded on nonpenetrating microwire arrays: applicability for neuroprostheses. *Neurosurg. Focus* **27**, E9 (2009).

54. Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J. & Hsiao, S. S. Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. *J. Neurosci.* **28**, 11526–11536 (2008).

55. Rowland, N. C. *et al.* Task-related activity in sensorimotor cortex in Parkinson's disease and

essential tremor: changes in beta and gamma bands. *Front. Hum. Neurosci.* **9**, (2015).

56. Dubey, A. & Ray, S. Cortical Electrocorticogram (ECoG) is a local signal. *J. Neurosci.* (2019).

57. Leonard, M. K. *et al.* Large-scale single-neuron speech sound encoding across the depth of human cortex. *Nature* **626**, 593–602 (2024).

58. Chartier, J., Anumanchipalli, G. K., Johnson, K. & Chang, E. F. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron* **98**, 1042–1054 (2018).

59. Cheung, C., Hamilton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor cortex. *Elife* **5**, e12577 (2016).

60. Silva, A. B., Littlejohn, K. T., Liu, J. R., Moses, D. A. & Chang, E. F. The speech neuroprosthesis. *Nat. Rev. Neurosci.* (2024) doi:10.1038/s41583-024-00819-9.

61. Muller, L., Hamilton, L. S., Edwards, E., Bouchard, K. E. & Chang, E. F. Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. *J. Neural Eng.* **13**, (2016).

62. Angrick, M. *et al.* Speech Synthesis from ECoG using Densely Connected 3D Convolutional Neural Networks. *Biorxiv* (2018).

63. Duraivel, S. *et al.* High-resolution neural recordings improve the accuracy of speech decoding. *Nat. Commun.* **14**, 6938 (2023).

64. Wang, Y., Korzeniewska, A., Usami, K., Valenzuela, A. & Crone, N. E. The dynamics of language network interactions in lexical selection: an intracranial EEG study. *Cereb. Cortex* **31**, 2058–2070 (2021).

65. Castellucci, G. A., Kovach, C. K., Howard, M. A., Greenlee, J. D. W. & Long, M. A. A speech planning network for interactive language use. *Nature* **602**, 117–122 (2022).

66. Liu, J. R., Zhao, L., Hullett, P. W. & Chang, E. F. Speech sequencing in the human precentral gyrus. *Nat. Hum. Behav.* (2025) doi:10.1038/s41562-025-02250-1.

67. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? (2007).

68. Flinker, A. *et al.* Single-Trial Speech Suppression of Auditory Cortex Activity in Humans. *J. Neurosci.* **30**, 16643 LP – 16650 (2010).

69. Mitchell, R. L. C. & Crow, T. J. Right hemisphere language functions and schizophrenia: the forgotten hemisphere? *Brain* **128**, 963–978 (2005).

70. Luks, T. L. *et al.* Atrophy in two attention networks is associated with performance on a Flanker task in neurodegenerative disease. *Neuropsychologia* **48**, 165–170 (2010).

71. van Turennout, M., Hagoort, P. & Brown, C. M. Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* vol. 23 787–806 at https://doi.org/10.1037/0278-7393.23.4.787 (1997).

72. Leonard, M. K., Bouchard, K. E., Tang, C. & Chang, E. F. Dynamic Encoding of Speech Sequence Probability in Human Temporal Cortex. *J. Neurosci.* **35**, 7203 LP – 7214 (2015).

73. Vitevitch, M. S., Luce, P. A., Charles-Luce, J. & Kemmerer, D. Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words. *Lang. Speech* **40**, 47–62 (1997).

74. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–32 (2013).

75. Oberhuber, M. *et al.* Four Functionally Distinct Regions in the Left Supramarginal Gyrus Support Word Processing. *Cereb. Cortex* **26**, 4212–4226 (2016).

76. Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M. & Lœvenbruck, H. What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behav. Brain Res.* **261**, 220–239 (2014).

77. Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498 (2019).

78. Makin, J. G., Moses, D. A. & Chang, E. F. Machine translation of cortical activity to text with an encoder–decoder framework. *Nat. Neurosci.* **23**, 575–582 (2020).

79. Sun, P., Anumanchipalli, G. K. & Chang, E. F. Brain2Char: a deep architecture for decoding text from brain recordings. *J. Neural Eng.* **17**, 66015 (2020).

80. Wilson, G. H. *et al.* Decoding spoken English from intracortical electrode arrays in dorsal precentral gyrus. *J. Neural Eng.* **17**, 66007 (2020).

81. Wandelt, S. K. *et al.* Representation of internal speech by single neurons in human supramarginal gyrus. *Nat. Hum. Behav.* (2024) doi:10.1038/s41562-024-01867-y.

82. Flinker, A. *et al.* Redefining the role of Broca's area in speech. *Proc. Natl. Acad. Sci.* **112**, 2871–2875 (2015).

83. Forseth, K. J., Hickok, G., Rollo, P. S. & Tandon, N. Language prediction mechanisms in human auditory cortex. *Nat. Commun.* **11**, 5240 (2020).

84. Hamilton, L. S., Edwards, E. & Chang, E. F. A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Curr. Biol.* **28**, 1860-1871.e4 (2018).

85. Nourski, K. V *et al.* Electrophysiology of the human superior temporal sulcus during speech processing. *Cereb. Cortex* **31**, 1131–1148 (2021).

86. Kriegstein, K. V & Giraud, A.-L. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* **22**, 948–955 (2004).

87. Khanna, A. R. *et al.* Single-neuronal elements of speech production in humans. *Nature* **626**, 603–610 (2024).

88. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).

89. Silva, A. B. *et al.* A Neurosurgical Functional Dissection of the Middle Precentral Gyrus during Speech Production. *J. Neurosci.* **42**, 8416 LP – 8426 (2022).

