A speech planning network for interactive language use

https://doi.org/10.1038/s41586-021-04270-z

Received: 29 September 2020

Accepted: 19 November 2021

Published online: 05 January 2022

Check for updates

Gregg A. Castellucci[1,2], Christopher K. Kovach[3], Matthew A. Howard III[3], Jeremy D. W. Greenlee[3] & Michael A. Long[1,2]✉

During conversation, people take turns speaking by rapidly responding to their partners while simultaneously avoiding interruption[1,2]. Such interactions display a remarkable degree of coordination, as gaps between turns are typically about 200 milliseconds[3]—approximately the duration of an eyeblink[4]. These latencies are considerably shorter than those observed in simple word-production tasks, which indicates that speakers often plan their responses while listening to their partners[2]. Although a distributed network of brain regions has been implicated in speech planning[5–9], the neural dynamics underlying the specific preparatory processes that enable rapid turn-taking are poorly understood. Here we use intracranial electrocorticography to precisely measure neural activity as participants perform interactive tasks, and we observe a functionally and anatomically distinct class of planning-related cortical dynamics. We localize these responses to a frontotemporal circuit centred on the language-critical caudal inferior frontal cortex[10] (Broca's region) and the caudal middle frontal gyrus—a region not normally implicated in speech planning[11–13]. Using a series of motor tasks, we then show that this planning network is more active when preparing speech as opposed to non-linguistic actions. Finally, we delineate planning-related circuitry during natural conversation that is nearly identical to the network mapped with our interactive tasks, and we find this circuit to be most active before participant speech during unconstrained turn-taking. Therefore, we have identified a speech planning network that is central to natural language generation during social interaction.

Conversational turn-taking can be divided into three major cognitive processes: perception of the opposing speaker's turn, planning of one's own turn and production of the speech comprising that turn[2,14] (Fig. 1a, b). Although each of these steps is composed of several subcomponents[2,8,14] (Extended Data Fig. 1a), speech planning is an especially multifaceted process encompassing various functions ranging from abstract conceptual and semantic operations to low-level articulatory programming and motor initiation[15]. Accordingly, many cortical regions have been linked to aspects of planning, including the inferior frontal gyrus[5,9,16,17], the premotor cortex[8], the superior temporal gyrus[6], the supplementary motor area[8,18] and the inferior parietal cortex[7]. However, because non-interactive tasks—such as picture naming[16,17], repetition[5,6,18] and reading[5,9]—have typically been used to identify such candidate regions, their relevance to speech preparation during interaction is unknown. In this study, we delineate the neural substrates underlying the planning processes relevant for rapid turn-taking by measuring cortical activity while participants engage in structured interactive tasks as well as unconstrained conversation.

Characterization of planning activity

Speech planning is an internal process with little or no behavioural correlate[19] that often overlaps with speech perception and production during natural conversation[2,20–22]. Therefore, we first sought to experimentally isolate neural activity related to planning during turn-taking. To do so, we employed an established question–answer paradigm in which a single word (that is, the 'critical information' or CI) initializes speech planning by providing the information necessary for a correct response[20] (Fig. 1c–f). An experimenter posed a battery of 39 to 94 CI questions (55.1 ± 20.5, mean ± s.d.; Supplementary Data 1) to 8 participants (neurosurgical patient volunteers; Extended Data Table 1), with the CI presented either near the middle ('early') or the end ('late') of the question. If speech planning is initiated by CI, late trials should show longer response latencies than early trials as they provide relatively less planning time[2,20]. As expected, we observed that the median reaction times were significantly longer in late trials compared with early trials (median: 782 ms versus 495 ms; $P < 0.05$ ($n = 8$ participants), signed-rank test; Extended Data Fig. 1b, c), indicating that the CI paradigm temporally isolated the prearticulatory processes relevant to rapid turn-taking in our neurosurgical cohort.

We next examined cortical responses while participants answered CI questions. Previous work using this task has reported widespread CI-related activity at the scalp using electroencephalography[20], and we sought to leverage the temporal (<10 ms) and spatial (<5 mm) precision[23] of intracranial electrocorticography (ECoG) to measure

[1]NYU Neuroscience Institute and Department of Otolaryngology, New York University Langone Medical Center, New York, NY, USA. [2]Center for Neural Science, New York University, New York, NY, USA. [3]Department of Neurosurgery, University of Iowa, Iowa City, IA, USA. ✉e-mail: mlong@med.nyu.edu

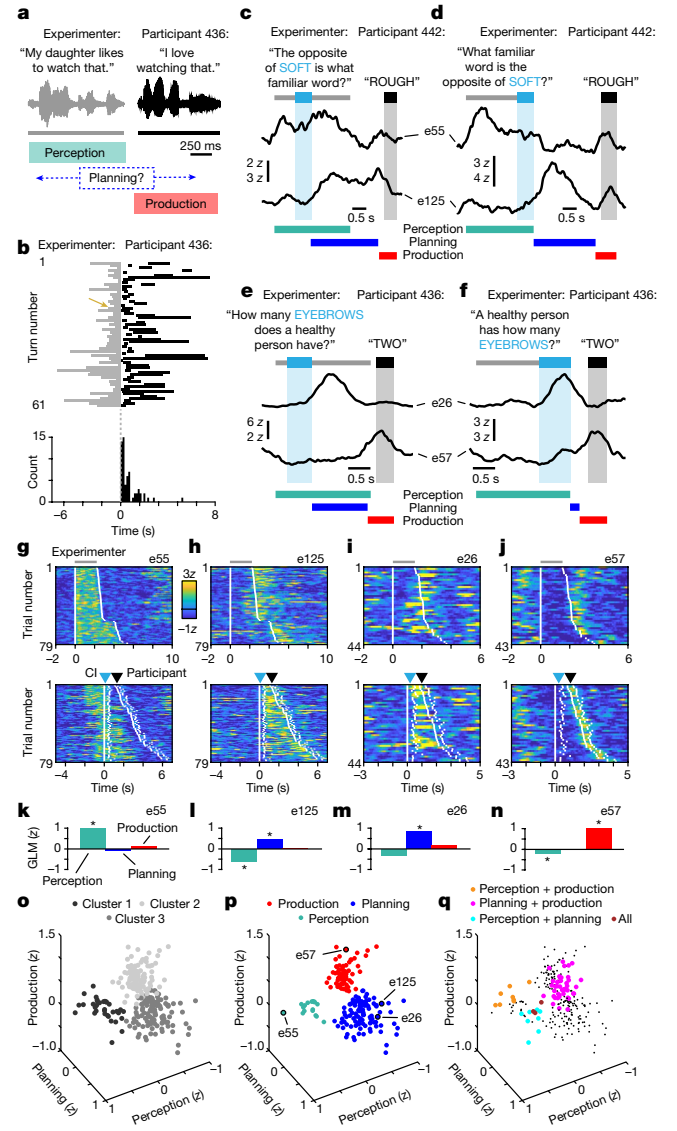


Fig. 1 | Cortical responses during interactive speech. a, Example interaction between the experimenter and participant 436 during conversation (top) with a diagram depicting the three behavioural phases of turn-taking (bottom). **b**, A schematic of all turn-taking interactions with participant 436 (top) and the associated distribution of interturn gaps (bottom). The gold arrow indicates an example interaction from **a**. **c**–**f**, Example early (**c**, **e**) and late (**d**, **f**) CI trials with high gamma activity from participants 442 (**c**, **d**) and 436 (**e**, **f**). The timing of the experimenter question, the CI and the participant answer are indicated with grey, light blue and black bars, respectively (coloured bars denote GLM regressor timing). **g**–**j**, Activity of the four example electrodes in **c**–**f** across all trials aligned to question (top) or CI (bottom): e55 (**g**), e125 (**h**), e26 (**i**) and e57 (**j**). The white ticks denote the onset and offset of the task epochs indicated with coloured arrows in **c**–**j**. In **c**–**j**, the waveforms are smoothed for display (500-ms mean boxcar filter). **k**–**n**, Perception, planning and production GLM weights for the example electrodes in **g**–**j**, respectively. Significant weights are indicated with asterisks. **o**, The distribution of all task-responsive electrodes in GLM weight space. The cluster membership is indicated by the greyscale colour. **p**, **q**, The distribution of electrodes displaying responses in one window of the CI task (**p**) or multiple windows (**q**). The response class is indicated by colour. In **q**, unmixed electrodes from **p** are denoted by small black points.

planning activity at distinct cortical sites. We implanted a total of 874 ECoG electrodes across the left, language-dominant hemispheres of 8 participants (64 to 224 electrodes in each; Supplementary Data 2) and found 790 electrodes (90.4%) to be suitable for further analysis (Methods). We assayed neural activity by examining the high gamma frequency band (approximately 70–150 Hz) of the local field potential and observed that the activity profiles of many electrodes were temporally locked to specific phases of the CI questions (Fig. 1c–f). Specifically, perception-related activity remained sustained throughout the duration of the experimenter's question (Fig. 1c, d), and production-related activity was largely restricted to the period comprising the participant's spoken answers (Fig. 1e, f). By contrast, planning-related activity was observed immediately following CI presentation and generally returned to baseline before participant response (Fig. 1c–f); this profile was consistent across trials (Fig. 1g–j) regardless of the CI position or question content (Extended Data Fig. 1d, e).

Cortical language circuitry is highly multimodal[24,25] and previous research has suggested that processing for interactive behaviours may be widely distributed rather than organized into discrete modules[26]. Therefore, it is possible that the dynamics exhibited by our electrodes do not cluster into distinct categories related to specific phases of spoken interactions but instead form a continuum at the population level. To differentiate these alternatives, we designed a general linear model (GLM) to quantify activity levels during the perception, planning and production windows of the CI task (coloured bars in Fig. 1c–f, Extended Data Figs. 1f, 2). We focused our analyses on significant increases in high gamma amplitude—a correlate of local neuronal activity[27]—and detected a total of 253 electrodes (32.0%) with significantly elevated responses during at least one of the defined task epochs (Fig. 1k–n, Supplementary Data 2). We then examined the organization of all task-responsive electrodes in three-dimensional space according to their perception, planning and production GLM weights (Fig. 1o) and found that significant positive responses for each category were invariably confined to separate clusters (*k*-medoids clustering; Fig. 1p). In contrast, a 'continuum model' assuming unimodal distributions of GLM weights (Extended Data Fig. 3a–c) always exhibited clusters containing multiple response classes (Extended Data Fig. 3d). In addition, we found that only 61 electrodes (24.1% of responsive sites) in our recorded data showed positive responses within more than one task window (Fig. 1q), significantly fewer than expected under the continuum model (interquartile range 44.3–48.6%; Extended Data Fig. 3e). Taken together, these results demonstrate that the neural responses recorded with our ECoG electrodes are organized into discrete classes related to speech perception, planning and production during turn-taking.

We next examined the relative frequency of neural responses (that is, significant positive GLM weights) related to each window of the CI task. We found that responses related to planning and production were most common, with 20.3% and 15.3% of electrodes showing significant increases in activity during the planning and production windows, respectively. These electrodes often exhibited significant negative perception-related GLM weights (Fig. 1k–n, Extended Data Fig. 4a, Supplementary Data 2), which is likely to reflect a decrease in activity during the perception window of the CI task (Extended Data Fig. 4b). Meanwhile, only 4.6% of electrodes were responsive during the perception window, probably resulting from sparse electrode coverage over the auditory-related temporal cortex[28] (Extended Data Fig. 5a–c). To ensure that the temporal overlap of the perception and planning windows in early CI trials (Fig. 1c, e) did not bias our analysis against detecting perception responses, we reanalysed our dataset with a reduced GLM lacking a planning component and found no net increase in perception-related electrodes (Extended Data Fig. 3f–h). These results demonstrate that our statistical approach and behavioural paradigm effectively delineated the cortical dynamics related to each phase of these interactions.

Spatial structure of planning responses

Although ECoG has been previously used to map cortical regions related to speech perception[28] and production[29], a similar high-precision

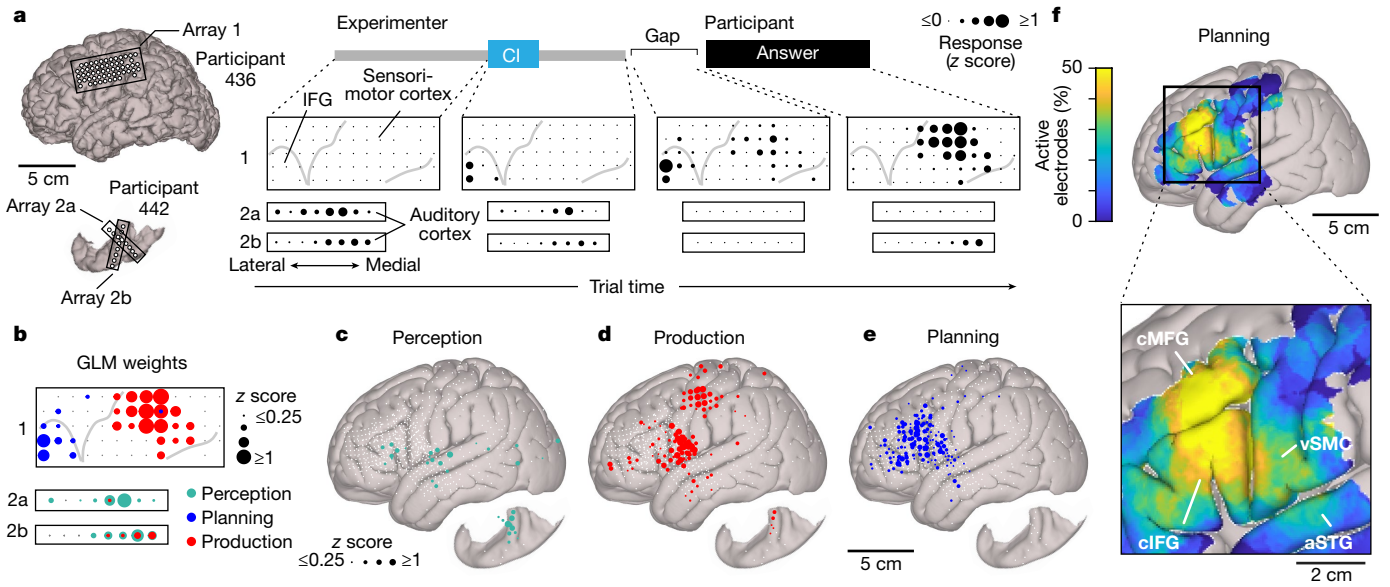


Fig. 2 | Spatial organization of speech planning responses. a, The average median high gamma activity across trials from electrodes on the left lateral cortical surface and superior temporal plane at different phases of the CI task. The electrode locations are shown on the left. Median values from time windows <50 ms were not included for analysis. **b**, GLM classification of the responses depicted in **a**. **c**–**e**, Canonical cortical surfaces displaying electrodes with responses related to speech perception (**c**), production (**d**) and planning (**e**) across all participants. Electrodes that did not display a significant response for a given process are indicated with small white circles. **f**, The proportion of electrodes displaying planning responses (1-cm-diameter spatial smoothing) (top) and regions displaying consistent planning responses in the CI task labelled on a magnified cortical surface (bottom).

