Article

Parallel encoding of speech in human frontal and temporal lobes

Patrick W. Hullett[1,2], Matthew K. Leonard [2,3], Maria Luisa Gorno-Tempini[1,2], Maria Luisa Mandelli[1,2] & Edward F. Chang [2,3]✉

Models of speech perception are centered around a hierarchy in which auditory representations in the thalamus propagate to primary auditory cortex, then to the lateral temporal cortex, and finally through lateral cortical areas to reach the frontal lobe. However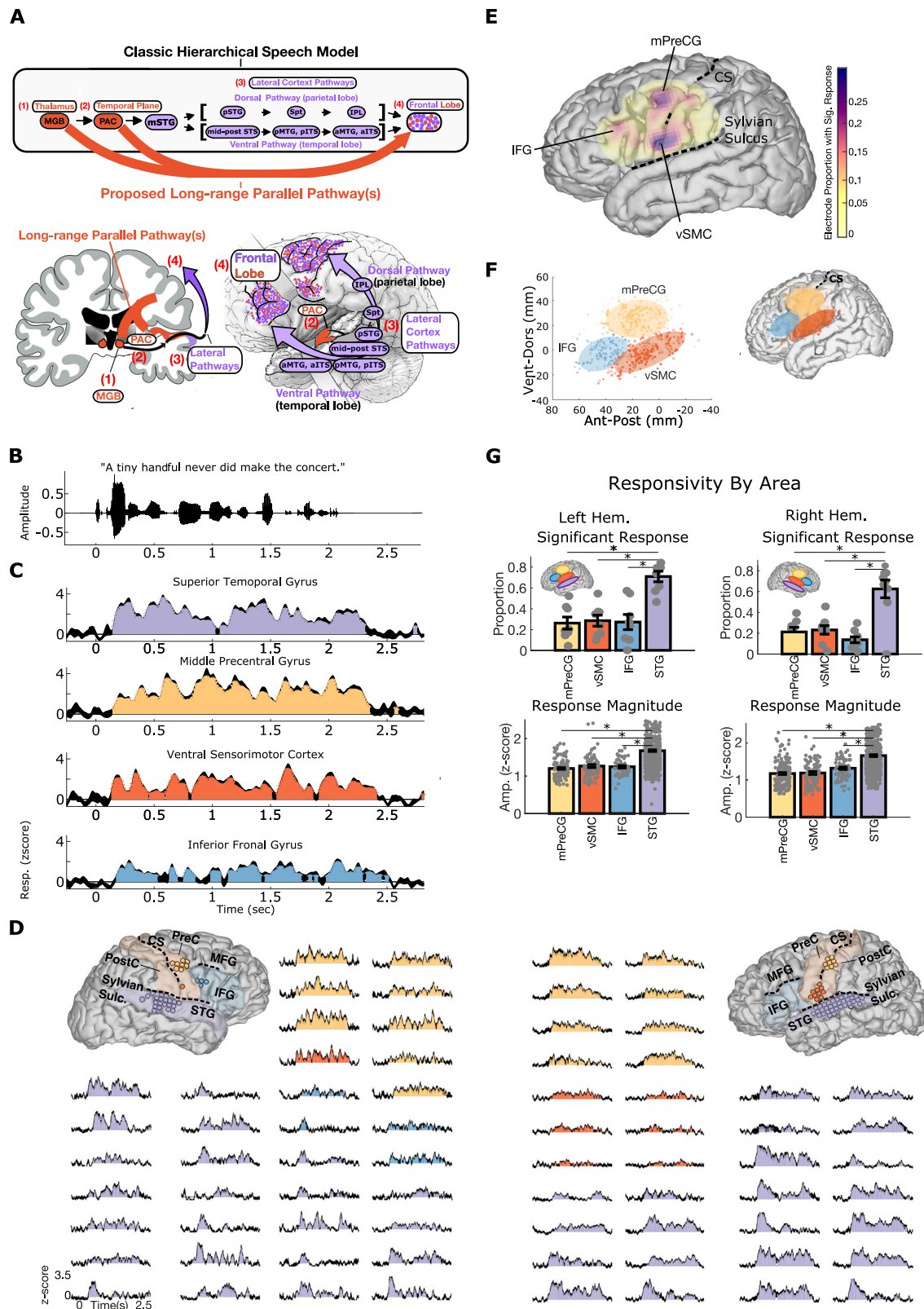, it is unclear if speech-evoked activity in frontal cortex strictly reflects downstream processing from this hierarchical pathway or whether there are long-range parallel connections from the low-level areas (thalamus, primary auditory cortex) to the frontal lobe. Here, we used high-density direct cortical recordings, high-resolution diffusion tractography, and hemodynamic functional connectivity to evaluate for evidence of direct parallel inputs to frontal cortex from low-level areas. We found that neural populations in the frontal lobe show speech-evoked responses that are synchronous with the shortest-latency responses in the superior temporal gyrus (STG) and encode spectrotemporal speech content indistinguishable from spectrotemporal encoding patterns observed in the STG. Additionally, we find white matter tractography and functional connectivity patterns that connect the auditory nucleus of the thalamus and the primary auditory cortex to the frontal lobe. Together, these results support the existence of robust long-range parallel inputs from low-level auditory areas to apical areas in the frontal lobe of the human speech network.

To understand how the brain performs remarkable computational feats like speech perception, it is necessary to understand the cortical regions involved and the network architecture that connects them. The speech network in the human brain is postulated to have a largely hierarchical organization with information flow from subcortical structures to primary auditory cortex, then to lateral temporal cortex through successive dorsal and ventral pathway areas to reach apical targets in the frontal lobe[1,2]. Thus, activity in the frontal lobe during speech perception is thought to reflect downstream computational processes inherited from successive stimulus transformations in areas along the dorsal and ventral lateral cortex pathways (Fig. 1A, purple).

In addition to this hierarchical structure of the speech network, within nearby levels, there are short-range, local, parallel connections[3,4]. However, long-range parallel connections from the bottom of the network (i.e., sensory nuclei in the thalamus or primary auditory cortex) to apical regions in the frontal lobe have yet to be described (Fig. 1A, orange arrows). In support of such long-range parallel pathways, studies in primates found evidence for frontal lobe auditory responses that have short latencies in the range of those observed in lateral temporal cortex[5,6]. Additionally, auditory activity in these areas showed evidence of spectrotemporal representations[7]. These data, although in primates, raise the possibility of frontal cortex responses that may reflect parallel inputs from the auditory thalamus

[1]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [2]Weill Institute for Neurosciences, University of California, San Francisco, CA, USA. [3]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA. ✉e-mail: edward.chang@ucsf.edu

and primary auditory cortex, forming a long-range parallel pathway to the frontal cortex.