90. Livezey, J. A., Bouchard, K. E. & Chang, E. F. Deep learning as a tool for neural data analysis: Speech classification and cross-frequency coupling in human sensorimotor cortex. *PLOS Comput. Biol.* **15**, e1007091 (2019).

91. No Title. https://brainmappinglab.org/.

92. Gupta, D., Hill, N. J., Adamo, M. A., Ritaccio, A. & Schalk, G. Localizing ECoG electrodes on the cortical anatomy without post-implantation imaging. *NeuroImage. Clin.* **6**, 64–76 (2014).

93. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

94. Iglesias, J. E. *et al.* SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. *Sci. Adv.* **9**, eadd3607 (2023).

95. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).

96. Vitevitch, M. S. & Luce, P. A. A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods, Instruments, Comput.* **36**, 481–487 (2004).

97. Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–10 (2014).

98. Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P. & Plemmons, R. J. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.* **52**, 155–173 (2007).

99. Zhang, H., Cisse, M., Dauphin, Y. N. & Lopez-Paz, D. MixUp: Beyond empirical risk minimization. *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.* 1–13 (2018).

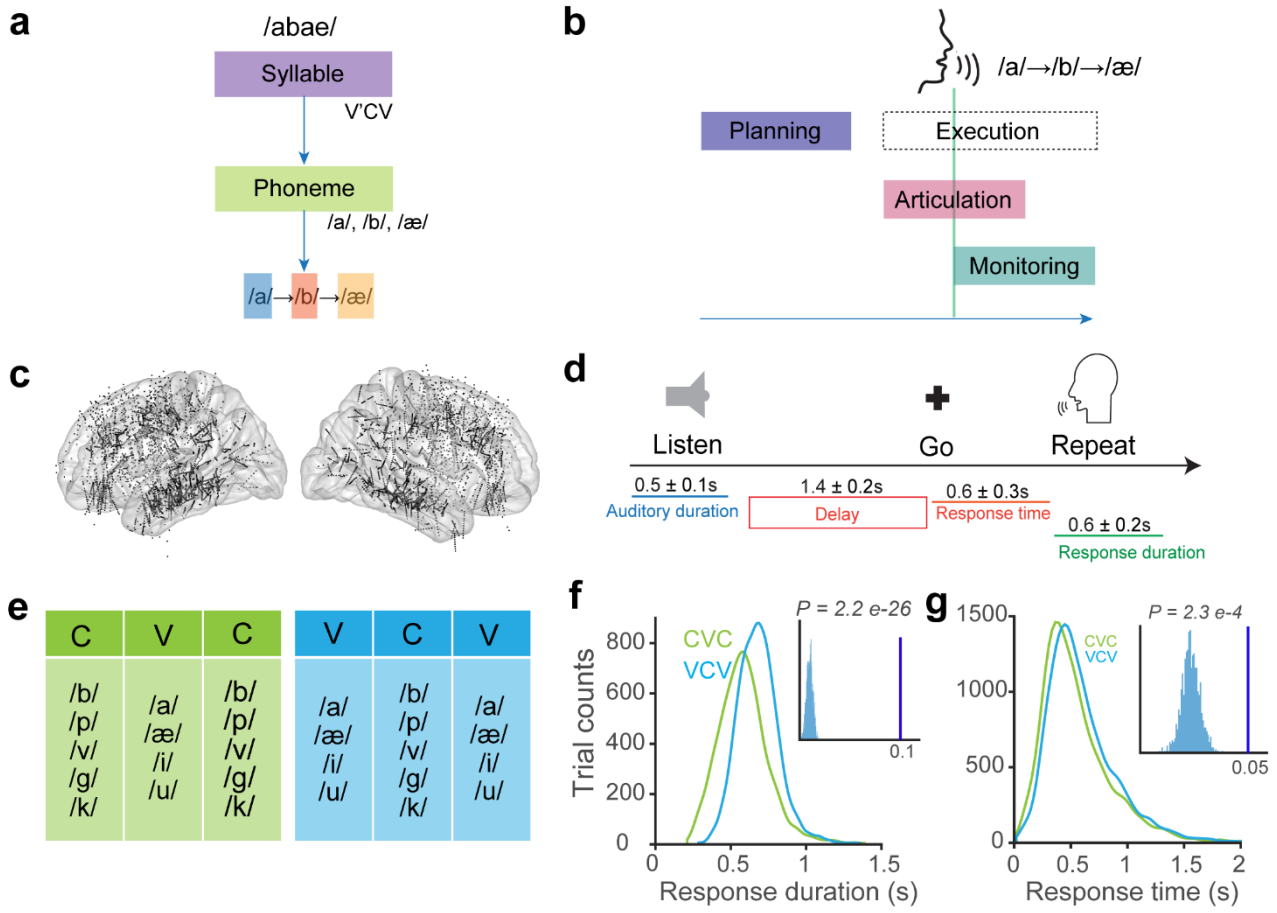

Figure 1. **Pseudo-word repetition task to study planning and execution for speech production. a.** Speech planning and execution at the sublexical level (post word formation) involves the composition of a syllable frame and the hierarchical incorporation of phonemes within this frame for the execution of the ordered speech sequence. **b.** Graphical representation of speech neural mechanisms from the planning of intended speech units to the execution of speech sound sequences. Speech production involves planning prior to execution, and execution itself which involves motor articulation followed by auditory monitoring of vocal feedback. **c.** Location of all 8106 recorded intracranial electrodes from 52 patients in both left and right hemispheres. **d.** Schematic of pseudo-word repetition task. Color bars indicate the duration of the auditory stimulus (blue), response time from the Go cue (orange), and response duration (green). **e.** Pseudo-word tokens used for the speech repetition task differed on their syllabic frame: consonant-vowel-consonant (CVC, green) and vowel-consonant-vowel (VCV, blue). The phonemes were chosen from a fixed for distribution for both consonants (5) and vowels (9) at each position within the pseudo-word. **f.** Histogram distributions of response duration for both CVC and VCV trials. The durations of VCV trials were significantly longer than the CVC trials (linear mixed effects [LME] modeling, $p = 2.2e\text{-}26$). The dark blue line indicates the beta estimate of the actual LME model, and the histograms (light blue) indicate the estimated distributions from permuted models (1000 permutations). **g.** Histogram distributions of response times for both CVC and VCV trials. The response time of CVC trials were significantly faster than the VCV trials (linear mixed effects [LME] modeling, $p = 2.3e\text{-}4$). The dark blue line indicates the estimate of the actual LME model, and the histograms (light blue) indicate the estimate distributions from permuted models (1000 permutations). Longer response times for di- vs. mono- syllabic utterances suggests that the syllable frame is used as a sub-lexical unit for speech planning.