spatial characterization for speech planning remains elusive. In addition, it is possible that planning-related loci are not spatially segregated[30] and share extensive territory with perception-related and/or production-related sites[24,25]. To resolve this issue, we first examined the activity recorded within individual participants and observed that electrodes responsive for the same windows of the CI task (that is, same GLM class) were spatially clustered (Fig. 2a, b, Extended Data Fig. 5a). Across participants, anatomically distinct cortical networks related to speech perception, planning and production during the CI task appeared at consistent cortical sites (Supplementary Data 2), regardless of their underlying clinical condition (Extended Data Fig. 5e). Specifically, electrodes that were responsive during perception (Fig. 2c) and production (Fig. 2d, Extended Data Fig. 5d) were largely restricted to well characterized sensory[28] and motor[29] structures, respectively. Meanwhile, 95% of planning-related electrodes were grouped in a spatially segregated frontotemporal region (Fig. 2e, Extended Data Table 2), with the highest densities of planning electrodes clustered in the caudal inferior frontal gyrus (cIFG) and the caudal middle frontal gyrus (cMFG) and a smaller number of planning sites located in the ventral speech motor cortex (vSMC) and the anterior superior temporal gyrus (aSTG) (Fig. 2f). We found that many electrodes within planning-related and production-related vSMC, IFG and cMFG also exhibited significant negative perception-related GLM weights (Extended Data Fig. 4c–e), which suggests that these structures are relatively less active during speech perception. However, a small number of electrodes within the IFG displayed perception-related responses (Fig. 2c), consistent with the previously established role for the IFG in language comprehension[31]. In summary, we have delineated a spatially coherent network whose activity is selectively linked to planning speech.

Speech selectivity of planning responses

To initiate speech planning in the CI task, a participant must first perceive the CI; therefore, this paradigm does not disentangle the cognitive processes involved in comprehending the CI from those occurring at planning onset. We addressed this issue using a command-response (CR) task in which participants were instructed to perform a range of actions varying in their linguistic relevance (Supplementary Data 1), including: hand movements (that is, button pressing; 'CR1', Fig. 3a), non-speech orofacial behaviours ('CR2', Fig. 3b), speech repetition ('CR3', Fig. 3c) and linguistic operations (that is, noun pluralization; 'CR4', Fig. 3d). Crucially, the structure of the CR tasks mirrors the CI task—a critical word or phrase is presented either early or late and its comprehension is required to generate a motor response.

Using this approach, we further characterized 124 electrodes from 6 participants who displayed significant planning-related responses in the CI task (Extended Data Table 1). Although individual planning electrodes could respond maximally in each of the CR tasks (Fig. 3a–d), only a subset displayed significant increases in preparatory activity for hand (18.6%) and non-speech orofacial movements (43.6%) whereas the majority were responsive for speech repetition (59.7%) and pluralization (80.7%) (Fig. 3e–h). At the population level, planning activity was greater in the CI task than in all CR conditions except pluralization ($P < 0.005$ and $P > 0.99$, respectively ($n = 124$ electrodes), Friedman test with Dunn–Šidák post-hoc tests; Fig. 3i). However, reaction times did not differ across the repetition task, pluralization task and CI questions ($P = 0.3114$ ($n = 6$ participants), Friedman test; Fig. 3j), indicating that the observed activity differences are unlikely to have resulted from task difficulty. Taken together, these results indicate that activity within the identified planning network reflects speech preparation rather than processes related to comprehension or task engagement.

Although planning electrodes were generally more active during speech as opposed to non-speech CR tasks (Fig. 3e–i), we hypothesized that individual regions may display variable degrees of speech selectivity[6,32]. Therefore, we examined planning activity levels in relation to cortical location and observed higher levels of preparatory activity broadly within the planning network as linguistic relevance increased (Fig. 3k–n). However, we found that planning electrodes within the cIFG were more linguistically selective than electrodes located in the precentral gyrus (PreCG) and the cMFG (Fig. 3o–r), as most PreCG and cMFG electrodes displayed significant planning responses in both speech tasks whereas electrodes in the cIFG were generally responsive for

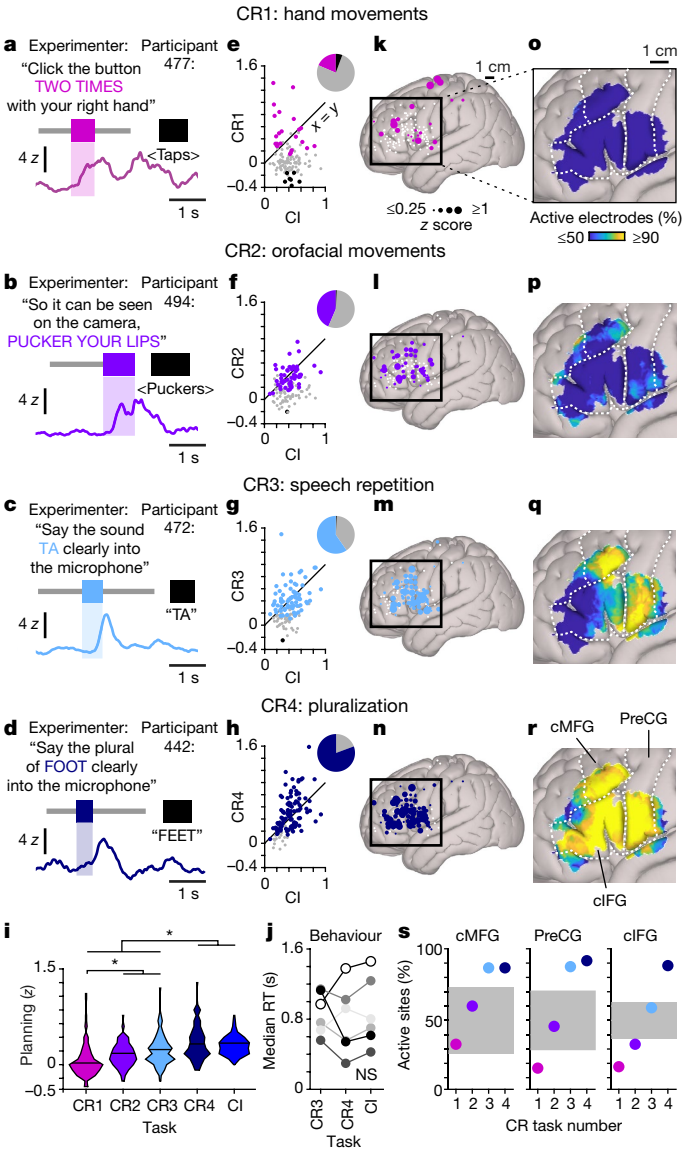


Fig. 3 | Linguistic selectivity of planning responses. a–d, High gamma activity in single trials of each CR task: non-speech hand (CR1; **a**) and orofacial (CR2; **b**) movements, speech repetition (CR3; **c**) and pluralization (CR4; **d**). The timing of experimenter question, CI and participant response are indicated with grey, coloured and black bars, respectively. The waveforms are smoothed for display (250-ms mean boxcar filter). **e–h**, The planning weights for each CR task plotted against planning weights from the CI task. The inset pie charts indicate the proportion of planning electrodes that were significantly responsive (coloured), not active (grey) or suppressed (black) for each CR task. **i**, The distribution of planning weights for all planning-related electrodes in the CR and CI tasks. The asterisks indicate significant differences between tasks ($P < 0.005$, Friedman test with Dunn–Šidák post hoc tests). **j**, Median reaction times (RT) in late trials of the CR3, CR4 and CI tasks did not differ significantly ($P = 0.34$, Friedman test; NS, not significant). **k–n**, Canonical cortical surfaces indicating electrodes with planning responses for each CR task. Electrodes that did not produce a significant response for a given task are denoted by small white circles. **o–r**, The proportion of planning electrodes showing significantly elevated responses in each CR task (1-cm-diameter spatial smoothing). The dashed lines indicate approximate boundaries of the cMFG, the PreCG and the cIFG. **s**, The percentage of electrodes located within each structure that displayed planning responses for each CR task. The distribution of percentages generated with shuffled data is indicated in grey.

the pluralization task only (Fig. 3q–s). These results therefore suggest that our planning circuit is further divided into subnetworks related to distinct speech planning processes.

Planning during natural conversation

We next aimed to compare the neural activity observed during task-based interaction to that arising during natural language use. To do so, we recorded neural activity while participants engaged in unconstrained conversation with the experimenter for 2.8 min to 16.3 min (Extended Data Fig. 6a) following completion of the structured tasks. In this context, our cohort exhibited grossly normal turn-taking behaviour, the temporal properties of which were comparable to that observed in the general population[2,3] (Extended Data Fig. 6b). These rapid interactions therefore provide an opportunity to investigate the neural mechanisms underlying naturalistic turn-taking.

During conversation, we observed that individual electrodes displaying planning-related responses in the CI task were often active before the onset of a participant's turn (Fig. 4a), which suggests that these sites are engaged in speech preparation during both behaviours. However, unlike our interactive tasks, speech planning during natural behaviour is highly flexible[22,33] and not temporally locked to an experimentally defined stimulus (that is, the CI); this behavioural variability thus renders a supervised analysis of neural activity largely unfeasible. To overcome this issue, we performed principal component analysis (PCA) on continuous electrode signals recorded while participants engaged in either the interactive tasks or natural conversation. This approach allowed us to identify sets of electrodes displaying correlated signals during task performance and conversation independently and assess whether network structure remains stable between the two contexts. During the tasks, we found that the electrode signals formed three main clusters in three-dimensional principal component (PC) coefficient space ($n = 6$ participants; Fig. 4b, Extended Data Figs. 6c, 7a–f), indicating that neural activity during task-based interactions is organized into distinct classes. We then asked whether this configuration is preserved in unstructured turn-taking and found that electrode signals formed nearly identical clusters during conversation (Fig. 4c, Extended Data Fig. 7a–f). Specifically, only 5 of 200 electrodes were differently clustered between the two behaviours across participants, which is significantly less than would be expected by chance (interquartile range 120–128; $P < 0.0001$, k-medoids clustering with permutation test; Extended Data Fig. 6d). Thus, the cortical network active during language generation in task-based interactions is similarly organized during natural turn-taking.

We subsequently assessed whether the PCA activity clusters observed during natural conversation corresponded to neural responses related to speech perception, planning and production. Although these functional classifications were originally made using a GLM—which detected elevated neural activity within predefined time windows of the CI task (Fig. 1)—we found that clusters in PC coefficient space were overwhelmingly composed of electrodes exhibiting a single class of task-related response, thus enabling us to functionally categorize each PCA cluster as 'perception', 'planning' or 'production' in 17 out of 18 cases across participants (Fig. 4c, Extended Data Fig. 7a–f). We observed a high degree of correspondence between cluster categorization and GLM-derived activity classes, such that 93.8%, 94.7% and 96.1% of electrodes in perception, planning and production clusters displayed analogous responses ($P < 0.0001$, permutation test; Extended Data Fig. 6e). Therefore, using our unsupervised PCA approach, we uncovered categories of neural activity during natural conversation that correspond to the distinct response classes and circuits isolated with the CI task (Extended Data Fig. 6f).

We next assayed time-varying PC activity (that is, PC scores) to examine the functional properties of electrodes within perception, planning and production clusters during task-based and natural turn-taking.