To test this hypothesis, we used electrocorticography (ECoG) to identify speech responses in the frontal and temporal lobes in awake participants while they passively listened to natural speech. We specifically asked whether there is evidence for: (1) short-latency speech-evoked responses that are synchronous between the frontal lobe and

the earliest latency responses in the superior temporal gyrus (STG) and (2) whether such activity is explained by spectrotemporal encoding models as would be expected for direct inputs from low-level areas such as the primary auditory cortex or the medial geniculate body of the thalamus. Here, we identify three areas in the frontal lobe cortex that are activated by speech with latencies that occur simultaneously or even before the earliest latencies in the superior temporal gyrus

Fig. 1 | Suprasylvian speech response characteristics. A The classic hierarchical speech pathway from the medical geniculate body (MGB) of the thalamus to the front lobe is shown in the top panel. The orange arrows illustrate the hypothesized long-range parallel pathways from the medial geniculate body (MGB) and primary auditory cortex (PAC) to the frontal lobe. **B** Acoustic waveform of a single example sentence presented to participants. **C** Example responses to a single sentence from the right hemisphere participant in (**D**). **D** Representative example of two participants showing example responses in suprasylvian cortex and superior temporal gyrus (STG) to the single sentence shown in (**B**). The brain reconstructions show the spatial locations of responses to this sentence. This shows high similarity in response characteristics in the frontal lobe and STG. **E** Spatial distribution of significant responses to speech across all participants. This shows three areas of peak response probability in IFG, middle precentral gyrus (mPreCG), and ventral sensorimotor cortex (vSMC) (electrodes collapsed to the left hemisphere). **F** Spatial clustering using a Gaussian mixture model demonstrates three areas of speech responses in IFG, mPreCG, and vSMC. The number of clusters was determined by Bayesian information criterion and silhouette criterion values (both converged on 3

clusters)[13]. **G** Suprasylvian area versus STG responsivity. The proportion of electrodes with significant responses and the mean response amplitude is lower in IFG, mPreCG, and vSMC than in STG ($p < 0.005$, two-sided Wilcoxon rank-sum test, Bonferroni corrected, n proportion samples for left/right hem. = 9/8 IFG, 8/8 mPreCG, 9/8 vSMC, 9/8 STG, n response amplitudes for left/right hem. = 52/84 IFG, 120/180 mPreCG, 98/91 vSMC, 645/548 STG). Each area has no left-right hemispheric differences regarding the proportion or magnitude of significant responses ($p > 0.05$, two-sided Wilcoxon rank-sum test). All data are presented as mean values ± SEM. CS central sulcus; Hem hemisphere; IFG inferior frontal gyrus; IPL inferior parietal lobe; ITS inferior temporal sulcus; MGB medial geniculate body; mPreCG middle precentral gyrus; MTG middle temporal gyrus; PostC postcentral gyrus; PreC precentral gyrus; Spt Sylvian fissure at the parieto-temporal boundary; STG superior temporal gyrus; STS superior temporal sulcus; Sulc. sulcus; vSMC ventral sensorimotor cortex. **A** contains brain illustrations from Kenneth Probst, reproduced with permission, licensed under a Creative Commons Attribution 4.0 https://creativecommons.org/licenses/by/4.0/deed.en.

(STG). Additionally, neural representations in these short-latency populations encode spectrotemporal content that is indistinguishable from populations in STG. Finally, to further establish evidence for anatomic connections that reflect parallel inputs from lower-level areas to the frontal lobe, we use white matter tractography and functional connectivity and find connections between frontal lobe regions and the auditory nucleus of the thalamus (MGB) and primary auditory cortex. Overall, these results demonstrate a fundamental divergence in the hierarchical architecture that is usually assumed to underlie cortical speech processing and shows multiple lines of evidence for long-range parallel inputs from low-level areas to frontal lobe cortical regions in the human speech network.

Results

Frontal lobe areas are activated by passive speech listening
To characterize the extent to which neural populations in frontal lobe respond to speech, ECoG participants passively listened to 10–40 min of natural speech, which consisted of prerecorded sentences (2–4 s each) from the phonetically transcribed TIMIT speech corpus[8]. We extracted activity in the high-gamma (70–150 Hz) range[9], which correlates with spiking activity[10], spike-based tuning properties[11], and reflects contributions from single neurons throughout the cortical depth[12]. At single electrodes, average responses to an example sentence (Fig. 1B) were highly robust and extended through the duration of the sentence in STG and three areas in the frontal lobe, including middle precentral gyrus (mPreCG; Fig. 1C; yellow), ventral sensorimotor cortex (vSMC; Fig. 1C; orange), and inferior frontal gyrus (IFG; Fig. 1C; blue). Figure 1D shows two example participants that illustrate typical electrode coverage in the left and right hemispheres and responses to a single sentence of speech that are qualitatively similar to responses in STG.