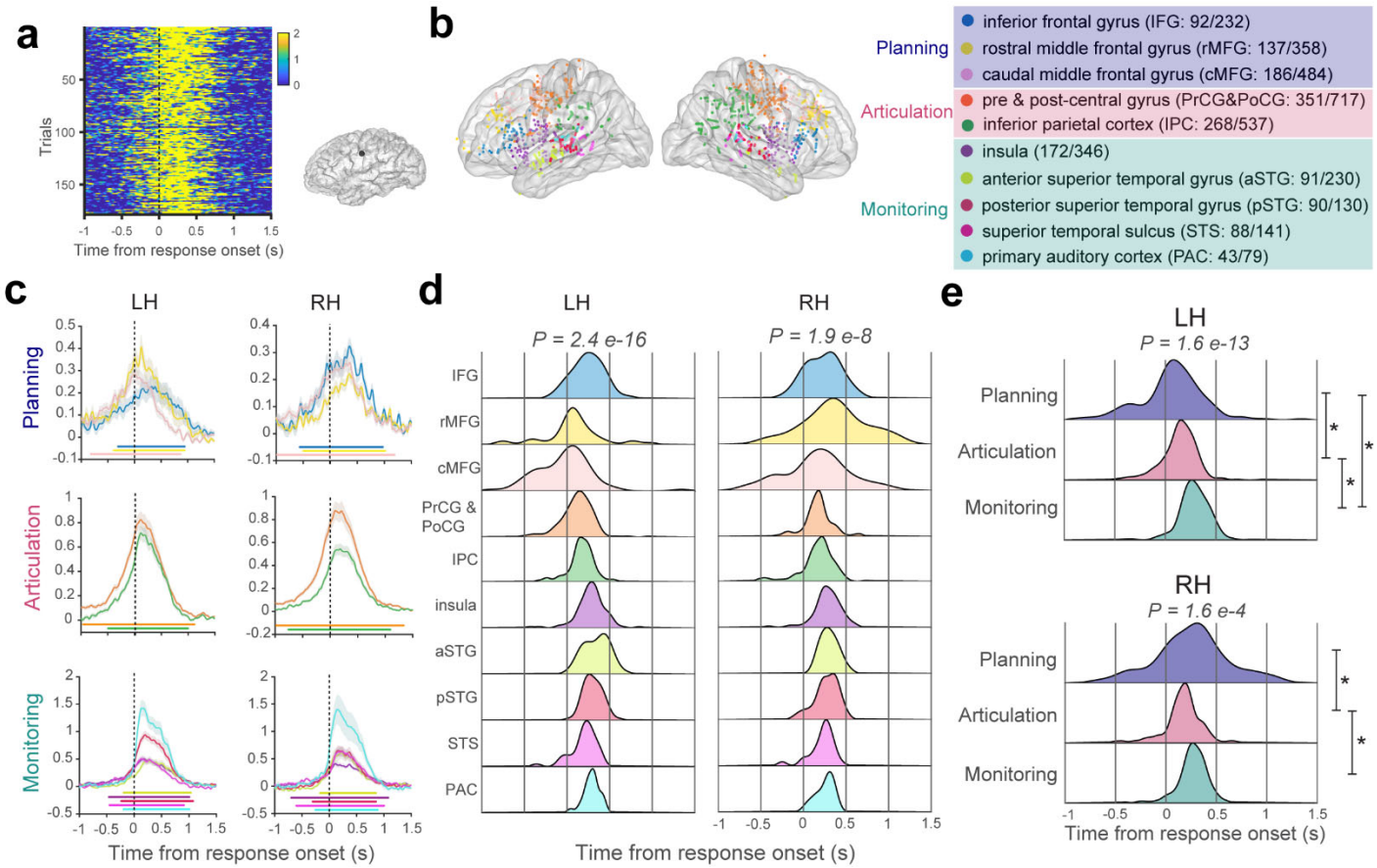

Figure 2. **Cortical activations from intracranial recordings during speech production. a.** Speech neural activations in the high-gamma band (HG: 70 – 150 Hz; z-score normalized with respect to baseline) of an example electrode in the pre-central gyrus that showed significant modulation during the speech utterance at the single trial level (FDR-corrected non-parametric permutation test). **b.** Location of all 3534 intracranial electrodes in both left (1701) and right (1605) hemisphere, with significant HG modulation during the speech utterance. The electrodes are color-coded with respect to anatomical regions-of-interest (ROIs), and are assigned to planning, articulation, and monitoring networks based on previous literature. The numbers in the brackets indicate the number of significant electrodes with respect to the total number of recorded electrodes per ROI. **c.** Mean HG activations of ROIs during response onset for both left and right hemispheres, averaged within patient. Shaded area denotes standard error. Colored horizontal bars indicate time-periods of significant activation above baseline (one-sided permutation test with cluster correction, $p<0.05$). **d.** Electrode distributions of peak HG time-points per ROI for each hemisphere. Significant differences in peak activations were observed in both left and right hemispheres, and there was a general trend of temporal progression of neural activations across ROIs around the speech utterance (one-way ANOVAs: left: $F_{9,717} = 10.87$, $p = 2.4e\text{-}16$, right: $F_{9,781} = 6.69$, $p = 1.9e\text{-}08$). **e.** Electrode distributions of peak HG time-points per ROI groups for each hemisphere. This temporal progression was also observed via significant differences in peak activations in both left and right hemispheres (one-way ANOVAs: left: $F_{2,724} = 30.7$, $p = 1.6e\text{-}13$, right: $F_{2,788} = 8.86$, $p = 1.6e\text{-}4$). Post-hoc t-tests indicate pairwise significance between the speech network: left hemisphere: $T_{Planning} < T_{Articulation}$, $t_{(365)} = 2.938$, $p = 0.0018$; $T_{Articulation} < T_{Monitoring}$, $t_{(507)} = 9.25$, $p = 3e\text{-}13$, $T_{Planning} < T_{Monitoring}$, $t_{(394)} = 9.84$, $p = 7.2e\text{-}21$, right hemisphere: $T_{Planning} > T_{Articulation}$, $t_{(511)} = 2.3$, $p = 0.0214$; $T_{Articulation} < T_{Monitoring}$, $t_{(528)} = 7.53$, $p = 2.2e\text{-}13$, $T_{Planning} < T_{Monitoring}$, $t_{(385)} = 2.36$, $p = 0.0188$. These analyses demonstrate the temporal progression of neural responses for planning, articulation, and monitoring for speech production in the left hemispheric speech network.