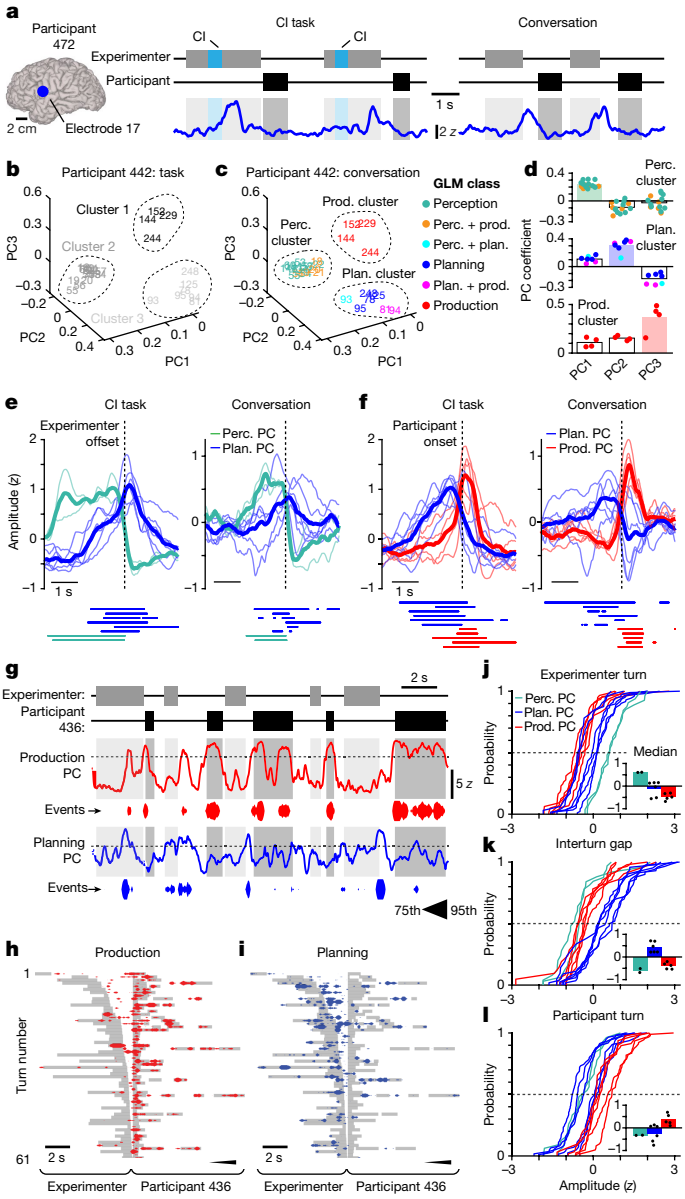


Fig. 4 | Planning activity during natural conversation. a, High gamma activity from an example planning electrode in participant 472 (left) during the CI task (middle) and unconstrained conversation (right). The waveforms are smoothed for display (250-ms mean boxcar filter). **b**, **c**, For all CI task-responsive sites in participant 442, the electrode activity distribution in PC coefficient space during the CI task (**b**) and conversation (**c**). The dashed outlines indicate clusters and the numbers denote electrode identity. **d**, The PC coefficients for electrodes in each cluster from **c**. The individual electrode coefficients are depicted as points, the averages are indicated with bars and functional PC categorization is designated by bar colour. **e**, **f**, The average activity of perception and planning PCs aligned to the offset of experimenter speech (**e**) and planning and production PCs aligned to the onset of participant speech (**f**) in the CI task (left) and conversation (right). The thin and thick lines indicate responses from individual participants and the average across participants, respectively. Periods of significantly elevated activity for each participant are denoted with bars under plots ($P < 0.05$, permutation test). **g**, Example production and planning PC activity during conversation with participant 436. The activity is displayed both as a waveform and by events whose thickness is scaled to PC amplitude (75th to 95th percentile). **h**, **i**, Timing of production (**h**) and planning (**i**) events for all interactions during conversation with participant 436. The events aligned to participant turn onset and ordered by gap duration. The grey boxes represent experimenter and participant turn timing and unfilled pauses during participant turns left as whitespace. **j**–**l**, Empirical cumulative distribution functions of median perception, planning and production PC activity in experimenter turns (**j**), interturn gaps (**k**) and participant turns (**l**). Each line represents data from an individual participant. The inset plots depict the median value for each distribution as dots and the average median value across PCs as bars.

Each cluster category was found to be highly weighted for an individual PC (Fig. 4d); therefore, we employed these maximally weighted PCs as an aggregate signal reflecting the overall activity of each cluster and further analysed two perception-related PCs, six planning-related PCs and five production-related PCs (Extended Data Fig. 7). Across participants, we observed that planning PCs displayed a significant peak in average activity near the offset of experimenter speech that continued into the interturn gap for both the task and conversation; conversely, perception PCs were active during the experimenter turn but sharply decreased in activity at the offset of experimenter speech ($P < 0.05$, permutation test; Fig. 4e). Planning PCs likewise exhibited a peak in activity before participant turn onset for both the task and conversation, whereas production PC activity was typically restricted to participant speech in both contexts ($P < 0.05$, permutation test; Fig. 4f). These results demonstrate that the functional properties of PCs related to speech perception, planning and production during unconstrained turn-taking are congruent with the responses observed in the structured CI task.

Finally, we examined planning PC activity during individual turn-taking interactions to provide a first description of speech planning-related dynamics during natural conversation. We observed planning-related PCs to be most active during experimenter turns and silent gaps before participant turns; these responses differed dramatically from production-related PCs, the activity of which was largely restricted to periods of articulation (Fig. 4g–i). Across participants, we likewise found that planning PCs displayed maximal activity levels during interturn gaps whereas the perception and production PCs were most active during the experimenter and participant turns, respectively (Fig. 4j–l, Extended Data Fig. 6g). Taken together, the functional properties displayed by the planning-related PCs are consistent with speech preparation during conversation, indicating that these signals represent the neural correlates of the planning processes enabling rapid turn-taking during natural language use.

Discussion

Conversational turn-taking requires a complex interplay of simultaneous speech perception and planning coupled with finely timed speech production. We used ECoG recordings to isolate planning-related dynamics and found these responses to be functionally and anatomically distinct from those underlying speech perception and production. We observed these discrete classes of neural activity during both structured interactions and unconstrained conversation, thus paving the way for future studies of naturalistic speech generation by providing a description of neural dynamics arising in real-world language use. Although previous work has suggested that neural circuitry related to human language is highly multimodal[24,25] and distributed[30], our findings indicate that separate cortical modules perform distinct functions within the language network during ethologically relevant interactive speech production[8,15]. This conclusion is consistent with perturbation studies of human language circuitry[16,17,34,35] as well as investigations of cortically dependent non-human vocalizations demonstrating the existence of anatomically distinct subcircuits within larger vocal production systems[36,37].

Our findings demonstrate that planning-related responses during interaction are largely restricted to a frontotemporal circuit centred on the cIFG and the cMFG. The caudal portion of the IFG (pars opercularis

and pars triangularis) is thought to represent Broca's region, which is classically considered to be crucial for speech production[10,38]. However, more recent work has demonstrated that this region is active before—rather than during—articulation[5,17]. Using our interactive paradigm, we also observe that responses in the cIFG occur primarily before speech initiation, and we show that this activity is specifically associated with planning speech. Unlike the cIFG, the cMFG has typically not been considered important for speech planning or articulation[12,39], although a case of pure apraxia of speech following its focal resection has been recently reported[11]. Here we show that the cMFG is involved in speech-selective planning during interaction. This region displays robust anatomical connectivity with the cIFG[40], as well as other language-related regions[11,41], and therefore represents an important cortical site for future study as its role in language generation has been largely overlooked.

It is likely that the cIFG and the cMFG perform distinct functions related to speech planning, as evidenced by their differing levels of speech selectivity exhibited during the CR tasks. However, future work is necessary to elucidate the specific planning subprocesses that are executed by these and other nodes within the identified planning network and to uncover how these subprocesses unfold in real time during turn-taking. Nevertheless, the isolation of this planning circuitry represents an important advance towards understanding how the human brain generates language in naturalistic contexts and complements parallel efforts in the wider field of neuroscience aimed at quantifying natural behaviour broadly[42]. Furthermore, this work helps lay a foundation for future studies investigating the biological basis of communication disorders—such as stuttering and apraxia of speech—where abnormalities in speech planning disrupt social language use[43,44].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04270-z.

1. Sacks, H., Schegloff, E. A. & Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language* **50**, 696–735 (1974).
2. Levinson, S. C. & Torreira, F. Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* **6**, 731 (2015).
3. Stivers, T. et al. Universals and cultural variation in turn-taking in conversation. *Proc. Natl Acad. Sci. USA* **106**, 10587–10592 (2009).
4. Schiffman, H. R. *Sensation and Perception: An Integrated Approach* (Wiley, 2001).
5. Flinker, A. et al. Redefining the role of Broca's area in speech. *Proc. Natl Acad. Sci. USA* **112**, 2871–2875 (2015).
6. 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).
7. Mirman, D., Kraft, A. E., Harvey, D. Y., Brecher, A. R. & Schwartz, M. F. Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cogn. Affect. Behav. Neurosci.* **19**, 1286–1298 (2019).
8. Guenther, F. H. *Neural Control of Speech* (MIT, 2016).
9. Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* **326**, 445–449 (2009).
10. Broca, P. Remarques sur le siege de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bull. Mem. Soc. Anat. Paris* **36**, 330–356 (1861).
11. Chang, E. F. et al. Pure apraxia of speech after resection based in the posterior middle frontal gyrus. *Neurosurgery* **87**, E383–E389 (2020).
12. Brass, M. & von Cramon, D. Y. The role of the frontal cortex in task preparation. *Cereb. Cortex* **12**, 908–914 (2002).
13. Sierpowska, J. et al. Involvement of the middle frontal gyrus in language switching as revealed by electrical stimulation mapping and functional magnetic resonance imaging in bilingual brain tumor patients. *Cortex* **99**, 78–92 (2018).
14. Levinson, S. C. Turn-taking in human communication-origins and implications for language processing. *Trends Cogn. Sci.* **20**, 6–14 (2016).
15. Indefrey, P. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* **2**, 255 (2011).
16. Schuhmann, T., Schiller, N. O., Goebel, R. & Sack, A. T. The temporal characteristics of functional activation in Broca's area during overt picture naming. *Cortex* **45**, 1111–1116 (2009).
17. Ferpozzi, V. et al. Broca's area as a pre-articulatory phonetic encoder: gating the motor program. *Front. Hum. Neurosci.* **12**, 64 (2018).
18. Alario, F. X., Chainay, H., Lehericy, S. & Cohen, L. The role of the supplementary motor area (SMA) in word production. *Brain Res.* **1076**, 129–143 (2006).
19. Ramanarayanan, V., Goldstein, L., Byrd, D. & Narayanan, S. S. An investigation of articulatory setting using real-time magnetic resonance imaging. *J. Acoust. Soc. Am.* **134**, 510–519 (2013).
20. Bogels, S., Magyari, L. & Levinson, S. C. Neural signatures of response planning occur midway through an incoming question in conversation. *Sci Rep.* **5**, 12881 (2015).
21. Ferreira, F. & Swets, B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* **46**, 57–84 (2002).
22. Wagner, V., Jescheniak, J. D. & Schriefers, H. On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 423–440 (2010).
23. Dubey, A. & Ray, S. Cortical electrocorticogram (ECoG) is a local signal. *J. Neurosci.* **39**, 4299–4311 (2019).
24. Cheung, C., Hamiton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).
25. Glanz Iljina, O. et al. Real-life speech production and perception have a shared premotor-cortical substrate. *Sci. Rep.* **8**, 8898 (2018).
26. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).
27. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
28. Flinker, A., Chang, E. F., Barbaro, N. M., Berger, M. S. & Knight, R. T. Sub-centimeter language organization in the human temporal lobe. *Brain Lang.* **117**, 103–109 (2011).
29. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).
30. Cogan, G. B. et al. Sensory-motor transformations for speech occur bilaterally. *Nature* **507**, 94–98 (2014).
31. Kotz, S. A. et al. Lexicality drives audio-motor transformations in Broca's area. *Brain Lang.* **112**, 3–11 (2010).
32. Fadiga, L. & Craighero, L. Hand actions and speech representation in Broca's area. *Cortex* **42**, 486–490 (2006).
33. Knudsen, B., Creemers, A. & Meyer, A. S. Forgotten little words: how backchannels and particles may facilitate speech planning in conversation? *Front. Psychol.* **11**, 593671 (2020).
34. Long, M. A. et al. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
35. Tate, M. C., Herbet, G., Moritz-Gasser, S., Tate, J. E. & Duffau, H. Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. *Brain* **137**, 2773–2782 (2014).
36. Long, M. A. & Fee, M. S. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).
37. Okobi, D. E., Jr, Banerjee, A., Matheson, A. M. M., Phelps, S. M. & Long, M. A. Motor cortical control of vocal interaction in neotropical singing mice. *Science* **363**, 983–988 (2019).
38. Tremblay, P. & Dick, A. S. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang.* **162**, 60–71 (2016).
39. Hosman, T. et al. Auditory cues reveal intended movement information in middle frontal gyrus neuronal ensemble activity of a person with tetraplegia. *Sci Rep.* **11**, 98 (2021).
40. Catani, M. et al. Short frontal lobe connections of the human brain. *Cortex* **48**, 273–291 (2012).
41. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
42. Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
43. Deger, K. & Ziegler, W. Speech motor programming in apraxia of speech. *J. Phon.* **30**, 321–335 (2002).
44. Jackson, E. S. et al. A fNIRS investigation of speech planning and execution in adults who stutter. *Neuroscience* **406**, 73–85 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Participants

Participants were patient volunteers undergoing surgical treatment at the University of Iowa Hospitals and Clinics for medically intractable epilepsy or brain tumours who consented to participate in research. Patient participants were undergoing treatment with implanted ECoG electrodes for seizure focus determination or for required awake craniotomy. All procedures were approved by the University of Iowa Institutional Review Board. Patients were confirmed to be left-lateralized for language production with Wada testing and/or were confirmed to be right-handed or ambidextrous (Extended Data Table 1).

Behavioural tasks

Three behavioural tasks were included in this experiment.

Critical information task. CI questions were adapted from an established Dutch stimulus set[20,45]. We developed 120 similarly structured English questions, which were divided among three categories (animals, body parts and antonyms). Each category features pairs of questions with the same CI in which the CI was presented either early or late (for example, Fig. 1c–f). The full list of CI questions is provided in Supplementary Data 1.

CI questions were read to the participant by an experimenter, who instructed the participant to respond as quickly as possible. Questions were presented in random order and varied slightly in exact wording from experiment to experiment. All CI questions could be answered with a single word, although participants occasionally responded with a short phrase.

Command-response tasks. The four CR tasks were performed in a randomized order interleaved with CI task trials in all cases except one, where it was completed as a separate block (Extended Data Table 1). A full list of all CR stimuli is presented in Supplementary Data 1. In CR1 (hand movements), participants performed movements that engage muscles unrelated to speech production. Specifically, participants pressed a spring-loaded button using their right hand a variable number of times, ranging from 1 to 4. In CR2 (orofacial movements), participants were instructed to carry out non-speech orofacial tasks, specifically smiling, mouth opening, tongue protrusion and lip puckering. The movements were recorded with a video camera and time-stamped with a common timescale to the ECoG recordings. In CR3 (speech repetition), participants repeated a combination of monosyllabic real words, monosyllabic pseudowords, and/or disyllabic pseudowords. In CR4 (pluralization), participants were instructed to produce the irregular and/or regular plural form of common monosyllabic nouns (for example, 'goose' → 'geese'). See Extended Data Table 1 for the specific tasks completed by each participant.