Across all 17 participants, we quantified the spatial distribution of responses to speech in the frontal and parietal cortex. We observed significant responses throughout the areas covered by ECoG grids, with three peaks of responsiveness in mPreCG, vSMC, and IFG (Fig. 1E, significant responses defined as electrodes with $p < 0.05$ for any response time bin, Bonferroni corrected for the number of time bins in each sentence, Wilcoxon rank-sum test). To further test whether these three peaks reflect spatially clustered responses, we used Gaussian mixture modeling to cluster the electrodes. The three areas that emerged reflect clusters of electrodes in mPreCG, vSMC, and IFG (Fig. 1F; cluster number = 3 determined using both Bayesian information criterion, $p < 0.05$, and Silhouette criterion[13] $p < 0.05$). Thus, in addition to STG, there are three regions within the frontal lobe cortex that have significant responses to natural speech.
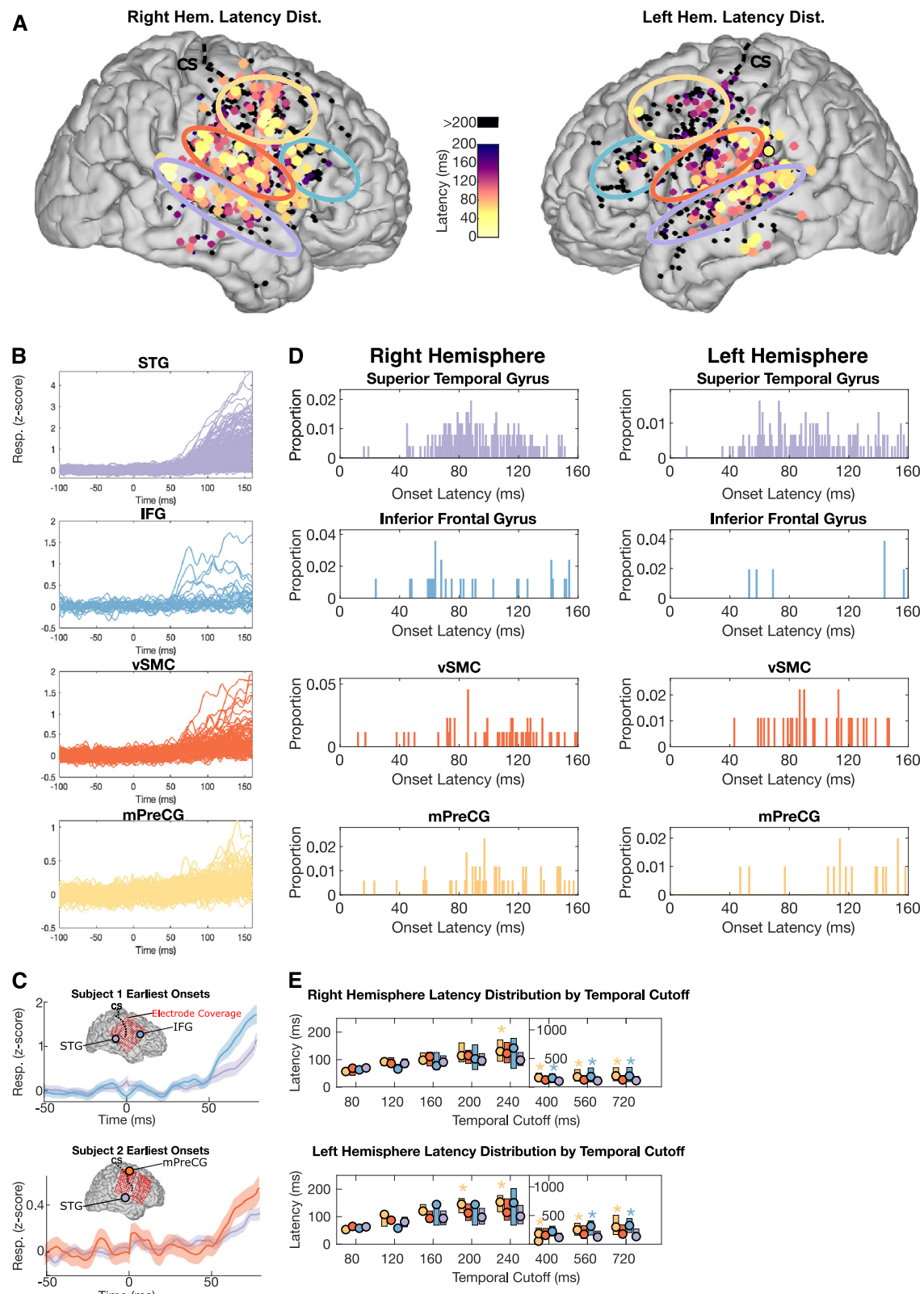
We compared the basic response properties in these three suprasylvian areas to STG, which is known to be a critical area that is

central to speech perception[14,15]. To characterize the overall responsivity, we first quantified the proportion of all electrodes within an area that showed a significant response to speech (Fig. 1G, top row). Although all three regions had electrodes with significant speech-evoked activity in both hemispheres, there was a lower proportion of speech-responsive electrodes in frontal lobe areas compared to the proportion of speech-responsive electrodes in STG ($p < 0.005$, two-sided Wilcoxon rank-sum test, Bonferroni corrected; Fig. 1G, top row). Additionally, while significantly greater than zero, the average response magnitude in suprasylvian areas was lower than STG ($p < 0.005$, two-sided Wilcoxon rank-sum test, Bonferroni corrected; Fig. 1G, bottom row). For both the proportion of significant electrodes and the response magnitude, there were no significant differences between hemispheres ($p > 0.05$; two-sided Wilcoxon rank-sum test), suggesting that frontal lobe speech-evoked activity is similar between right and left hemispheres[16,17].

Frontal lobe areas have short-onset latencies like the superior temporal gyrus
Having established the existence of robust speech-evoked activity throughout bilateral frontal lobe cortex during passive listening, we asked whether any of these neural populations had short-onset latencies that are synchronous or faster than the earliest onset latencies in lateral temporal cortex. If present, this would be consistent with parallel pathways from lower-order auditory areas to frontal and temporal lobes. To address this question, we were careful to only select participants with extensive coverage across STG and these frontal lobe areas within a single participant (Supplementary Fig. 1). The seventeen participants included in this study is a subset of all participants with grid-based coverage over a 10-year span in our center. Additionally, for higher spatial resolution, all participants have specialized high-density recording grids with 4 mm inter-electrode distance in contrast to 10 mm inter-electrode distance used at other centers[18]. Direct visual examination of speech-evoked activity showed near-synchronous shortest latency responses across frontal areas and STG (Fig. 2A, B, D). Furthermore, there were within-participant examples of frontal electrodes with onset latencies synchronous or earlier than the shortest onset latencies in the temporal lobe (Fig. 2C). To quantify this, we calculated the response onset latencies across all participants, similar to prior characterizations in humans and primates[19,20] (Fig. 2A, D). This showed electrodes throughout the temporal, frontal, and parietal lobes with short response latencies, often less than 80 ms.

For each region, we quantified the latency distribution, which showed similar short-latency responses across all areas (Fig. 2D). To test for significant differences in short-onset latencies between areas, we used a temporal cutoff to define the left-sided tail of each

distribution as the short-onset latencies of interest (Fig. 2E). The temporal cutoff ranged from 80 to 720 ms, with significance testing at each cutoff. As shown, onset latencies were not significantly different between frontal lobe areas (IFG, mPreCG, vSMC) and STG for latencies up to 200 ms in the left hemisphere and 240 ms in the right hemisphere (Fig. 2E, $p > 0.05$, two-sample Kolmogorov–Smirnov test, Bonferroni corrected). This data shows short-onset latencies within the same hemisphere are not significantly different between frontal lobe areas and STG. Thus, in response to natural speech, IFG, mPreCG, and vSMC are active beginning at the same time as neural populations in STG, consistent with parallel inputs to STG and frontal lobe areas.