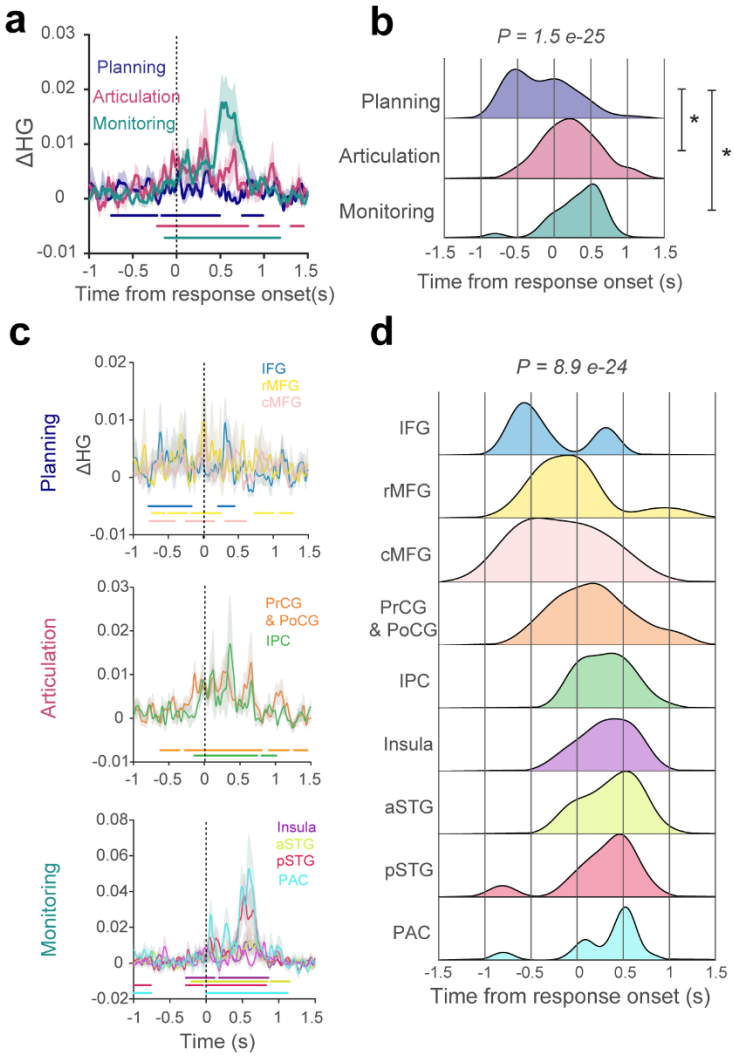

Figure 3. **Neural code for syllables in the left hemisphere.** Syllables are theorized to be composed during planning and used as frames for articulation and monitoring networks during speech execution**.** We analyze the neural code for the syllable frame through the use of a syllable discrimination framework: ROI-grouped HG activations of VCV trials were contrasted (subtracted) with CVC trials. **a.** This analysis was carried out for each speech network and time points with significant syllable discrimination was found for planning, articulation and monitoring networks. Plots show mean ΔHG contrasts of VCV versus CVC for each network aligned to the response onset. Shaded area denotes standard error. Horizontal bars indicate durations of significant contrasts (across vs. within frame, one-sided permutation test with cluster correction, $p<0.05$). **b.** Histograms show electrode distributions of peak ΔHG contrasts per network: planning, articulation, and monitoring. Significant differences in peak ΔHG (one-way ANOVA, $F_{2,495} = 64.3$, $p = 1.5e\text{-}25$), followed by post-hoc t-tests indicate pairwise significance in syllabic frame contrasts (ΔHG) between the networks indicate a temporal progression of syllable activations across ROI networks ($T_{Planning} < T_{Articulation}$, $t_{(333)} = 3.3$, $p = 0.0011$; $T_{Articulation} < T_{Monitoring}$, $t_{(403)} = 1.3$, $p = 0.2$ (not significant), $T_{Planning} < T_{Monitoring}$, $t_{(352)} = 3.78$, $p = 1.7e\text{-}4$). **c.** Contrasting syllables per ROIs within each speech network indicate time points with significant syllable discrimination around the speech utterance. Plots show mean ΔHG contrasts of VCV versus CVC for each ROI aligned to the response onset. Shaded area denotes standard error. Horizontal bars indicate durations of significant contrasts (across vs. within syllables, one-sided permutation test with cluster correction, $p<0.05$). **d.** Histograms indicate electrode distributions of peak ΔHG contrasts per ROI. Significant differences in peak ΔHG were found (one-way ANOVA, $F_{8,489} = 18.1$, $p = 8.9e\text{-}24$). These differences followed a general temporal progression of syllable activations, around the utterance These analyses show that the neural coding of the syllabic information occurs in a temporal progression from planning to execution in left-hemispheric speech ROIs**.** This progression supports a model for planning and execution where syllable frames are first enacted in the planning stage, before being executed via articulatory and feedback speech networks.