Unconstrained conversation. Following completion of the CI and CR tasks, the experimenter engaged the participant in unconstrained conversation. Participants were given no specific instructions during this portion of the experiment other than to speak naturally.

Data acquisition

For awake craniotomy patients, local field potentials were recorded either with a custom 64-channel grid engineered at the University of Iowa or with commercially available subdural grids manufactured by Ad-Tech Medical or PMT. Signals were amplified, bandpass filtered (0.7–800 Hz) and sampled at 2,034.5 Hz using a multichannel amplifier and digital acquisition system (PZ2 preamplifier and RZ2 processer; Tucker-Davis Technologies). For chronically implanted epilepsy patients, electrophysiological signals from subdural electrode grids and strips (AdTech) were bandpass filtered (0.1–500 Hz) and recorded at 2,000 Hz with a multichannel amplifier and digital acquisition system (ATLAS system, Neuralynx). Analogue input channels synchronized with the neural recordings additionally marked the timing of participant button presses and the output of a microphone that captured the speech acoustics of the experimenter and participant. Input channels were sampled at 48,828 Hz by the TDT system with OpenEx software and 16,000 Hz by the Neuralynx system with Pegasus software and downsampled offline to 12,000 Hz. In addition to the electrical signals, a video of the participant was also acquired at 24 fps during all experiments. The video was synced to the electrophysiological data after the experiment and provided a secondary high-quality audio recording channel of the experiment, which was sampled at 48 kHz.

Behavioural analysis

The audio acquired with the electrophysiological acquisition system and/or camera was then annotated by a trained phonetician (G.A.C.) to determine the onsets and offsets of all experimenter questions/commands, CI and participant responses for the CI and CR tasks. Trials where the participant failed to respond, requested additional information (for example, asking for the question to be repeated), or produced a filled pause or disfluency (for example, saying 'um' or stuttering) before responding were excluded from further analysis. Trials in the CR task where a participant failed to accurately complete the command (for example, repeating a word when instructed to say the plural form of the word) were excluded from analysis. In the CR task, the onset and offset of hand movements were determined using the electrical signal from the button that was acquired simultaneously with the local field potentials and were defined as the onset of the first press and offset of the last press of the trial, respectively. For the CR trials requiring orofacial movement responses, the onset and offset of movements were estimated as the timing of the first and last frames of the video where movements were clearly observed.

The audio of the unconstrained conversation was also annotated to provide timestamps for the onsets and offsets of all experimenter and participant turns. Turns were considered continuous segments of speech that were uninterrupted by the other speaker. However, experimenter turns were occasionally truncated at silent pauses (that is, unfilled pauses) if the experimenter spoke continuously for periods in excess of 10 s. Silent pauses produced by the participant were annotated within the participant's turns. In addition, during the participant's speech, the experimenter sometimes produced backchannels (for example, 'okay', 'yes', to display interest) that would not result in termination of the participant's turn. In such cases, we did not consider these events turn-taking on behalf of the participant as they did not yield the floor. Interactions where the participant responded with a backchannel were noted (for example, Extended Data Fig. 6a) but were defined as turn-taking events on behalf of the participant. Finally, interactions where speaker overlap rendered reliable annotation impossible were not considered for analysis.

Anatomical reconstructions

Electrode localization in intraoperative patients. In intraoperative cases, electrode localization and coregistration was performed using intraoperative photographs and perioperative T1-weighted magnetic resonance (MR) images obtained over the course of the clinical workup. MR images were processed using the 'recon-all' pipeline in FreeSurfer to generate cortical surface meshes and to obtain surface-based coregistration to an anatomical template space[46]. Intraoperative photographs of the craniotomy and electrode arrangement were then aligned to renderings of the surface meshes according to a visual comparison of gyral anatomy. Image-based localization was carried out by two raters (G.A.C. and C.K.K.) and, when discrepant, reconciled by an additional rater (J.D.W.G.). Finally, a subset of electrode locations was identified in image RAS coordinate space by selecting the vertex within the surface mesh nearest the electrode location, as observed in the aligned intraoperative photograph. The remaining locations were determined

through thin-plate spline (TPS) interpolation, according to the geometry of the electrode grid.

Electrode localization in patients with chronically implanted electrodes. Patients with chronically implanted electrodes underwent CT and MR imaging before and immediately after implantation. Electrode locations were identified in post-implantation images based on characteristic metallic artefacts: localized magnetic susceptibility-related voids in MR images and punctate radiodensities in CT images. Electrode coordinates were transformed to the pre-implantation image through an initial linear image coregistration followed by a manually guided TPS warping. Coordinates aligned to corresponding anatomical landmarks in pre- and post-implantation imaging, identified through visual comparison of the linearly coregistered images, served as control points in the TPS warping

Electrode coregistration. For all patients, anatomical categorization of electrode sites was guided by surface-based coregistration and segmentation implemented in Freesurfer[47] using the Desikan-Killiany-Tourville (DKT) atlas[48,49]. Following automatic parcellation, the location of each electrode according to the DKT atlas labels was confirmed by three raters (G.A.C., C.K.K. and J.D.W.G.) and corrected if necessary. Electrode locations were then transformed into Montreal Neurological Institute (MNI) space using symmetric diffeomorphic registration implemented in the ANTs toolbox[50] to obtain a nonlinear coregistration between the patient's pre-operative T1 image and an MNI-aligned template brain (CIT168 template)[51].

For analyses where electrodes from individual participants were rendered onto a canonical cortical surface (for example, Fig. 2c), electrodes were plotted on the gyral surface of the asymmetric version of the 6th-generation template brain provided with the FSL software package. For analyses of electrode response density (for example, Fig. 2f), the proportions of electrodes displaying a given response profile in 1-cm-diameter regions of the canonical cortical surface were calculated for areas with electrode coverage from at least three participants.

ECoG signal pre-processing

Before pre-processing each participant's electrophysiological data, non-responsive channels were first identified and excluded (Supplementary Data 2). Such channels were clearly identifiable as they displayed signal variances markedly lower than active channels. Next, artefactual periods were identified in the active channels by iteratively thresholding at an absolute *z*-score threshold of 9 (calculated across the entire recording) and blanking all signals within 2 s of the suprathreshold period. The *z*-scores were then recomputed, and the procedure was repeated with undiscarded samples until all samples fell below the threshold. Any CI or CR trials occurring within 1 s of a blanked period were not considered for analysis.

Next, the electrophysiological data were pre-processed by filtering stationary and non-stationary line noise using adaptive thresholding applied to coefficients of the demodulated band transform[52] with a bandwidth parameter of 0.25 Hz. The data were then high-pass filtered at 5 Hz with a finite impulse response (FIR) filter and re-referenced to the common average signal on a per-grid basis[53]. Finally, high gamma band amplitude was calculated for each channel by averaging the analytical envelope across 8 frequency bands of logarithmically increasing centre frequency (73–144 Hz) and standard deviations (4.68–6.62 Hz)[29] and resampled at 500 Hz. As artefacts were commonly observed at the beginning and end of recordings, these periods (the first and last roughly 5 s) were blanked in both the raw and high gamma signals. Finally, the high gamma signals for each channel were *z* scored across the entire recording duration. Only electrodes located on the cortical surface were included in the analyses described in this paper. Any sites within white matter, not contacting the brain, or deep within cortical or subcortical structures were not considered, and any electrodes determined to have been located on seizure foci or tissue included in the subsequent resection were excluded from further analysis.

Active site detection using generalized linear modelling

To identify electrodes displaying signals significantly correlated to speech perception, planning and production in the CI task, we used mass univariate generalized linear modelling in an approach similar to the statistical parameter mapping procedure used in functional MR imaging[54] and ECoG[55] studies. For each participant, we first constructed four block regressors using the temporal structure of each CI trial (depicted in Extended Data Fig. 1f): (1) a regressor that was active from experimenter question onset to participant answer offset, which was included to regress out any activity that was globally correlated with trial engagement but not with any specific feature of the task (for example, attention); (2) a regressor that was active from question onset to question offset to identify activity correlated with speech perception; (3) a regressor that was active from CI offset to 200 ms before answer onset (as articulatory movements not immediately resulting in an acoustic consequence can occur hundreds of milliseconds before the acoustic onset of speech[19,56]) to identify activity correlated with planning; and (4) a regressor that was active from 200 ms before answer onset until answer offset to identify activity correlated with participant speech production. A 250-ms buffer where no regressors were active was included before trial onset and after trial offset. The regressors for each trial along with the associated high gamma signals (downsampled to 100 Hz for computational efficiency) were then concatenated and GLM fitting was performed. This analysis was carried out using the 'glmfit' function in MATLAB 2020a (MathWorks) with a normal distribution and identity link function specified (therefore, general linear modelling was specifically performed).

To assess significance, we used a resampling method where the GLM was iteratively performed on a randomly selected subset of 80% of trials over 100 repetitions with replacement. This resampling technique was then repeated to apply the trial structure to random epochs of high gamma data whose duration matched the actual data. This allowed us to generate empirical distributions of actual regressor weights, weight *P* values and full-model correlation coefficients as well as null distributions for analogous values generated with shuffled signals. For an electrode's individual regressor weights to be considered significant, two criteria had to be met: (1) the two-tailed 95% confidence interval obtained from the empirical distribution of individual regressor weights could not include 0 or the mean shuffled weight for that regressor and (2) the logscale distribution of the weight *P* values could not include the Bonferroni-corrected α value of 0.05 at the 95th percentile.

To confirm that the responses of individual electrodes were well fit by the specific temporal structure of the CI questions, we used a jittering analysis. Specifically, we performed the resampling technique described above on high gamma signals that were incrementally shifted in uniform 500-ms steps from zero lag to an absolute maximum jitter of 10 s, and the mean full model *R* value over 100 repetitions was recorded at each step for all electrodes (Extended Data Fig. 2a). We then collected random subsets (with replacement) of *R* values from the jittered models such that all participants contributed data from an equal number of electrodes ($n = 44$, 75% of the count from the participant with the fewest electrodes). To quantify the temporal precision of each electrode's fit to the CI task data, we subtracted the median *R* value of large jitters (−10 s to −4 s) from the maximum *R* value of small jitters (−2 s to 2 s), a value we refer to as delta (*D*). We included only large negative jitters in the calculation of *D* to avoid inadvertently fitting activity related to the subsequent trial at high positive jitters with the short duration regressor corresponding to the participant's answer (about 0.5 s). We observed that the pooled distribution of *D* values across participants was bimodal, with electrodes whose activity was fit poorly to the temporal structure of the CI task forming a cluster that was distinct from electrodes whose activity fit the task

structure well. To isolate these two distributions, we then fit the D distribution with a mixture of two Gaussians after excluding values above the 95th percentile to reject outliers (Extended Data Fig. 2b). We then used the Gaussian representing the distribution of poorly fit electrodes to empirically define a D threshold by calculating the value at which 95% would be excluded (1.96 standard deviations from the mean) (Extended Data Fig. 2b). This procedure was repeated 1,000 times and the median D threshold was calculated across iterations. Any electrode whose D value was below this threshold value (0.126) was excluded from further analysis. After thresholding, we found that the number of rejected sites within DKT-defined regions was significantly anticorrelated with the number of responsive electrodes within each region (Spearman $R = -0.8311$, $P < 0.0001$; Extended Data Fig. 2c–e). Therefore, electrodes whose response profiles were not specifically matched to CI task structure were not evenly distributed across the brain, but were more prevalent in regions outside of the language-related circuitry identified with this task.

Electrodes that met all significance testing criteria were then classified based on their mean regressor weights. Electrodes possessing a single positive significant perception, planning and production weights were deemed unmixed perception, planning and production sites, respectively. Electrodes possessing a combination of significant positive perception, planning and production weights were categorized as 'mixed'. Significantly negative regressor weights did not affect electrode classification.

Reduced GLM analyses

As we observed far fewer perception electrodes than planning electrodes, we performed two additional GLM analyses to assess whether the overlapping structure of the perception and planning periods during early CI questions biased the GLM to detect planning-related activity at the expense of perception-related activity. To do so, we first performed the GLM analysis described above but omitted the planning regressor. For this reduced model, we held the D threshold and individual electrode D values constant to maintain consistent exclusion criteria with respect to the full model. We then compared the number of perception-related electrodes detected with the reduced model to that detected with the full model (Extended Data Fig. 3c). We then performed a parallel analysis with a reduced GLM that omitted the perception regressor. With these reduced models, we assessed (1) the number of planning electrodes (as detected with the full model) that gained a significant perception response when the planning regressor was omitted (Extended Data Fig. 3d), and (2) the number of perception electrodes (as detected with the full model) that gained a significant planning response when the perception regressor was omitted (Extended Data Fig. 3e). This analysis allowed us to assess whether the planning regressor also possessed explanatory value for perception-related activity and vice versa.

Analysis of CI question types

To assess whether planning-related electrodes displayed significantly elevated activity following CI in both 'early' and 'late' CI trials (Extended Data Fig. 1d) as well as in each of the three question categories (Extended Data Fig. 1e), we averaged high gamma signals aligned to CI for each of these categories. We then calculated the peak-to-trough amplitude of each mean response by subtracting the 25th percentile value of the z-scored high gamma activity during the 2 s before CI onset from the 75th percentile value of the z-scored high gamma activity during the 2 s following CI onset. We then performed the same procedure using random timepoints rather than CI onset times to generate parallel shuffled measures of the peak-to-trough response amplitudes for each condition. Data from all mixed and unmixed planning electrodes were pooled across participants to generate distributions of actual and shuffled response amplitudes for each condition, and differences between these distributions were assessed using a signed-rank test.

CI clustering analyses in GLM weight space

To assess whether functionally relevant classes of electrodes (that is, perception, planning and production) formed distinct clusters in GLM weight space, k-medoids clustering was performed on all responsive electrodes pooled across all participants with the 'kmedoids' function in MATLAB. Clustering was specifically performed in three-dimensional GLM weight space (that is, perception, planning and production weights), and three clusters were assumed.