Short-onset latency inputs to the frontal and temporal lobe encode the same spectrotemporal speech information

A feature of most functional-anatomical models of speech is frontal lobe areas receive inputs from the temporal lobe through dorsal and

Fig. 2 | Frontal lobe areas have short-onset latencies like the superior temporal gyrus. A Spatial map of response latencies show short-onset latencies in STG, IFG, mPreCG, and vSMC (onset defined as first 1-ms time bin in which the response *p*-value is <0.05 for 15 consecutive 1-ms bins[19,20]; Wilcoxon rank-sum test). **B** All responses across participants with onset latencies less than 200 ms. **C** Two example participants showing the electrode sites with shortest onset latency in the STG and frontal lobe. As shown, the earliest onset latencies within a subject are similar between the frontal lobe and STG. **D** Latency distribution by area for onset latencies less than 160 ms (the proportion y-axis is relative to the full distribution of latencies up to 1600 ms). This shows the earliest onset latencies are similar across areas and is consistent with low-latency parallel inputs to each area. **E** Latency distribution boxplots for all onset latencies up to the temporal cutoff specified on the x-axis (80–720 ms). There is no significant difference between STG and IFG, mPreCG, or vSMC up to 240 ms in the right hemisphere and 200 ms in the left hemisphere (* = *p* < 0.05, two-sample Kolmogorov–Smirnov test, Bonferroni corrected, n latencies for left/right hem. = 52/84 IFG, 120/180 mPreCG, 98/91 vSMC, 330/274 STG). All box plots show the median (circle), 25th and 75th percentiles (box). These indistinguishable short-onset latency times are consistent with a subset of parallel inputs to the frontal cortex and STG rather than strictly hierarchical inputs from STG to the frontal cortex. CS central sulcus; IFG inferior frontal gyrus; mPreCG middle precentral gyrus; STG superior temporal gyrus; vSMC ventral sensorimotor cortex.

ventral pathways and may reflect qualitatively different (and perhaps higher-order) representations of speech[1,2]. Given the existence of parallel short-latency responses in the frontal and temporal lobes, we asked if the spectrotemporal information encoded in these responses fundamentally differed between lobes. To test this, we computed spectrotemporal receptive fields (STRFs, Fig. 3A) for short-onset latency sites (<200 ms) in STG, IFG, mPreCG, and vSMC. In all three frontal lobe areas, responses to speech were well-predicted by STRFs (mean r = 0.39 ± 24, Fig. 3B), with some electrodes reaching r = 0.75, demonstrating that frontal lobe neural population activity is well-explained by spectrotemporal models.

Next, we asked whether the spectrotemporal content encoded in short-onset (<200 ms) frontal lobe electrodes is similar to STG. Based on visual examination of short-onset STRFs from each frontal lobe area and STG, there was a high degree of similarity in spectrotemporal encoding (Fig. 3C). Additionally, the STRF features correlated with spectrotemporal structure seen in phonetic features similar to STG[21]. For example, the bottom row of Fig. 3C shows broadband excitatory tuning followed shortly by broadband inhibitory tuning. This is the canonical spectrotemporal structure that is tuned for plosives in phonetic feature space[21]. To quantify this we examined the STRF weights and computed three key parameters (Fig. 3D): (1) the best frequency (first column)[22,23], (2) spectral tuning (second column)[24], and (3) temporal tuning (third column)[24,25]. For each of these parameters, there was no significant difference across regions (*p* > 0.05 Kruskal–Wallis test and cosine similarity permutation test of mean spectral and temporal tuning vectors[26]).

While individual STRF tuning properties were not different between STG and frontal lobe areas, the joint combination of these tuning properties in each STRF characterizes the spectrotemporal processing at each site. To jointly characterize spectrotemporal processing for each electrode across regions, we transformed the STRFs into their modulation transfer functions (MTFs)[23,27,28]. MTFs are the modulation domain representation of spectrotemporal processing characterized by STRFs[28,29]. They summarize neural processing in terms of spectral and temporal modulation tuning and are frequently used to characterize processing within the auditory system. As shown in Fig. 3E, the ensemble (mean) modulation transfer function across frontal lobe areas and STG is similar. To quantify this, the mean temporal modulation tuning and spectral modulation tuning distributions derived from the MTFs were calculated (Fig. 3F), and no significant differences in temporal or spectral modulation tuning were observed across frontal lobe areas and STG (*p* > 0.05, cosine similarity permutation test of mean spectral and temporal modulation tuning vectors[26]). Together, these results indicate that the short-latency responses in frontal lobe areas and STG encode the same spectrotemporal representations of speech, consistent with parallel spectrotemporal speech representations in the temporal and frontal lobes.