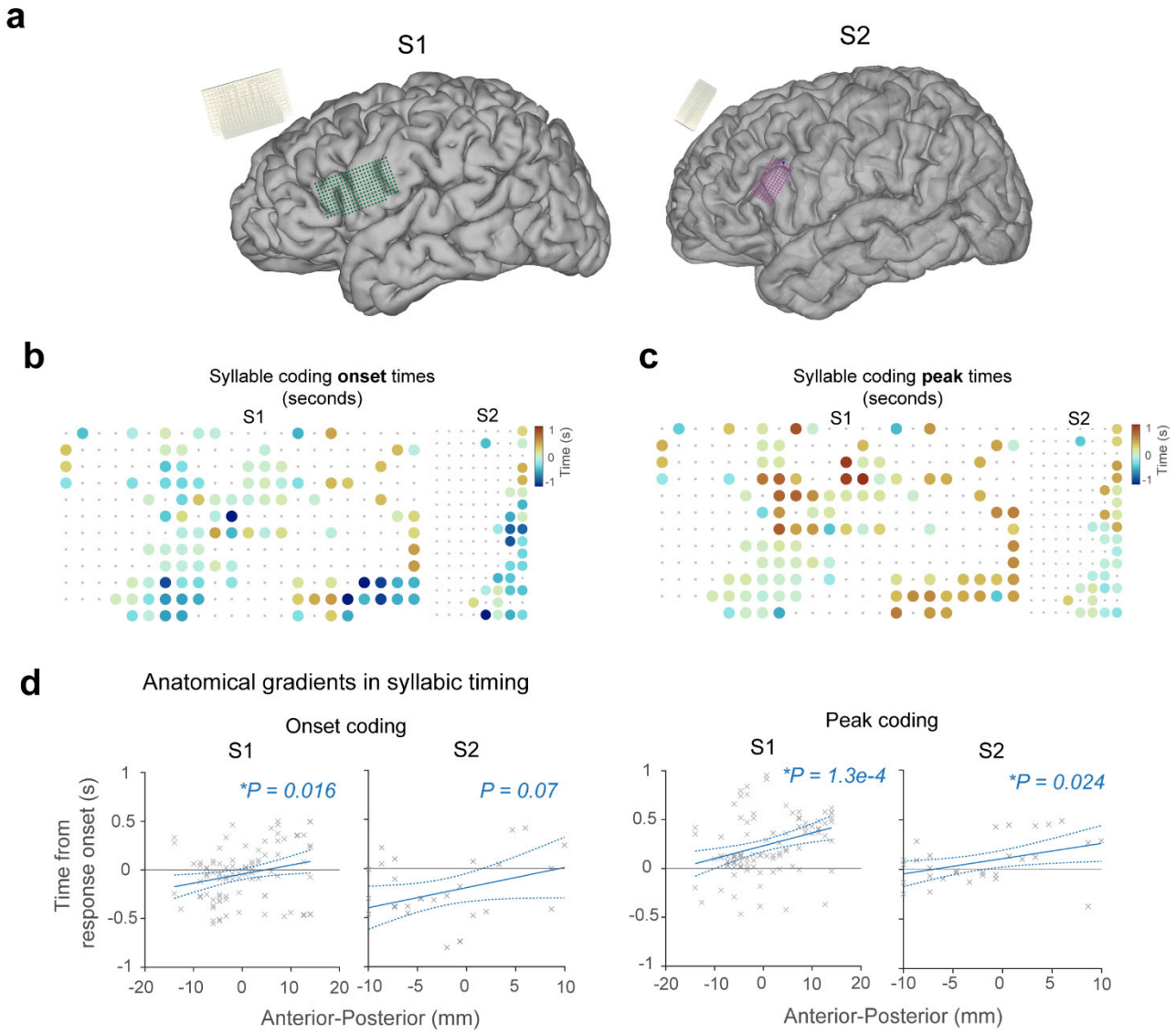

Figure 4. **μECoG captures spatio-temporally selective syllabic coding sites for speech planning and execution. a.** High-density μECoG electrode arrays were implanted over planning (anterior) and execution (posterior) networks in two awake patients. Electrode arrays had either 256 electrodes with 1.72 mm inter-electrode distance for S1 (green) or 128 electrodes with 1.33 mm inter-electrode distance for S2 (pink). **b.** Peak HG activations show fine-scale spatio-temporal patterns during response onset. Large circles indicate electrodes with significant HG activations with heatmap colors corresponding to signal power (FDR-corrected permutation test, $p < 0.05$, S1: 201/256 significant/total electrodes, S2: 63/128). Non-significant electrodes are denoted by grey dots. Arrows indicate electrode array orientation from anterior to posterior. **b, c.** Neural code for the syllable was extracted by contrasting HG activations per syllabic structure (ΔHG: VCV – CVC) for each electrode. Heat maps show spatial activation patterns of onset and peak times of ΔHG activations for electrodes with significant syllable contrasts. Large circles reflect electrodes with significant syllable contrasts (one-side permutation test with cluster correction, $p < 0.05$, S1: 100/201, S2: 37/63). Non-significant electrodes are indicated by grey dots. These significant contrasts indicate syllable coding on an electrode specific level and demonstrate spatially selectivity on the underlying neuroanatomy. **d.** A spatial correlation analysis of the syllable contrast onsets along the anterior-posterior axis shows a spatial progression of temporal coding for syllable codes starting in the anterior planning section and moving towards the posterior execution network (S1: $R^2 = 0.05$, $F_{2,92} = 6.02$, $p = 0.016$, S2: $R^2 = 0.08$, $F_{2,28} = 6.02$, $p = 0.07$). This spatio-temporal progression was also seen in the peak times (F-statistic test against the constant model, S1: $R^2 = 0.09$, $F_{2,98} = 10.9$, $p = 1.3e\text{-}4$, S2: $R^2 = 0.13$, $F_{2,28} = 5.68$, $p = 0.024$). Dotted lines indicate the 95% confidence interval of the linear model fit. The presence of a spatial gradience, extracted using high-density μECoG electrode arrays, demonstrate a potential anatomical substrate for the transition from planning to execution for speech production.