Using a simulation, we then assessed whether the observed clustering of functional electrode classes in GLM weight space could have resulted from subdividing continuously varying patterns of neural activity. To do this, we generated 253 simulated electrodes (that is, the number of observed responsive electrodes in our dataset) that were randomly assigned a perception weight, planning weight and production weight. The simulated weights for each regressor were generated from independent unimodal normal distributions whose means and standard deviations matched the observed values across responsive electrodes (perception, -0.119 ± 0.313 z; planning, 0.245 ± 0.238 z; production, 0.220 ± 0.377 z). Simulated weights that surpassed the 5th percentile of observed significant weights for each regressor (perception, 0.229 z; planning, 0.164 z; production, 0.205 z) were considered 'responsive' in the model and subjected to further analysis. This method allowed us to match the range of regressor weights of our simulated responsive electrodes to that observed in the actual data while assuming that regressor weights reflect underlying non-categorical, unimodal continua of neural activity as opposed to the discrete response profiles we hypothesized. Next, simulated responsive electrodes were classified as unmixed and mixed perception, planning and production sites in the same manner as was performed for the actual data, and we quantified (1) the proportion of mixed electrodes and (2) the proportion of 'misclustered' unmixed electrodes. To identify misclustered electrodes, each simulated cluster first received an assignment as perception, planning or production; this assignment was determined such that the maximal number of simulated unmixed perception, planning and production electrodes fell into the perception, planning and production cluster, respectively, and each cluster contained at least one unmixed electrode whose type matched its assignment (that is, the planning cluster must contain at least one unmixed planning electrode). The proportion of misclustered unmixed sites was then calculated by summing the number of simulated unmixed perception, planning and production electrodes falling outside of the perception, planning and production clusters, respectively, and dividing by the total number of simulated unmixed electrodes. This analysis was repeated over 100,000 iterations to generate empirical distributions of mixed and misclustered unmixed electrode proportions under the unimodal regime to compare against the observed results. Data from all iterations was considered unless the model failed to generate at least one unmixed electrode of each class.

Assessing the speech selectivity of planning sites

To quantify the specificity of planning responses, another GLM analysis was performed on the CR task data. This analysis was identical to the GLM performed for the CI task, except four planning regressors (CR1–CR4) and three production regressors (CR1 (hand movements), CR2 (non-speech orofacial movements), and CR3 and CR4 (speech)) were employed. As the temporal precision of each site was quantified with the CI task, no jittering analysis was performed. Also, as the CI in this task was longer in duration (for example, "open your mouth" versus a single word), the CI onset was defined as the onset of the planning regressor. This GLM analysis was performed on all unmixed planning and mixed planning–production electrodes (referred to as 'planning electrodes' in the context of the CR task). Mixed planning sites possessing a significant positive perception GLM weight were not included to avoid biasing the analysis with unintended acoustic differences between

the four CR trial types. The mean planning regressor weight was then used as a metric for selectivity for each of the four CR response types.

We assessed the speech selectivity of planning electrodes within individual regions by calculating the proportion of planning electrodes in (1) the caudal IFG (that is, pars triangularis and pars opercularis), (2) the caudal MFG and (3) the precentral gyrus that displayed a significant positive planning weight in each of the four CR tasks; the boundaries of these regions were defined according to the DKT atlas[48,49] (Extended Data Table 2, Supplementary Data 2). To assess the significance of these proportions, we performed a resampling analysis where the planning weights observed across all planning electrodes for each CR task were shuffled and randomly assigned to the electrodes in each of these three regions. The proportion of shuffled significant positive responses were then calculated within each region and the process was repeated 100,000 times to determine the distribution of proportions that would be expected by chance; owing to differences in electrode coverage, this null distribution varied between regions. Any proportions greater than the 97.5th percentile and lower than the 2.5th percentile of each region's null distribution were therefore considered significantly different than chance.

Principal component analysis

Calculation of task-derived and conversation-derived PCs. To reduce the dimensionality of neural activity, PCA using singular value decomposition was performed on the unaveraged high gamma signals from all CI-task responsive electrodes (that is, each electrode is a variable with timepoints as observations). Specifically, we divided the task and conversation epochs of each participant's dataset—which occurred sequentially during the experiments—and performed PCA (implemented with the 'pca' function in MATLAB) on the data from each period independently for all participants. The task period was defined as the epoch beginning 5 s before the onset of the first CI or CR trial and ending 5 s after the last CI or CR trial. In participants where the task began before the start of the electrophysiological recording or during a blanked artefactual period, the task period was defined to begin at the first unblanked sample. The conversation period was defined as the period beginning 5 s before the first experimenter turn onset and ending 5 s after the last participant turn offset. In cases where conversation was still ongoing when the recording was terminated or during a final artefactual period, the conversation period was defined to end at the last unblanked sample. Finally, before performing PCA, each electrode's high gamma signal was smoothed with a 250-ms mean boxcar filter and *z*-scored within the task and conversation epochs independently.

Clustering of task-responsive electrodes in PC coefficient space. To assess network organization in an unsupervised manner, we mathematically identified clusters of task-responsive electrodes for each participant possessing at least two unmixed electrodes of different classes ($n = 6$; Extended Data Table 1) using *k*-medoids clustering (implemented with the 'kmedoids' function in MATLAB). To standardize this analysis across participants, clustering was specifically performed in three-dimensional PC coefficient space using the scalar coefficients (that is, loadings) of the first three PCs while assuming three clusters, as visual inspection revealed at least three main electrode clusters in all cases (Extended Data Fig. 7a–f). The coefficients of the first three PCs were used because (1) these PCs individually explained at least 5% of the variance for the task data in all participants and for the conversation data in 7/8 participants while also cumulatively explaining at least one-third of the variance in all participants for both behaviours, (2) these PCs fell significantly above the linear decay phase in the scree plots for the task and conversation datasets in 7/8 and 6/8 participants, respectively, and (3) when all participant data were pooled, the first three PCs individually explained at least 5% of the variance and cumulatively explained at least one-third of the variance on average for both the task and conversation while also falling significantly above the linear decay phase (Extended Data Fig. 6c). Note that the linear decay phase was estimated for each participant using the 95% confidence interval of a line fit to the centre of the scree plot (that is, data from the middle 50% of PCs centred at the half maximum PC) via the 'fitlm' function in MATLAB—thus allowing for the 'elbow' of the scree plot to be empirically defined. This process was performed across participants by similarly fitting the pooled data from the middle 50% of the first 13 PCs (the number possessed by the participant with the fewest task-response electrodes, and consequently, PCs) from all participants (Extended Data Fig. 6c).

To assess whether the correlational structure among electrodes was stable between the task-based behaviour and natural conversation, we then identified electrode clusters in the conversation time period that were analogous to those observed during the task period. Specifically, analogous clusters were defined as the task cluster and conversation cluster that possessed the highest number of common electrodes. For all six participants, each task-related electrode cluster was analogous to a single, unique cluster in the conversation data such that 18 pairs of analogous clusters were identified (Extended Data Fig. 7). We then calculated the number of electrodes that 'switched' clusters between the task and conversation datasets (that is, electrodes not remaining within a pair of analogous clusters), and we determined whether the observed number of electrode switches for each participant and the total across participants were significantly less than expected by chance using a permutation test where the conversation cluster membership was shuffled for each participant over 1,000 iterations (Extended Data Fig. 6d).

Functional categorization of electrode clusters. To determine whether electrode clusters in PC coefficient space corresponded to the perception, planning and production electrode classes defined using the CI task and the GLM, we examined the GLM-defined classification of electrodes within each cluster. To do this, we tallied the number of unmixed perception, planning and production electrodes (that is, electrodes with a single significant positive GLM weight) in each cluster and functionally categorized clusters according to which electrode class was most numerous. Across the six participants, 17/18 analogous cluster pairs were thus defined as either a perception, planning or production cluster using this method (analogous clusters always received the same designation); one cluster pair received two categorizations (Extended Data Fig. 7d) and was not included in further analyses.

We next assessed whether the unsupervised clustering analysis successfully recovered the functional classes of neural activity identified with the GLM by assessing the degree to which electrodes within a cluster in PC coefficient space exhibited GLM-defined neural activity congruent with the functional categorization of that cluster. Specifically, we calculated the percentage of electrodes across participants in conversation-related perception, planning and production clusters that displayed significant GLM-defined perception, planning or production activity, respectively. To determine whether the observed values across participants were higher than expected by chance, we assessed significance using a permutation test where the GLM activity classes of each electrode were shuffled over 1,000 iterations (Extended Data Fig. 6e).

Functional categorization of PCs. To examine the neural activity related to speech perception, planning and production during conversation, we assessed time-varying PC scores (that is, the linear combination of electrodes summed in time according to their PC coefficients) as a proxy for the aggregate activity of electrodes within perception-, planning- and production-related clusters. To first determine which PCs corresponded to each electrodes class, we examined the distribution of PC coefficients displayed by electrodes within each cluster category for each participant. Specifically, we determined which of the first three PCs displayed the highest average coefficient value for electrodes in perception clusters, planning clusters and production clusters, and defined these PCs as perception, planning and production PCs, respectively. In two participants where the vast majority of

task-related responses were defined as planning (472, 100%; 510, 97.5%), we considered the PC with the highest average coefficient across all planning-related electrodes as the planning PC (Extended Data Fig. 7g, h). Finally, to avoid assessing PCs representing mixed response profiles, we did not include any clusters containing ≥50% mixed response electrodes (for example, a planning cluster containing five unmixed planning and five mixed planning–production electrodes) in this analysis (n = 3/17 clusters; Extended Data Table 1, Extended Data Fig. 7a–f). Using this approach, we isolated two perception PCs, six planning PCs and five production PCs across eight participants in both the task and conversation datasets (Extended Data Fig. 7).

We then assessed the average response profiles of the perception, planning and production PCs by calculating (1) the mean planning and perception PC scores aligned to experimenter speech offset in the CI task and conversation, and (2) the average planning and production PC scores aligned to participant speech onset in the CI task and conversation. Periods of significantly elevated average PC activity (that is, PC scores) were identified using a permutation test where average responses for each PC were calculated when aligned to trial-number-matched random timepoints over 1,000 iterations; any time bins of the average response that fell above the 95th percentile value of the shuffled average responses were considered significant.

Lastly, we quantified PC activity with respect to unconstrained turn-taking behaviour in natural conversation by calculating the median z-scored PC score amplitude during all (1) experimenter turns, (2) intraturn gaps and (3) participant turns. Consistent with the previous analyses, participant turn onset was defined as 200 ms before the acoustic onset of participant speech to account for silent articulatory movements. Owing to this correction, any gaps less than 200 ms in duration were not considered for analysis. Finally, for all participants with multiple functionally categorized PCs (n = 4), the difference in median amplitude between PCs was assessed for significance using a signed-rank test (Extended Data Fig. 6g).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data used in these analyses are not publicly available owing to concerns regarding patient privacy; however, the corresponding author will provide deidentified primary data upon request.

Code availability

The corresponding author will provide the MATLAB code used in this study for analysis of ECoG and behavioural data upon request.

45. Bogels, S., Casillas, M. & Levinson, S. C. Planning versus comprehension in turn-taking: fast responders show reduced anticipatory processing of the question. *Neuropsychologia* **109**, 295–310 (2018).
46. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
47. Fischl, B. et al. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**, 11–22 (2004).
48. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 171 (2012).
49. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
50. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
51. Tyszka, J. M. & Pauli, W. M. In vivo delineation of subdivisions of the human amygdaloid complex in a high-resolution group template. *Hum. Brain Mapp.* **37**, 3979–3998 (2016).
52. Kovach, C. K. & Gander, P. E. The demodulated band transform. *J. Neurosci. Methods* **261**, 135–154 (2016).
53. Liu, Y., Coon, W. G., Pesters, A., de, B. P. & Schalk, G. The effects of spatial filtering and artifacts on electrocorticographic signals. *J. Neural Eng.* **12**, 056008 (2015).
54. Friston, K. J. et al. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1995).
55. Qian, T., Wu, W., Zhou, W., Gao, S. & Hong, B. in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2347–2350.
56. Tilsen, S. et al. Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS ONE* **11**, e0146813 (2016).

Acknowledgements We thank A. Flinker, E. Jackson, J. Krivokapić, D. Schneider, N. Tritsch and members of the Long laboratory for comments on earlier versions of this manuscript; A. Ramirez-Cardenas, H. Chen, K. Ibayashi, H. Kawasaki, K. Nourski, H. Oya, A. Rhone and B. Snoad for help with data collection; and F. Guenther and N. Majaj for helpful conversations. This research was supported by R01 DC019354 (M.A.L.), R01 DC015260 (J.D.W.G.) and Simons Collaboration on the Global Brain (M.A.L.).

Author contributions G.A.C. and M.A.L. conceived the study and designed the experiments; G.A.C., C.K.K., J.D.W.G. and M.A.L. conducted the research; G.A.C., C.K.K. and M.A.L. performed data analyses; G.A.C., C.K.K. and M.A.L. created the figures; G.A.C. and M.A.L. wrote the initial draft of the manuscript; G.A.C., C.K.K., M.A.H., J.D.W.G. and M.A.L. edited and reviewed the final manuscript. J.D.W.G. and M.A.L. acquired funding; J.D.W.G., M.A.H. and M.A.L. supervised the project.

Competing interests The authors declare no competing interests.