While short-onset (<200 ms) electrodes in the frontal lobe are well-predicted by spectrotemporal speech representations, it is possible that they reflect higher-order linguistic features that are correlated with the speech spectrogram. To test this alternative, we compared encoding performance for STRFs to semantic-based encoding models generated with word vector representations of speech derived from the FASTTEXT data set[30]. We expected that both spectrotemporal and semantic encoding models would perform well in each region, however, they would differ from each other as a function of short or long onset latency. Specifically, we hypothesized that the STRF would be a better model for electrodes with short-onset latencies since these populations reflect direct inputs from low-level areas such as the medial geniculate body or primary auditory cortex.
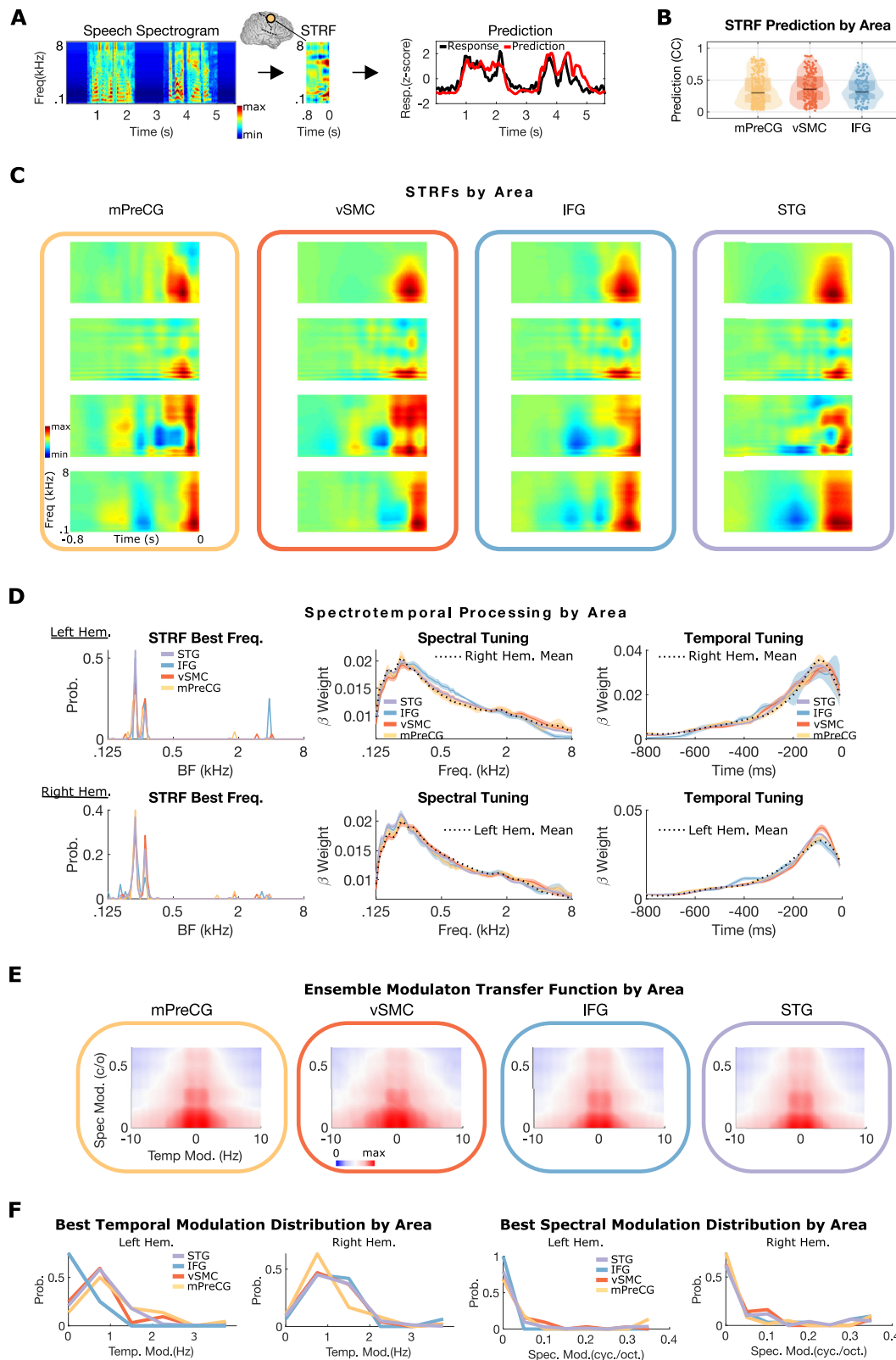
Confirming our hypothesis, we found that electrodes with short-onset latencies (<200 ms) in IFG, mPreCG, and vSMC show significantly higher STRF model predictions than semantic model predictions, consistent with dominant spectrotemporal representations at those sites (Fig. 4A, *p* < 0.01, Wilcoxon rank-sum test). At longer latencies (>200 ms), spectrotemporal and semantic encoding models were not significantly different, further supporting the notion that these short latency electrodes reflect spectrotemporal representations. Overall, these data demonstrate that neural populations in the frontal lobe with short-onset responses encode spectrotemporal representations of sound similar to populations in more low-level auditory areas and STG.

White matter tractography and functional connectivity show connections from thalamus and primary auditory cortex to frontal lobe areas

To establish whether the short-latency responses with STG-like spectrotemporal encoding observed in frontal lobe cortex reflects a direct anatomical pathway from lower-level auditory areas, we used diffusion tensor imaging (DTI) tractography. Based on the known connectivity patterns with STG, we hypothesized that the medial geniculate body (MGB) within the thalamus and the primary auditory cortex within Heschl's gyrus are key areas that might have parallel projections to the frontal lobe cortex.

Using data from 842 individuals from the Human Connectome project[31], we calculated white matter tractography in each hemisphere with the MGB and Heschl's gyrus as seed regions. To localize the MGB, we used MNI coordinates in conjunction with the terminal point of white matter tracts from the inferior colliculus to the thalamus (Fig. 5A). We restricted the suprasylvian target region of interest to the total area spanned by the lateral frontal and parietal cortex. We identified tracts from MGB that projected to IFG, mPreCG, and right vSMC (Fig. 5A). We also identified tracts from Heschl's gyrus that projected to vSMC and left IFG.

To test whether identified tracts relate to functional measures, we also calculated resting-state functional connectivity between the medial geniculate body or Heschl's gyrus to all of the cortex using fMRI. For the MGB seed region, there was significant functional connectivity with IFG, mPreCG, and vSMC (Fig. 5B, *p* < 5 × 10⁻⁵ after peak-level family-wise error (FWE) correction for multiple comparisons). For the Heschl's gyrus seed region, there was significant functional connectivity to mPreCG and vSMC (Fig. 5C, *p* < 5 × 10⁻⁵ after peak-level family-wise error (FWE) correction for multiple comparisons).

Finally, although limited to the supplementary material due to the scarcity of data (Supplementary Fig. 2), we show single pulse stimulation in the primary auditory cortex with cortico-cortical evoked potentials in these frontal lobe areas consistent with the DTI results above. Together, white matter tractography, resting-state functional connectivity, and single-pulse stimulation demonstrate that short-latency evoked activity that encodes speech-relevant spectrotemporal content in frontal lobe cortex is consistent with a parallel auditory pathway that functions alongside the classical hierarchical speech network mediated by STG.

Discussion

We examined the nature of short-latency, speech-evoked activity in frontal lobe areas that are not typically associated with auditory

Fig. 3 | Sites with short-onset latencies in frontal cortex encode the same spectrotemporal information as STG. A Example middle precentral gyrus STRF with the predicted and actual response to a single sentence. Predicted responses are obtained by convolving the spectrogram with the STRF and are proportional to the similarity between the stimulus's spectrotemporal content and the STRF structure. **B** STRF prediction value distributions for all short-latency electrodes with a significant response to speech in mPreCG, VSMC, and IFG (suprasylvian cortex global mean Pearson correlation coefficient 0.39 ± 0.24; tested on held-out data). STRFs, which are predictive of neural responses, characterize spectro-temporal encoding at each site and are consistent with the presence of spectro-temporal representations in these areas. All box plots show the median, 25th and 75th percentiles. Notches approximate the 95% confidence interval of the median. n site prediction values = 46 IFG, 115 mPreCG, 110 vSMC, and 389 for STG. **C** Examples showing the high degree of STRF similarity between frontal lobe areas and STG.