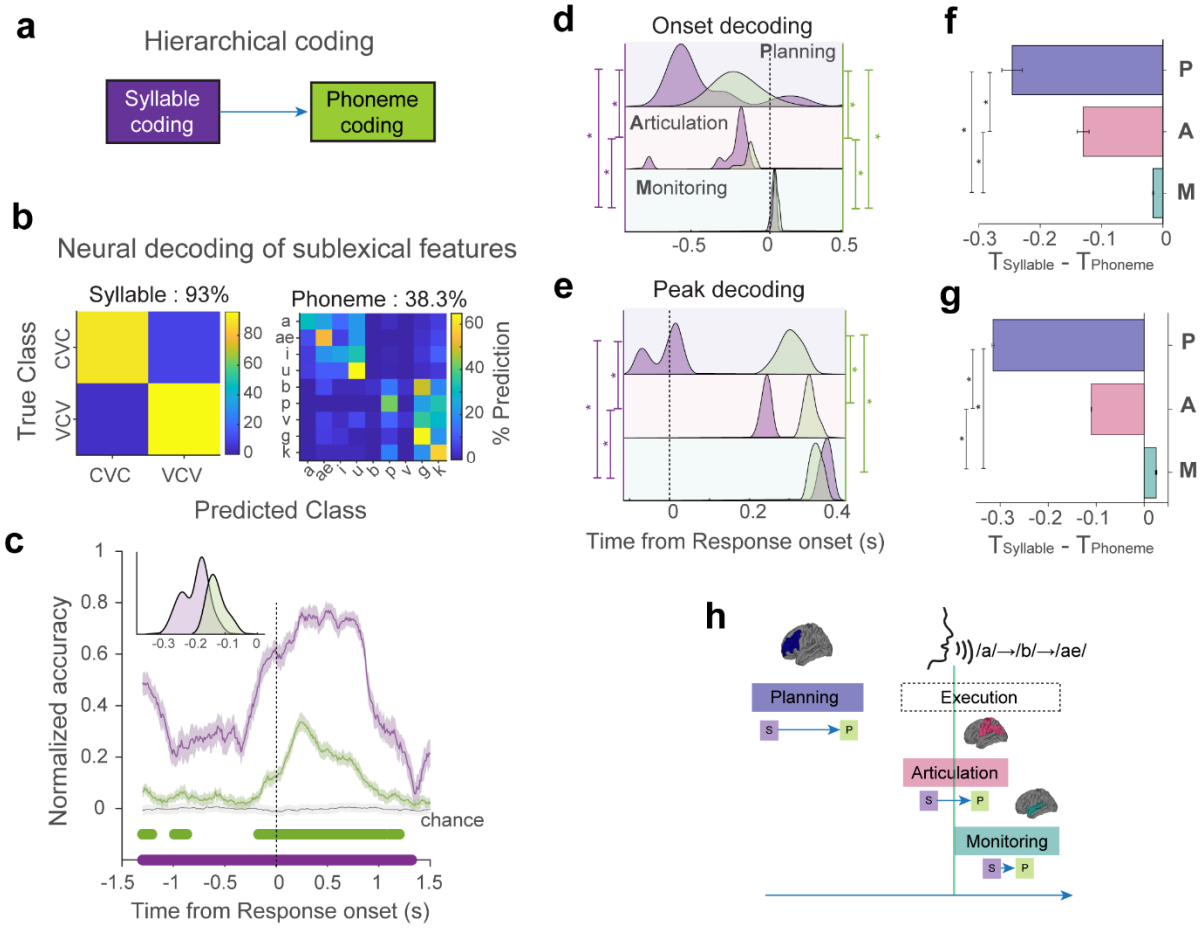

Figure 5. **Left-hemispheric ROIs demonstrate two-tiered hierarchical organization of syllables and phonemes. a.** Graphical representation of hierarchical speech representations with syllables coded at higher levels and phonemes are coded at lower levels. The hierarchical organization allows syllable frames to be coded earlier in time, followed by phonemes into the syllable frame. **b.** Neural codes for syllables and phonemes were characterized by neural decoding analyses. Heatmaps indicate confusion matrices (SVD-LDA, 10-fold cross-validation) for syllables (2-way; chance 50%) and phonemes (9-way, chance: 11.11%) in speech significant electrodes (2895 total electrodes, 69 mean electrodes per subject). **c.** The temporal evolution of syllables and phoneme activations were obtained using time-resolved neural decoding (200 ms window; 10 ms step size). Solid lines indicate average across 20 instances of decoding iterations, and shaded bars indicate standard errors across instances. Grey traces indicate mean and standard errors of chance decoding which was obtained from randomly shuffling syllable and phoneme labels (250 shuffled decoding models). Horizontal bars indicate durations of significant decoding (temporal permutation cluster test, $p < 0.05$; one-sided permutation test against 250 shuffled decoding models). Colored histograms indicate distributions of decoding onsets of syllables (violet) and phonemes (green), with syllables detectable earlier than phonemes (one-sided Mann-Whitney U test, $p = 3.14e\text{-}8$). **d, e.** Colored histograms indicate distributions of onset decoding times for both syllable and phoneme for each ROI network. Significant differences in decoding distributions indicate a temporal progression across the ROI networks for both syllables and phonemes (one-way ANOVA with post-hoc Tukey HSD test, $p < 0.01$, for exact p-values see Main text). Syllable coding was earlier in time than phoneme coding ($p < 0.05$, one-sided Mann-Whitney U test; except peak decoding at monitoring ROIs, $p = 0.0785$), demonstrating a network-level organization in accordance with the proposed hierarchical representation of syllables and phonemes (see **Fig. 1a**). **f, g.