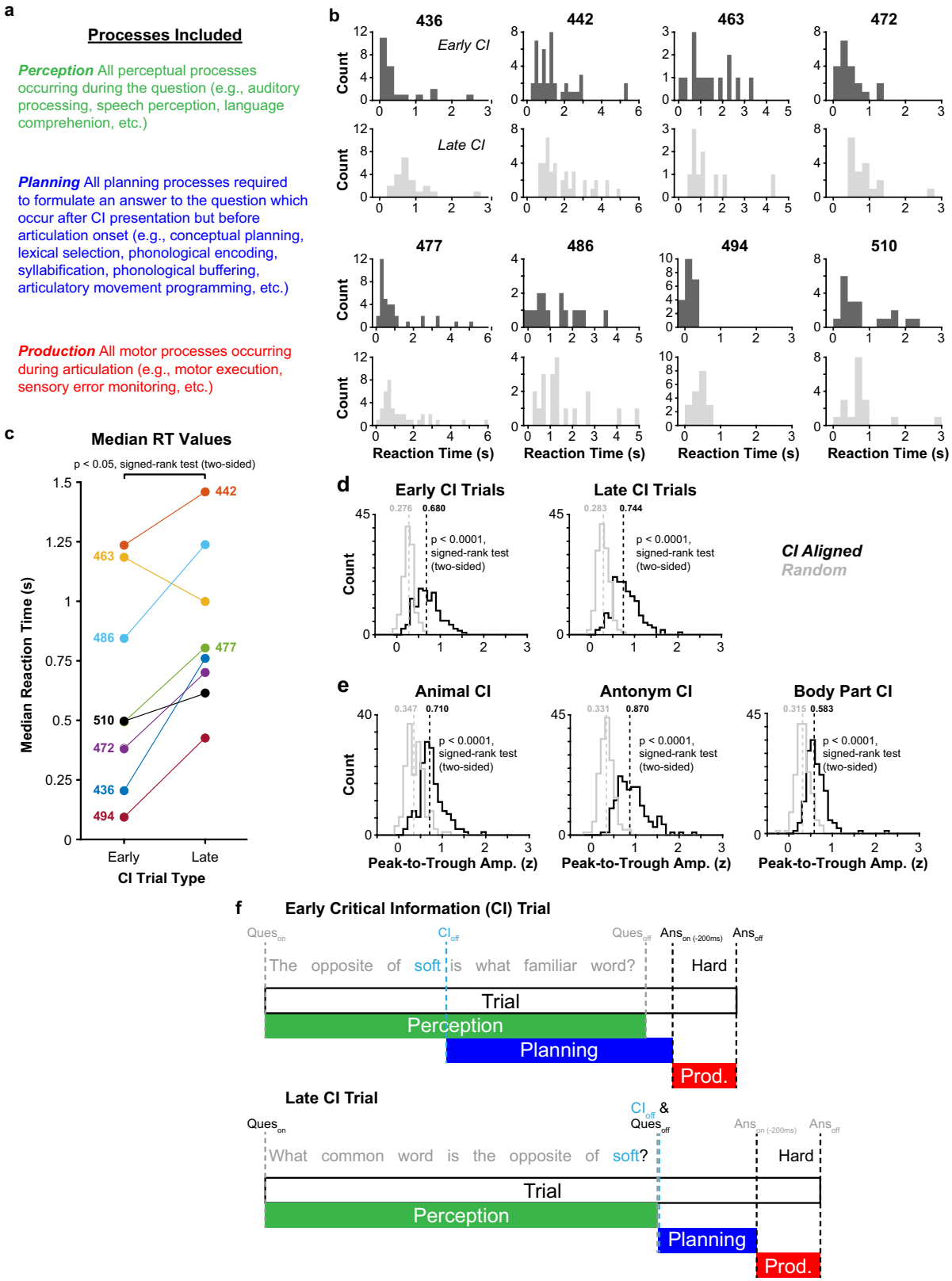
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04270-z.

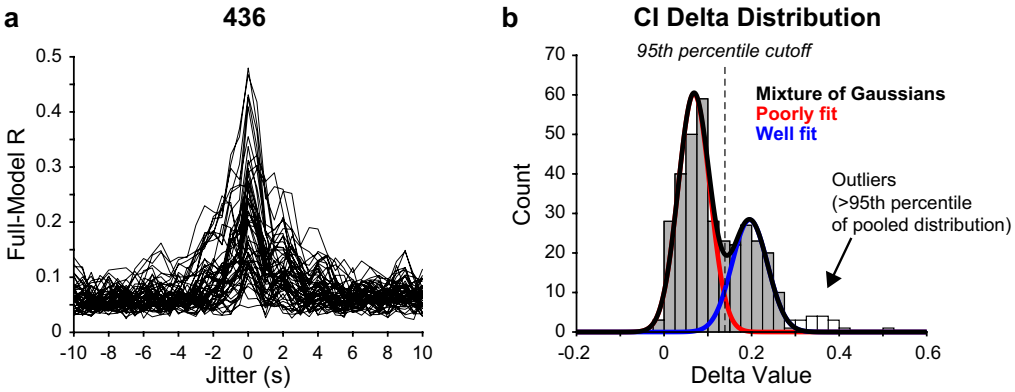
Correspondence and requests for materials should be addressed to Michael A. Long.

Peer review information *Nature* thanks Gregory Cogan, Uri Hasson and Frederic Theunissen for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Behaviour during the CI task. a, Description of subprocesses assumed to occur during the perception, planning, and production windows of the CI task. **b**, Histograms of reaction times (RT) in early and late CI trials for all participants. **c**, Median RT values for early and late CI trials for all participants. **d**, **e**, Histograms depicting the distribution of average peak-to-trough response amplitudes for all electrodes displaying planning-related responses when aligned to CI onset in early and late trials (**d**) and different CI question types (**e**); median values for each distribution are indicated. Observed data (in black) are compared with a null distribution (in grey) consisting of randomly chosen timepoints (Methods). **f**, Schematics displaying GLM regressor structure for an early (top) and a late (bottom) variant of an example CI task question.



c

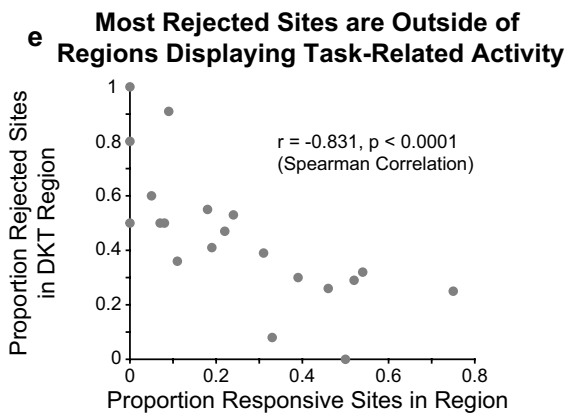
Jittering analysis summary			
Participant	Electrodes Rejected	Electrodes Passing*	Total Electrodes Analyzed**
436	15	29	44
442	75	26	101
463	13	40	53
472	24	17	41
477	49	34	83
486	16	20	36
494	87	68	155
510	14	40	54
Total	293	274	567

*All electrodes with at least one significant regressor weight, including those which were not included for further analysis, due either: 1) being located over seizure foci or tumors, or 2) displaying only negative regressor weights (n = 5) or only global trial-related activity (n = 0) (see Methods).

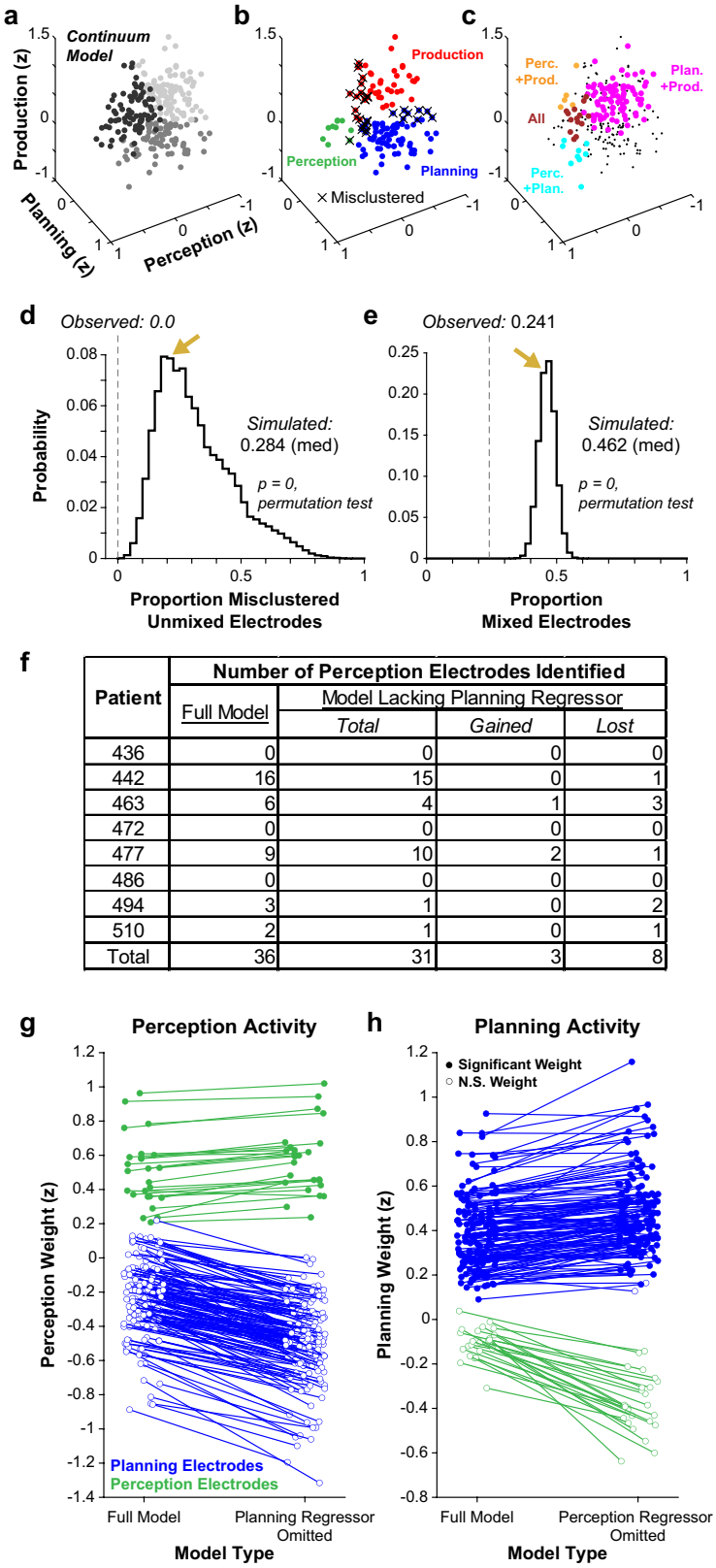
**Any sites with at least one significant regressor weight not over a tumor or seizure foci.

d

Antomical information for electrodes rejected by the jittering analysis					
DKT Parcellation	# Electrodes Rejected	# Responsive Electrodes	Total Electrodes in Region	Prop. Rejected	Prop. Active
Left lateral orbitofrontal cortex	1	0	1	1.00	0.00
Left superior frontal gyrus	10	1	11	0.91	0.09
Left fusiform gyrus	4	0	5	0.80	0.00
Left inferior parietal lobule	12	1	20	0.60	0.05
Left lateral occipital cortex	6	2	11	0.55	0.18
Left supramarginal gyrus	9	4	17	0.53	0.24
Left pars orbitalis	7	1	14	0.50	0.07
Left superior parietal lobule	6	1	12	0.50	0.08
Left temporal pole	4	0	8	0.50	0.00
Left rostral middle frontal gyrus	43	20	91	0.47	0.22
Left superior temporal gyrus	42	20	103	0.41	0.19
Left pars triangularis	35	28	89	0.39	0.31
Left middle temporal gyrus	19	6	53	0.36	0.11
Left caudal middle frontal gyrus	12	20	37	0.32	0.54
Left postcentral gyrus	14	18	46	0.30	0.39
Left inferior temporal gyrus	1	4	12	0.08	0.33
Left precentral gyrus	37	66	128	0.29	0.52
Left pars opercularis	28	50	109	0.26	0.46
Left transverse temporal gyrus	3	9	12	0.25	0.75
Left insular cortex	0	2	4	0.00	0.50
Total	293	253	783		



Extended Data Fig. 2 | GLM temporal jittering analysis. a, Full model R values for GLM fits of jittered high gamma activity from participant 436; each line represents data from an individual electrode. **b**, Example distribution of pooled D values with the fit of two Gaussians overlaid (black). The Gaussian distributions corresponding to well fit (blue) and poorly fit electrodes (red) as well as the 95th percentile of the D distribution for poorly fit electrodes (dashed line) are indicated. D values above the 95th percentile of the pooled distribution were deemed outliers (white bars) and not fitted. **c**, Table summarizing the number of electrodes rejected by the jittering analysis in each participant. **d**, Table reporting the anatomical locations of electrodes rejected by the jittering analysis and electrodes displaying significant activity in the CI task. **e**, Scatterplot depicting the proportion of rejected electrodes within a region as a function of the proportion of responsive electrodes in a region.



f

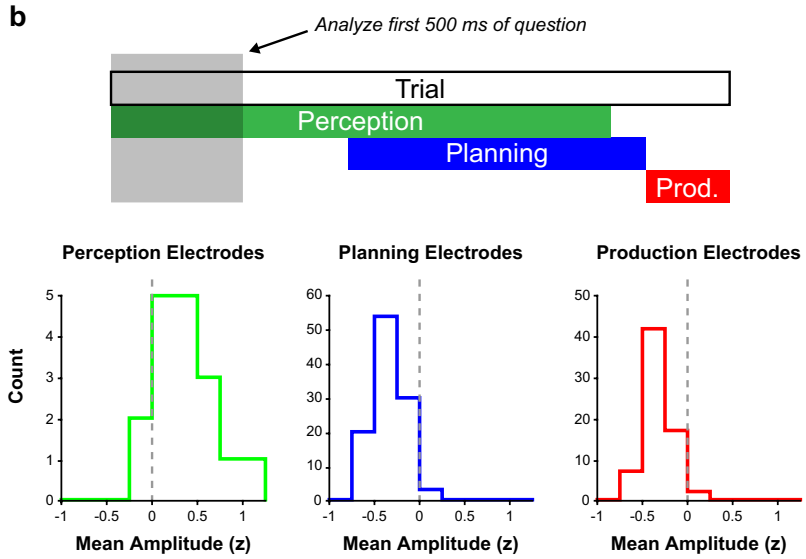
Patient	Number of Perception Electrodes Identified			
	Full Model	Model Lacking Planning Regressor		
		Total	*Gained*	*Lost*
436	0	0	0	0
442	16	15	0	1
463	6	4	1	3
472	0	0	0	0
477	9	10	2	1
486	0	0	0	0
494	3	1	0	2
510	2	1	0	1
Total	36	31	3	8