D There is no significant difference in short-onset (<200 ms) STRF tuning para-meters across frontal lobe sites and STG, including: best frequency tuning (p > 0.05 Kruskal–Wallis test), spectral, or temporal tuning (p > 0.05 cosine similarity per-mutation test of mean spectral and temporal tuning vectors). Spectral and tem-poral tuning plots show the mean ± s.e.m for each area in each hemisphere; the total contralateral hemisphere mean tuning is plotted as a dotted line for reference. **E** Ensemble modulation transfer functions by area. The average modulation tuning for each area has a high degree of similarity. **F** There is no significant difference in temporal or spectral modulation tuning between frontal lobe areas and STG (p > 0.05, cosine similarity permutation test of mean spectral and temporal mod-ulation tuning vectors[26]). BF best frequency; CC Pearson correlation coefficient; c/o cycles per octave; Hem. hemisphere; IFG inferior frontal gyrus; mPreCG middle precentral gyrus; Spec. Mod. spectral modulation; STRF spectrotemporal receptive field; Temp. Mod. Temporal Modulation; vSMC ventral sensorimotor cortex.
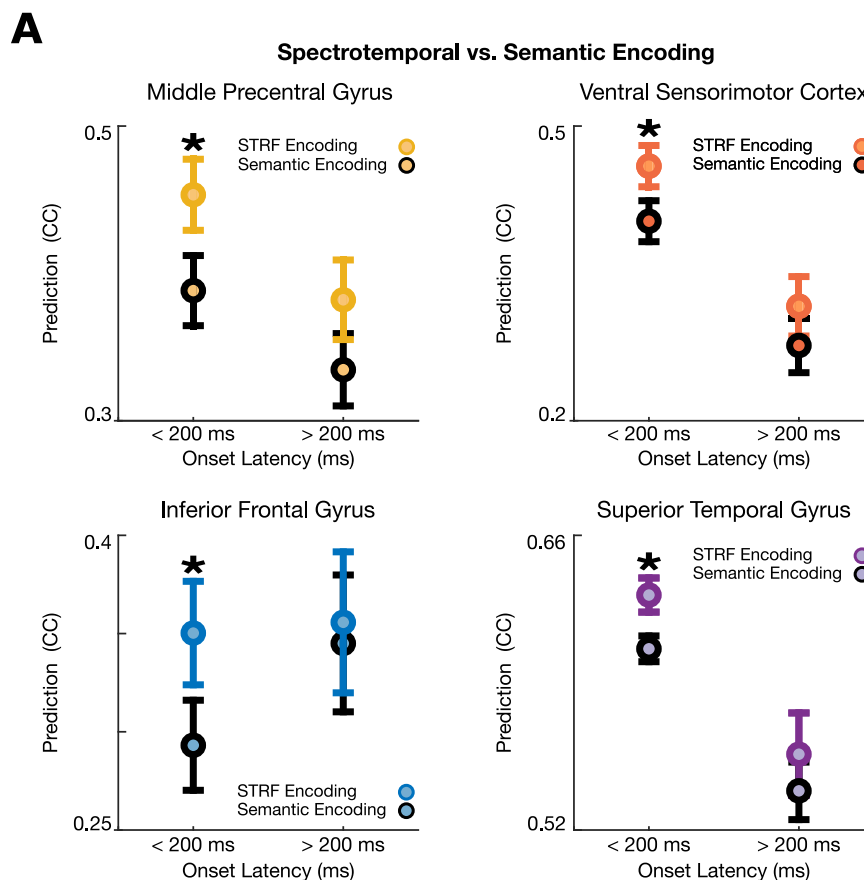

A

Spectrotemporal vs. Semantic Encoding

Middle Precentral Gyrus — *Ventral Sensorimotor Cortex* — *Inferior Frontal Gyrus* — *Superior Temporal Gyrus*

Fig. 4 | Spectrotemporal encoding models are better than semantic encoding models for short-onset latency sites in the frontal lobe and STG. A Short-onset latency sites in each area encode low-level spectrotemporal information. Sites with onset latencies less than 200 ms show significantly higher STRF model predictions than semantic model predictions consistent with dominant spectrotemporal over semantic coding at these short-onset latency sites. This is consistent with short-onset latency sites encoding spectrotemporal speech information in parallel

(p < 0.001, Wilcoxon rank-sum test, two-sided). Longer onset latency sites in the same areas show no difference between spectrotemporal or semantic encoding models. All data are presented as mean values ± SEM. n site CC prediction values for early (<200 ms)/late (>200 ms) onset latencies = 46/31 IFG, 115/68 mPreCG, 110/43 vSMC, 389/72 STG. CC Pearson correlation coefficient; STRF spectrotemporal receptive field.

processing. Using high-density ECoG, we found that a subset of neural populations in the frontal lobe cortex exhibited response latencies that were synchronous with or preceded the earliest response latencies in STG. Spectrotemporal representations in these populations were lar-gely indistinguishable from those found in STG, encoding key spectral and temporal modulation rates for speech. Finally, we identified white matter tracts that connect early auditory regions like MGB and Heschl's gyrus with frontal lobe areas and were associated with resting-state functional connectivity. Together, these results indicate the

existence of long-range pathways between low-level areas (medial geniculate body, primary auditory cortex) and neural populations in the frontal cortex, which work in parallel to the more well-established hierarchical pathway via lateral cortex areas.