** The difference in activation timing of the hierarchical representation varied for each ROI group, with maximum difference between syllable and phoneme activation during planning (250-350 ms), followed by articulation (110 – 130 ms), and then minimally (17 – 25 ms) during monitoring (pairwise t-test, planning vs. articulation, onset: t(798) = -10.7, p = 5.5e-25, peak: t(798) = 71, p = 0; articulation vs. monitoring: onset: t(798) = 18.2, p = 1.9e-62, peak: t(798) = 101, p = 0; planning vs. monitoring: onset: t(798) = -8.04, p = 3.3e-15, peak: t(798) = 117.8, p = 0). **h.** The temporal difference in hierarchical organization between syllables and phonemes suggests the existence of a temporal compression between syllables and phonemes from planning to execution, allowing for both distinct syllable and phoneme plans prior to execution, and their integration during articulation.

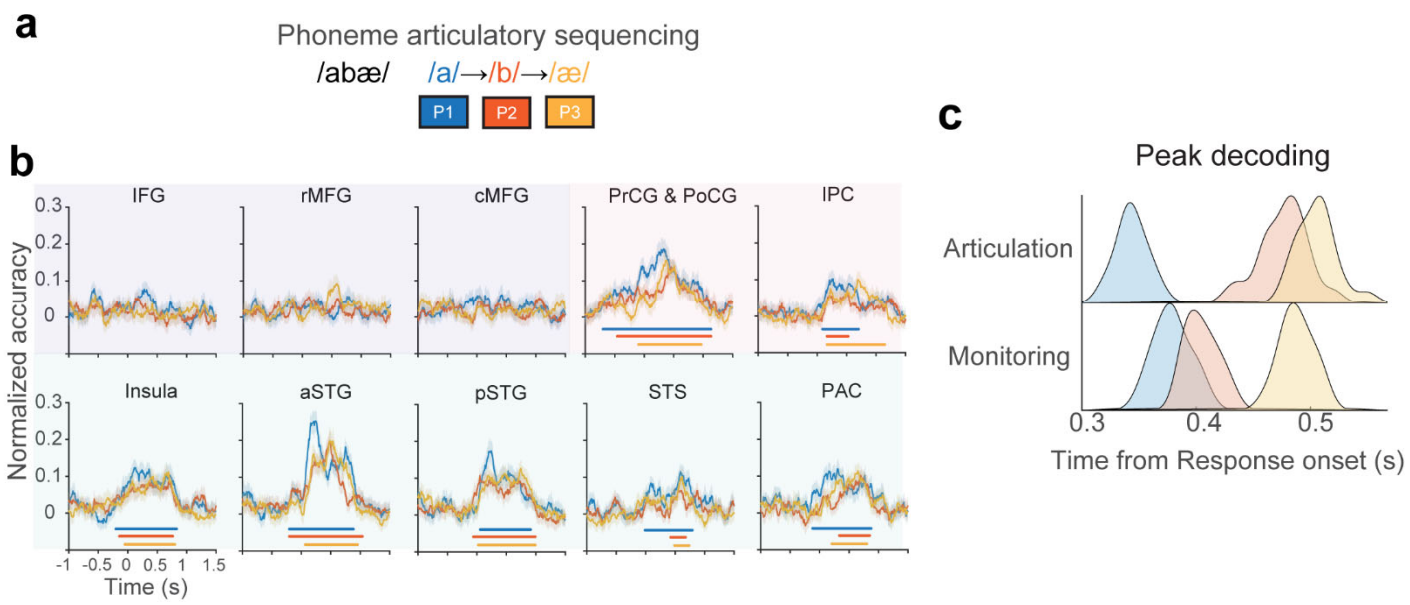

Figure 6. **Phonological units are sequentially articulated during execution. a.** Graphical representation of sequential execution that depicts phonological units articulated in their correct order during the utterance. **b.** Time-resolved decoding models were trained on neural signals aligned to the utterance to decode phonological units at each position within the pseudo-word. Each colored trace indicates temporal evolution of positional phonological units in time for an ROI (mean across 20 decoding instances). Results are shown for all ROIs in the left hemisphere, grouped by planning, articulation, and monitoring networks. Significant phonological unit decoding for all positions was observed only in execution ROIs (articulation and monitoring), suggesting that temporal sequencing is carried out as part of the execution of speech sounds rather than during planning. Horizontal bars indicate durations of significant decoding (temporal permutation cluster test, $p < 0.05$; one-sided permutation test against 250 shuffled decoders). Background shading indicates ROI networks for planning, articulation and monitoring. **c.** Colored histograms show distribution times of decoding peaks for each positional phonological unit within the ROI network. Positional phonological unit codes were decoded in the correct serial order (pairwise one-sided Mann-Whitney U test, $p < 0.05$). The presence of ordering demonstrates the serial execution of positional phonological unit codes by the execution network.