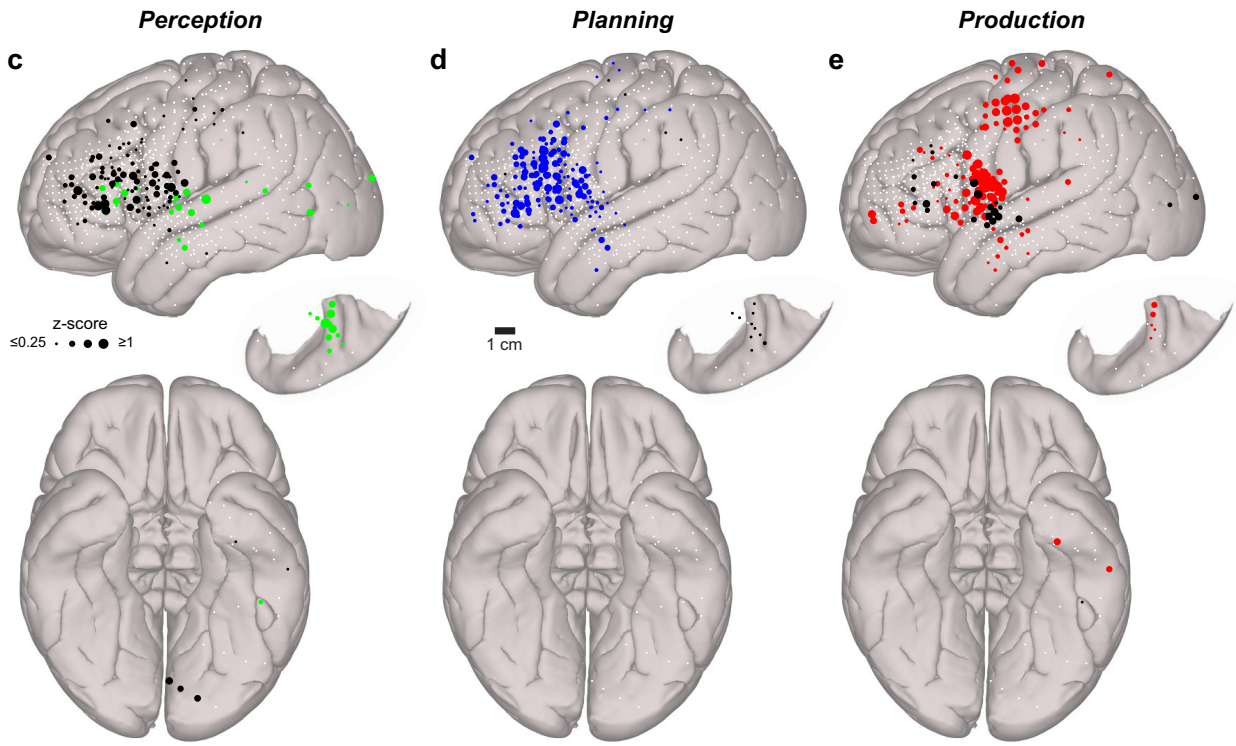
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Analysis of neural activity in the CI task. a, Scatterplot depicting the distribution of all simulated task-responsive electrodes from the continuum model in three-dimensional GLM weight space; cluster membership indicated by greyscale colour. **b**, **c**, Distribution of simulated electrodes from the continuum model displaying responses in one window (i.e., unmixed) of the CI task (**b**) or multiple windows (**c**); response class indicated by colour in **b** and **c** and unmixed electrodes denoted by small black points in **c**. In **b**, simulated unmixed electrodes located outside the cluster primarily containing electrodes of the same type (i.e., 'misclustered') are indicated with an 'X'. **d**, **e**, Histograms depicting the distribution of the proportion of misclustered electrodes responsive during a single task window (i.e., unmixed electrodes) (**d**), and the proportion of electrodes displaying more than one significant positive weight (i.e., mixed electrodes) (**e**) across 100,000 iterations of the continuum model simulation. The median of each distribution as well as the values observed in the actual data (dashed line) are indicated. Gold arrows indicate the bin of each distribution containing the measurements corresponding to the example iteration depicted in panels p, r, and t of Fig. 1. **f**, Table reporting the number of electrodes displaying perception-related responses using either the full model or the reduced GLM lacking a planning regressor. **g**, **h**, Scatterplots depicting perception (**g**) and planning (**h**) GLM weights in the full model and reduced models lacking a planning regressor or perception regressor, respectively. Significant positive weights are denoted with filled points and nonsignificant or significant negative weights are denoted with unfilled points; the *x*-coordinates of each point are randomly jittered by 25% to better visualize filled versus unfilled status. No planning electrodes displayed significant perception responses in the reduced GLM lacking a planning regressor, and no perception electrodes displayed significant planning responses in the reduced GLM lacking a perception regressor.

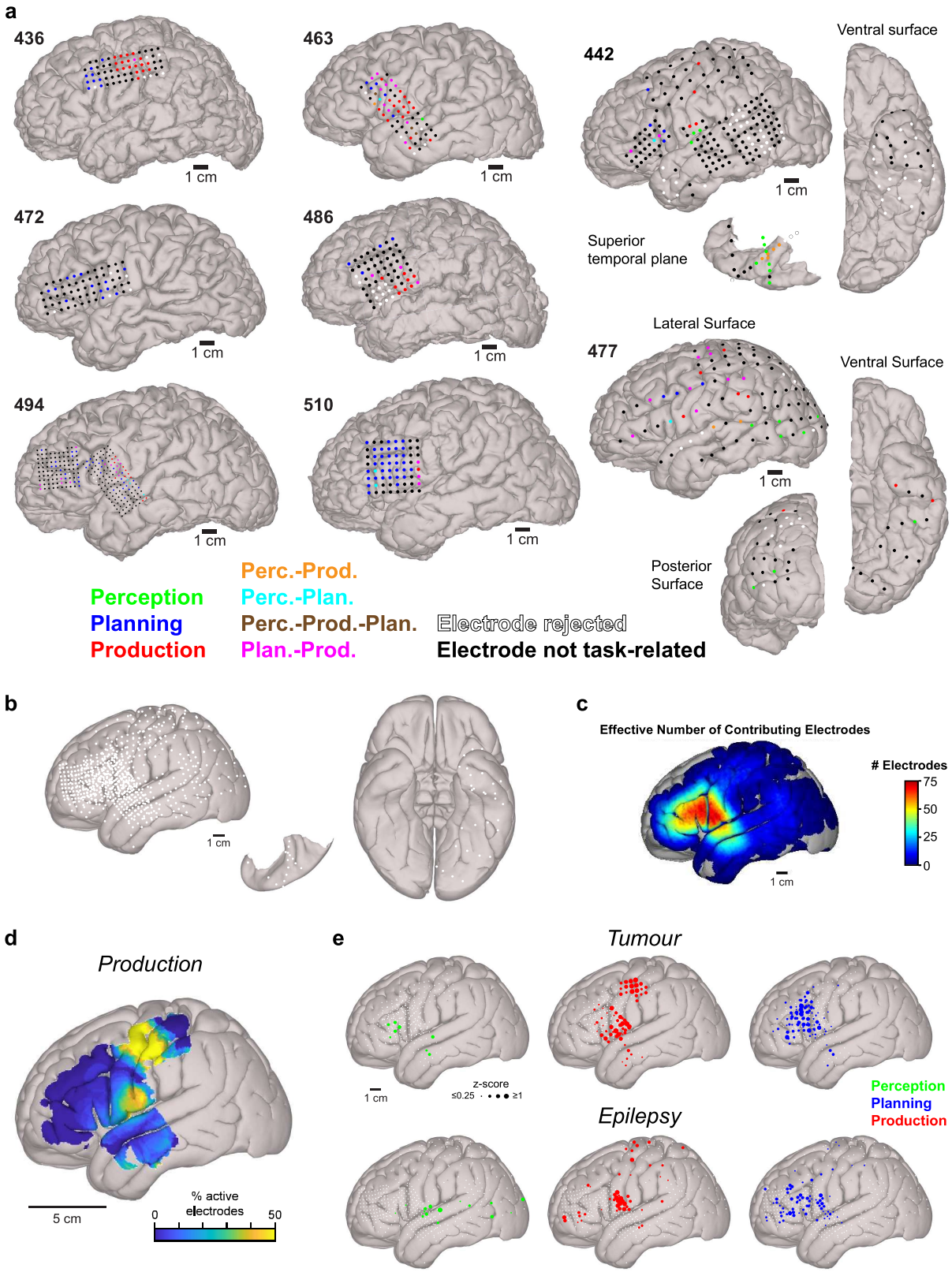
a Summary of significant GLM weights in task-responsive electrodes

GLM Class	# Electrodes with Sig. Perception Weight		# Electrodes with Sig. Planning Weight		# Electrodes with Sig. Production Weight	
	Positive	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*
Perception	36	0	11	10	11	4
Planning	11	75	160	0	45	17
Production	11	41	45	8	121	0





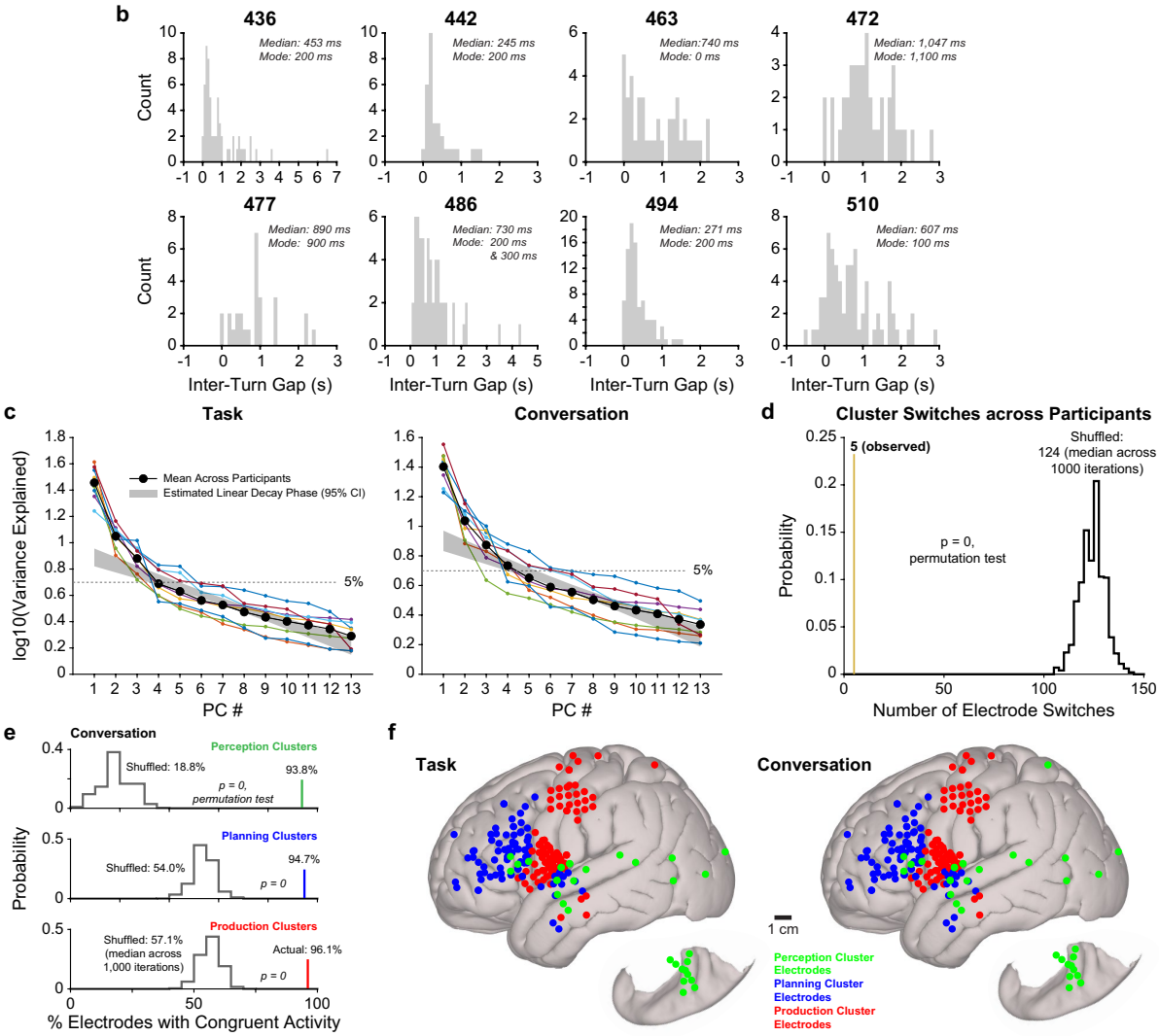
Extended Data Fig. 4 | Additional analyses of task-related activity changes. **a**, Table reporting the number of perception, planning, and production-related electrodes displaying significant positive and negative weights for each GLM regressor. **b**, Histogram depicting mean high gamma amplitude in the first 500 ms of CI questions for all unmixed perception, planning, and production electrodes. **c**, **d**, Canonical cortical surfaces displaying electrodes with significant positive (coloured) or negative (black) GLM weights in the perception (**c**), production (**d**), and planning (**e**) windows of the CI task across all participants. Electrode diameter is scaled to the absolute magnitude of the GLM weight, and electrodes not displaying a significant weight for a given regressor are indicated with small white circles.