Sensory systems, including vision and hearing, are known to have hierarchical structure, with serial feedforward connections from low-level areas to mid-level and then to apical areas within the frontal cortex[1,2,32]. In addition to this hierarchical structure, locally, within nearby levels, there are short-range parallel connections[3,4].
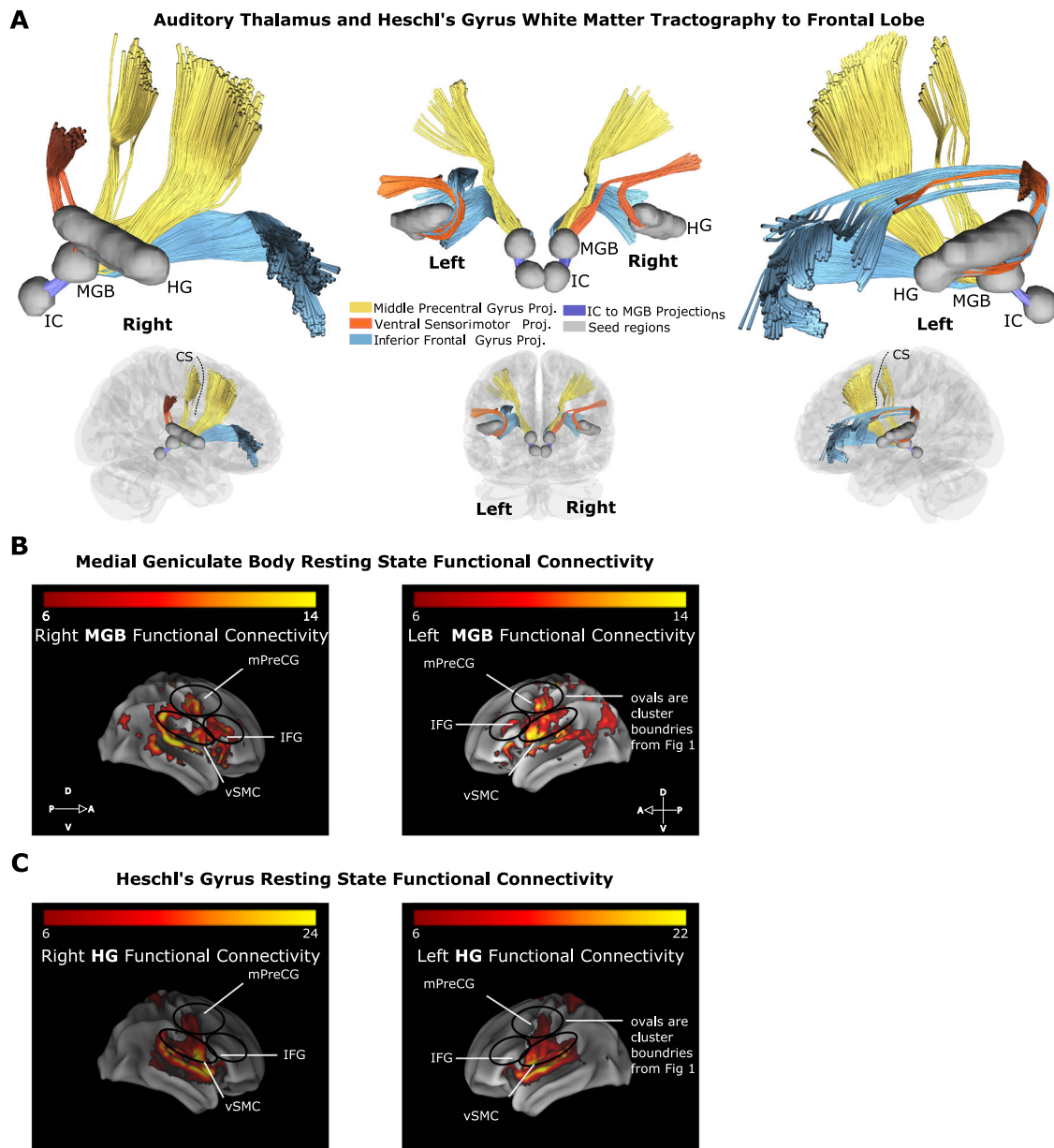
Fig. 5 | White matter tractography and functional connectivity show connections from thalamus and primary auditory cortex to frontal lobe areas. A White matter tractography between the medial geniculate body (MGB), Heschl's gyrus (HG), and suprasylvian cortex (IFG, middle precentral gyrus (mPreCG), and ventral sensorimotor cortex (vSMC)). White matter tracts are color-coded by the suprasylvian site of termination. This analysis demonstrates tractography data consistent with direct MGB-to-frontal cortex and HG-to-frontal cortex white matter tracts. **B** Resting-state functional connectivity of the medial geniculate body and **C** Heschl's gyrus showing functional connectivity to frontal lobe areas, middle precentral gyrus, ventral sensorimotor cortex, and inferior frontal gyrus (connectivity maps derived by performing one-sample t-tests, one-sided, with a statistical threshold of $p < 5 \times 10^{-5}$ after peak-level family wise error (FWE) correction for multiple comparisons). Colorbar represents *T*-score. Overall, these data provide additional structural and functional evidence for parallel pathways to frontal lobe areas from low-level areas, such as Heschl's gyrus and the medial geniculate body. CS central sulcus; FC functional connectivity; HG Heschl's gyrus; IFG inferior frontal gyrus; IC inferior colliculus; MGB medial geniculate body; mPreCG middle precentral gyrus; vSMC ventral sensorimotor cortex.

However, the presence of long-range parallel connections in humans from low-level areas, such as the primary auditory cortex or MGB directly to apical areas in frontal cortex have not been described. The present results demonstrate that, in addition to short-range parallel connections, long-range parallel connections exist from the thalamus and primary auditory cortex to areas in the frontal lobe. To what extent these different proposed parallel pathways exist is an important question for further work and could be examined further in participants with depth electrode coverage or tracer studies in primates.

It is important to emphasize the exploratory nature of the tractography results. While focusing on the robust ECoG findings in combination with DTI data, we can present broad interpretation that respects the complexity of the neural architecture involved and produces a model that can be tested with more robust structural analysis of these pathways. DTI-based methods have inherent limitations, including issues of resolution, crossing fibers, and challenges in differentiating close anatomical structures such as the MGB and adjacent thalamic nuclei[33,34]. Although steps within the scope of the study were taken to mitigate these limitations, these findings require further

validation via post-mortem ultra-high-resolution MRI, virtual lesioning[35], tracer injections in primates, and MGB single pulse stimulation in primates[36]. Additionally, future work in which preoperative high-resolution DTI and fMRI recordings in study participants could show that the individual electrode responses are predicted by single-subject connectivity for additional support of the model.