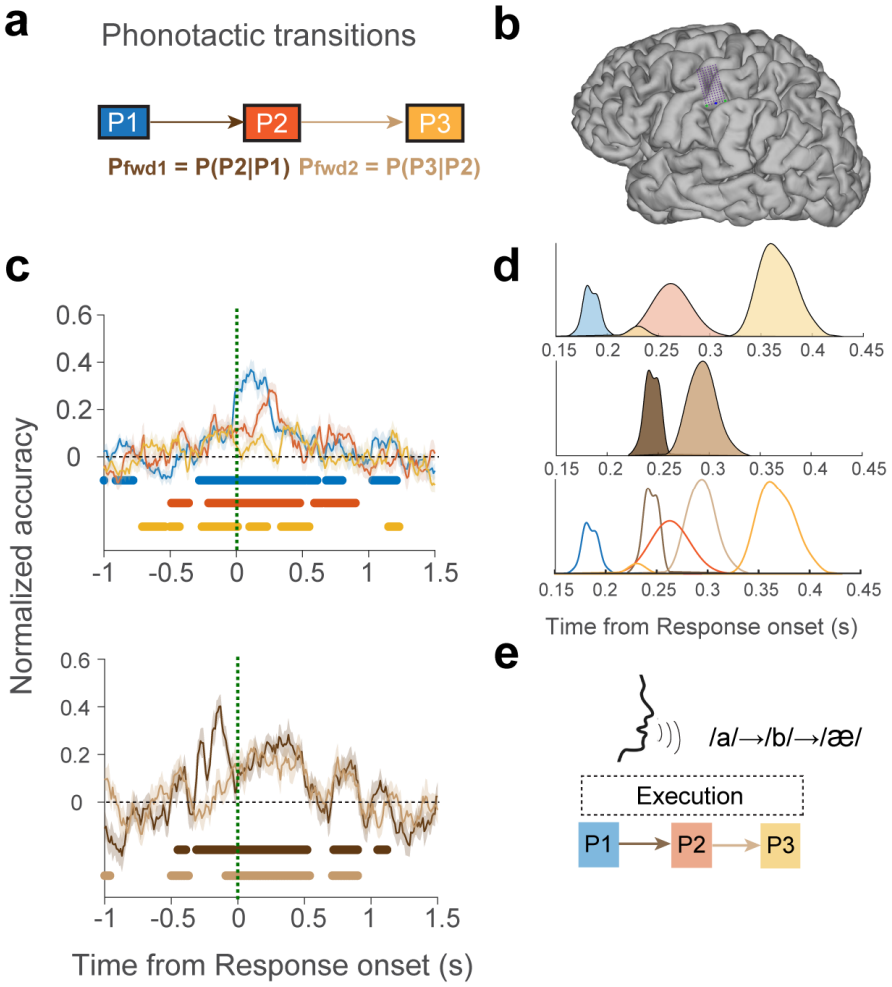

Figure 7. **Continuous phonotactic transitions are present in motor cortex during articulation. a.** Graphical schematic demonstrating that phonological unit sequence execution may involve both discrete phonological units as well as their underlying transitionary statistics (phonotactics). Forward phonotactic transition from C1 to V2 is denoted by P_{fwd1}, and from V2 to C3 is denoted by P_{fwd2}. **b.** μECoG electrode array was implanted over sensorimotor cortex in an awake patient. The electrode array had 128 electrodes with 1.33 mm inter-electrode distance (purple). **c.** Time-resoled decoding revealed accurate phonological unit sequencing for CVC trials (C1: blue, V2: orange, C3: yellow). Temporal decoding models trained to predict the phonotactic transitions (3-way: low, medium, and high phonotactic transitions) demonstrated significant decoding for P_{fwd1} (dark brown) and P_{fwd2} (light brown). Horizontal bars indicate durations of significant decoding (temporal permutation cluster test, $p < 0.05$; one-sided permutation test against 250 shuffled decoders). **d.** Colored histograms indicate distribution times of peak decoding onsets for positional phonological units and phonotactic transitions, revealing significant temporal ordering of phonological units and phonotactic transitions (pairwise one-sided Mann-Whitney U test, $p < 0.05$, $T_{C1} < T_{V2}$, p = 1.6e-08; $T_{V2} < T_{C3}$, p = 4.2e-07, $T_{Pfwd1} < T_{Pfwd2}$, p = 6.6e-08). Further, the timing of the forward transitions were embedded between the phonological unit positions (pairwise one-sided Mann-Whitney U test, $p < 0.05$, $T_{C1} < T_{Pfwd1}$, p = 3.7e-08; $T_{Pfwd1} < T_{V2}$, p = 6.6e-08, $T_{V2} < T_{Pfwd2}$, p = 6.2e-08, $T_{Pfwd2} < T_{C3}$, p = 7e-07). **e.** This significant temporal embedding demonstrates the existence of both discrete and continuous motor coding during the execution of sequences during speech production.