Extended Data Fig. 5 | Anatomical analysis of responses. a, Cortical reconstructions for all participants displaying the location of all electrodes; the size of each electrode depicts the actual size of its recording area on the cortical surface. GLM classification is indicated by electrode colour. **b**, Canonical cortical surfaces showing electrode locations from all participants as standard-sized white circles. **c**, Number of electrodes sampling each area of the canonical cortical surface (1 cm diameter spatial smoothing) after pooling electrodes from all participants. **d**, Proportion of electrodes displaying significant production-related responses in the CI task (1-cm-diameter spatial smoothing). **e**, Canonical cortical surfaces displaying electrodes with significant responses related to speech perception, production, and planning in patients with tumour (top) and patients with epilepsy (bottom) separately; electrode diameter scaled to GLM regressor weight. Electrodes not displaying a significant response for a process are depicted as small white circles.

a Additional summary statistics for turn-taking behavior during natural conversation

Patient	Conversation Duration (s)	Backchannels	Total Turns	Backchannel Rate	Mean Turn Duration (s)	StDev (s)
436	371.656	3	61	4.9%	1.834	1.992
442	466.296	15	32	46.9%	6.819	12.081
463	563.421	5	41	12.2%	8.318	9.804
472	296.065	0	41	0.0%	2.545	2.581
477	168.553	2	27	7.4%	2.024	1.567
486	428.520	5	58	8.6%	1.652	1.629
494	350.111	28	90	31.1%	1.163	1.111
510	977.880	8	71	11.3%	3.2634	3.659

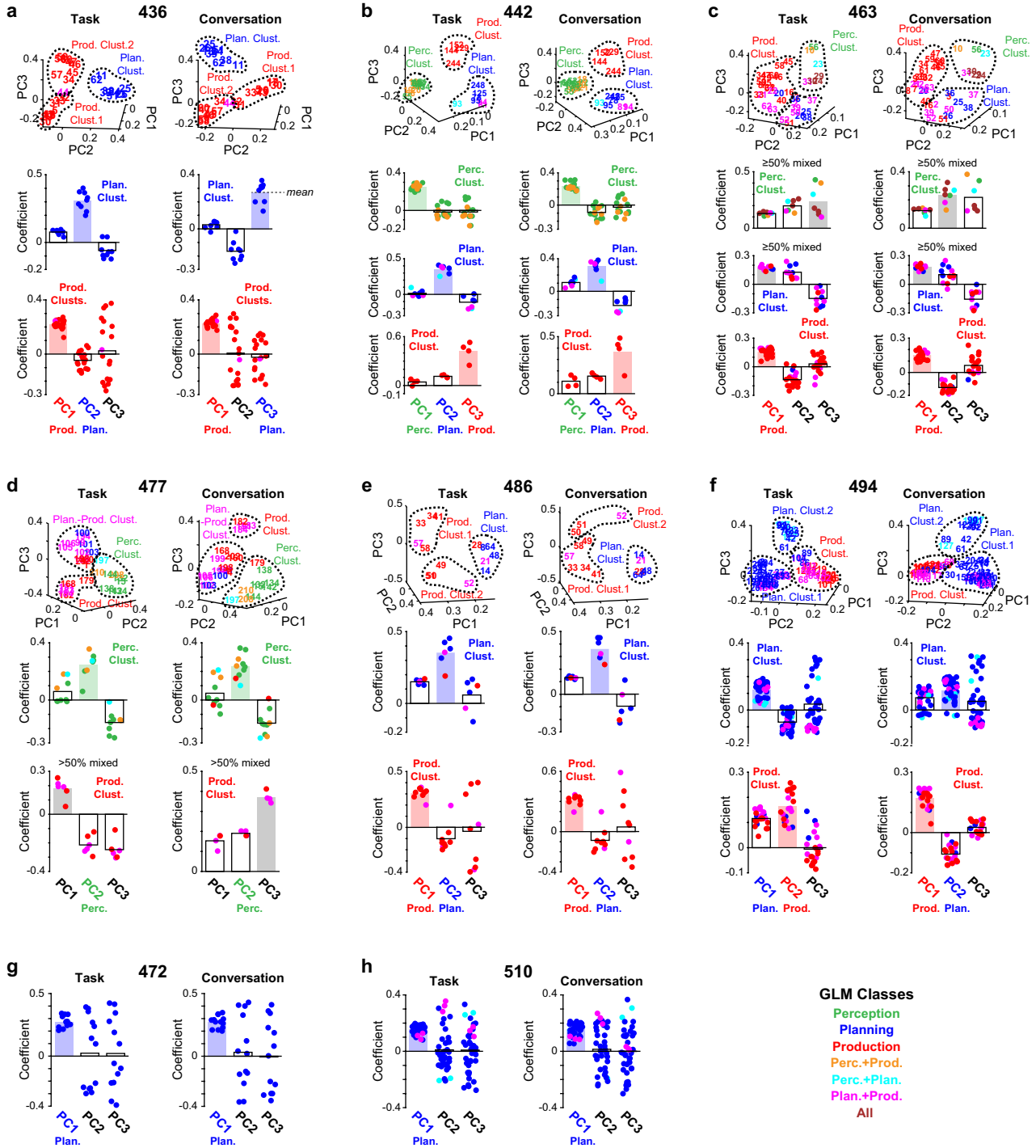


g Summary statistics for PC activity during conversation

Part.	Planning PC Median Value			Production PC Median Value			Paired Sign-Rank Result (Planning vs. Production)		
	Exp. Turn	*Gap*	*Part. Turn*	*Exp. Turn*	*Gap*	*Part. Turn*	*Exp. Turn*	*Gap*	*Part. Turn*
436	0.153	0.473	-0.583	-0.678	-0.525	0.688	**4.16E-08**	**2.66E-05**	**4.84E-11**
442	-0.512	0.221	0.059	-0.331	-0.538	0.096	0.145	0.322	0.443
463				-0.485	-0.374	0.265			
472	-0.555	0.713	0.070						
486	-0.034	0.223	-0.347	-0.307	-0.309	0.231	**5.35E-07**	**1.15E-05**	**9.35E-07**
494	0.151	0.216	-0.766	-0.541	-0.202	0.544	**5.28E-15**	**7.28E-07**	**9.49E-16**
510	0.132	0.697	0.129						
Part.	Perception PC Median Value			Production PC Median Value			Paired Sign-Rank Result (Perception vs. Production)		
	Exp. Turn	*Gap*	*Part. Turn*	*Exp. Turn*	*Gap*	*Part. Turn*	*Exp. Turn*	*Gap*	*Part. Turn*
442	0.603	-0.157	-0.150	-0.331	-0.538	0.096	**3.53E-06**	0.414	0.070
477	0.612	-0.687	-0.323						
Part.	Perception PC Median Value			Planning PC Median Value			Paired Sign-Rank Result (Perception vs. Production)		
	Exp. Turn	*Gap*	*Part. Turn*	*Exp. Turn*	*Gap*	*Part. Turn*	*Exp. Turn*	*Gap*	*Part. Turn*
442	0.603	-0.157	-0.150	-0.512	0.221	0.059	**7.95E-07**	0.259	**0.004**
477	0.612	-0.687	-0.323						

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Additional conversation-related analyses. a, Table reporting additional turn-taking behavioural measures for each participant. **b**, Histograms of gap durations (time between experimenter turn offset and participant turn onset) during unconstrained conversation for each participant; bins are centred on 100 ms increments with a width of 100 ms. **c**, Scree plots for the PCA analysis of high gamma signals in the task (left) and conversation (right) periods of the recordings; data from each participant are represented by thin lines and the average across participants is denoted with a thick black line. The 95% confidence interval of the linear decay phase across participants (Methods) is also indicated. **d**, The observed number of electrodes whose cluster membership was not stable (i.e., switched clusters) between the task and conversation with a histogram depicting the distribution of electrode cluster switches expected by chance. **e**, The observed percentage of electrodes in perception, planning, and production clusters (in conversation-derived PC coefficient space) displaying significant perception, planning, and production responses (per the GLM), respectively, with histograms depicting the percentages expected by chance for each cluster type. **f**, Canonical cortical surfaces displaying the locations of all electrodes in perception, planning, and production clusters across participants (n = 6) in the task (left) and conversation (right). **g**, Table reporting summary statistics for PC activity (i.e., time-varying PC score) during unconstrained conversation for each participant.



Extended Data Fig. 7 | PCA results for individual participants. a–f, For 6 participants possessing sufficient numbers of electrodes belonging to multiple GLM classes (Methods): scatterplots depicting electrode distributions in PC coefficient space in the task and conversation periods (top row). Bar graphs depicting the PC coefficients for all electrodes in perception, planning, or production clusters from the PCA performed on task data and conversation data (bottom rows). Participant number given at top of each panel. **g**, **h**, For 2 participants possessing mainly planning electrodes (Methods, Extended Data Table 1): bar graphs depicting the PC coefficients for all planning-related electrodes from the PCA performed on task data and conversation data. In the bar graphs, the functional categorization of PCs is indicated by filled bars coloured either green (perception), blue (planning), or red (production). Any clusters rejected due to a high proportion (50%) of mixed electrodes are indicated with grey filled bars.

Extended Data Table 1 | Participant information

Code	Age	Sex	Type	Handedness	Wada Testing	Pathology/Diagnosis	Tumor location/seizure loci	Tasks Completed	Data Rejected	Notes
436	58	F	Acute	RH	N/A	Grade II oligodendroglioma (tumour)	Left middle frontal gyrus	CI questions; conversation	N/A	N/A
442	35	F	Chronic	70+	LH language dominant	epilepsy	Multifocal onset	CI questions; CR1, CR2, CR3a, CR3c, CR4a, CR4b; conversation	N/A	CR and CI trials not interleaved.
463	65	M	Acute	ambidextrous	N/A	Glioblastoma (tumour)	Left parahippocampal gyrus	CI questions; CR1, CR2, CR3a, CR4b; conversation	Excluded from CR analysis as only 2 pluralization trials (CR4) were completed correctly; planning and perception electrode clusters rejected from PC functional analysis (≥50% mixed electrodes)	N/A
472	32	M	Acute	RH	N/A	Vascular malformation, focal cortical dysplasia TIIIc (tumour)	Left temporal pole	CI questions; CR1, CR2, CR3a, CR4b; conversation	Excluded from PCA clustering analysis (n = 0 unmixed perception and n = 0 unmixed production electrodes)	N/A
477	24	F	Chronic	80+	LH language dominant	epilepsy/mild gliosis	Left parietal seizure focus	CI questions; CR1, CR2, CR3a, CR3b, CR4a, CR4b; conversation	Production electrode cluster rejected from PC functional analysis (≥50% mixed electrodes); mixed planning-production electrode cluster rejected from PC functional analysis	N/A
486	53	M	Acute	RH	LH language dominant	Glioblastoma (tumour)	Left inferior frontal gyrus	CI questions; CR1, CR2, CR3a, CR4b; conversation	N/A	N/A
494	30	M	Acute	70+	LH language dominant	Hippocampal sclerosis (epilepsy)	Left hippocampus	CI questions; CR1, CR2, CR3a, CR4b; conversation	N/A	N/A
510	54	M	Acute	RH	N/A	Anaplastic Oligodendroglioma (tumour)	Left inferior frontal gyrus	CI questions; CR1, CR2, CR3a, CR4b; conversation	Excluded from PCA clustering analysis (n = 0 unmixed perception and n = 1 unmixed production electrodes).	N/A

Extended Data Table 2 | DKT parcellation of all CI task-responsive electrodes

DKT Label	# Perception Sites	# Production Sites	# Planning Sites	All **
Left superior frontal gyrus	0	0	1	11
Left rostral middle frontal gyrus*	0	3	20	91
Left caudal middle frontal gyrus*	0	5	19	37
Left pars opercularis*	4	14	43	109
Left pars triangularis*	3	6	27	89
Left pars orbitalis	0	1	1	14
Left precentral gyrus*	1	53	29	128
Left postcentral gyrus	1	16	3	46
Left superior parietal lobule	0	1	0	12
Left supramarginal gyrus	0	4	2	17
Left inferior parietal lobule	1	0	0	20
Left superior temporal gyrus*	11	5	14	103
Left middle temporal gyrus	0	6	1	53
Left inferior temporal gyrus	2	2	0	12
Left transverse temporal gyrus	9	5	0	12
Left lateral occipital cortex	2	0	0	11
Left insular cortex	2	0	0	4
Total	36	121	160	769

*Regions that combine to include 152 of 160 (95.0%) of all planning electrodes (inferior and middle frontal gyri, superior temporal gyrus and precentral gyrus).
**Total does not include electrodes that are located outside of the anatomical structures listed in this table.

nature portfolio

Corresponding author(s): Michael A. Long

Last updated by author(s): Nov 9, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	As now stated in the Methods, we used commercially available acquisition software from Neuralynx (Pegasus) and TDT (OpenEx) for data collection.
Data analysis	Custom code written in MATLAB 2020a was used to analyze the data for this study. All statistical tests were performed using the software included in MATLAB 2020a. Accessibility of data analysis code can be found in the Code Availability statement.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

De-identified data will be provided upon reasonable request to the corresponding author. A Data Availability Statement has now been added to the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	8 participants, 790 total electrodes. No formal sample size calculations were performed, but data were limited by patient availability.
Data exclusions	As detailed in the Methods, approximately 10% of electrodes were either faulty or positioned over resected tissue (i.e., tumor or siezure loci). Those data were excluded from further analysis.
Replication	Data were consistent across all participants, and data related to each individual is included in Extended Data Fig. 5. In two follow up studies using related tasks, we find the same result.
Randomization	Randomization was not relevant to this study as no clinical interventions were applied.
Blinding	Blinding was not relevant to this study as no clinical interventions were applied.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Adult (> 18 years old) patients were recruited regardless of race, gender, etc. with no exclusions. Demographic variables are provided for each participant in Extended Data Table 1.
Recruitment	We included adult neurosurgical patients treated at the University of Iowa Health Center. We exclusively recoded from the left language dominant hemisphere. Underlying condition was either epilepsy or tumor, and this clinical designation was not found to affect the results (Extended Data Fig. 5E).
Ethics oversight	All experiments were approved by the University of Iowa Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.