Perhaps a surprising result was how similar spectrotemporal representations encoded by early onset responses are between the frontal lobe and STG. Based on the model of long-range parallel pathways proposed here, frontal neuronal populations and STG would receive the same low-level spectrotemporal representations in MGB or primary auditory cortex to serve as raw snapshot of the of incoming speech signal. This is analogous to work from our group showing primary auditory cortex in Heschl's gyrus, and STG receive rapid parallel inputs that convey spectrotemporal representations of speech[4]. Rapid snapshots of raw spectrotemporal information to frontal lobe areas may help with top-down modulatory functions carried out by the frontal lobe[37-40] or real-time spectrotemporal feedback during speech production[41,42]. Further comparisons between the information encoded in fast frontal lobe responses and longer latency frontal lobe responses (reflecting inputs from canonical hierarchical pathways) is one future direction to investigate the function of these rapid parallel inputs to frontal lobe areas. A second future direction is specialized orthogonal stimuli designed to test which lower-level speech representation is dominant (spectrotemporal versus phonemic versus phonetic feature representations)[21,23,43]. Lastly, due to experimental time constraints, our standard protocol is focused on presenting speech sounds. It is less common to collect data that could speak to questions of whether these early frontal lobe responses are specific to speech. Future work using simple tones, clicks, or white noise bursts could investigate this important question.

It may be surprising to find such long-range connections, particularly as part of a network involved in processing such a complex stimulus. However, three important frontal lobe characteristics suggest a long-range parallel auditory pathway. First, these frontal areas are highly heterogeneous, and sub-populations of neurons in primates appear to be responsive to the acoustic properties of sound, similar to the tuning properties seen in lower-level areas[7]. Second, prior work in primates has suggested the presence of short-onset latencies in the frontal lobe to sound[5,6], and in humans, electrical stimulation of primary auditory cortex elicits short-latency responses in the inferior frontal gyrus[44]. Lastly, in rhesus macaques, non-primary auditory belt regions have tracer-defined anatomic projections to the frontal lobe[45]. The present work shows short-latency neural populations with acoustic representations that are indistinguishable from STG, consistent with long-range parallel transmission of low-level representations of speech directly to the frontal cortex and is further supported by structural evidence of white matter tracts connecting the thalamus and primary auditory cortex to the frontal lobe.

Regarding the potential function of a parallel long-range pathway, an intriguing possibility is that these frontal lobe auditory populations facilitate implementation of processes that are canonically associated with interactions between frontal areas and lower-level auditory areas[37-39,46]. For example, there is substantial data supporting the role of areas like IFG and ventral precentral gyrus in top-down modulation of lower-level areas during speech perception[37-40]. An important aspect of top-down modulation is the necessity for real-time feedback, and consistent with this, experimental evidence has demonstrated the remarkable speed with which it occurs; for example, there is behavioral and neural evidence for nearly instantaneous perceptual warping in phenomena like phoneme restoration[38,47,48]. Whereas most hierarchical models of speech processing in humans assume that the frontal lobe receives inputs from the dorsal and ventral pathways, near instantaneous frontal cortex-mediated perceptual restoration may be facilitated by long-range parallel input to frontal areas that provide a snapshot of the raw incoming sensory data for real-time top-down modulation.

Similarly, a parallel auditory pathway to neural populations throughout sensorimotor speech cortex may play a role in real-time acoustic feedback for articulatory control[41,42]. Neighboring—or possibly even overlapping—neural populations that directly control articulation[49-53] show sensory responses during passive listening[53,54], which could be used in both speech-motor planning and for real-time feedback from the acoustics generated by articulatory output. In either potential role (top-down modulation or real-time acoustic feedback), this long-range parallel pathway gives frontal lobe cortical areas real-time access to the primary spectrotemporal input signal in addition to higher-level representations the frontal lobe receives from other areas, such as the lateral temporal cortex.

We do not disregard the importance of canonical processing pathways, or the fundamental hierarchical nature of the brain[1,2,55]. Indeed, for speech, it is well-established that structures like the arcuate fasciculus support a major pathway between auditory, temporal and frontal cognitive and motor regions[56-58]. Instead, we propose that parallel inputs provide additional computational resources that support the rapid processing required for sounds like speech. Further work is necessary to understand how these pathways work together and whether they have distinct targets within the frontal cortex (perhaps suggested by the distributed and relatively sparse nature of short-latency responses in the present results). However, taken together, the existence of a direct, long-range parallel auditory pathway to the frontal cortex supports the notion that the degree to which neural systems underlying speech perception are largely hierarchical needs to be examined more closely.

Methods

Participants and neural recordings

ECoG arrays (interelectrode distance = 4 mm) were placed subdurally in 17 patient volunteers (9 right hemisphere, 8 left hemisphere) undergoing a neurosurgical procedure for the treatment of medication-refractory epilepsy. All participants were native English speakers; all were fluent in English. All participants had normal hearing and no communication deficits. All experimental protocols were approved by the University of California, San Francisco, Institutional Review Board and Committee on Human Research. Each participant gave written informed consent before participating in the study. The location of array placement was determined by clinical criteria alone. Participants were asked to passively listen to 10–40 min of natural speech while ECoG signals were recorded simultaneously. Signals were amplified and sampled at 3052 Hz. After the rejection of electrodes with excessive noise or artifacts, signals were referenced to a common average, and the high-gamma band (70–150 Hz) was extracted as the analytic amplitude of the Hilbert transform[9]. Signals were subsequently downsampled to 100 Hz. For onset latency analysis the high-gamma band was also extracted using Morlet wavelet decomposition and downsampled to 1000 Hz. The resulting signal for each electrode was z-scored based on the mean and standard deviation of activity during the entire block.

Stimuli

Speech stimuli were delivered binaurally through free-field speakers at approximately 70 dB average sound pressure level. The frequency power spectrum of stimuli spanned 0–8000 Hz. The stimulus set consisted of prerecorded (2–4 s) sentences from the phonetically transcribed TIMIT speech corpus with one-second silent intervals between each sentence presentation[8]. To quantify response characteristics to individual sentences for responsivity and onset latency analysis we analyzed responses to ten unique sentences (each unique sentence repeated ten times). Each participant was also presented an additional 115–489 unique, non-repeated sentences for receptive field

analysis. The total speech corpus included 286 male and 116 female speakers, with 1–3 sentences spoken per speaker, and unique lexical content for each sentence.

Responsivity and spatial analysis

Ten TIMIT sentences were presented randomly ten times in each participant (100 sentence presentations total). Neural data was aligned by the onset of sound for each sentence. The mean evoked high-gamma to each of the ten sentences was then computed. To test for speech-evoked responses, for each time bin of the evoked high-gamma after sound onset, a Wilcoxon rank-sum test was performed to test for a significant difference from baseline ($p < 0.05$, Bonferroni corrected for the number of sample bins in the sentence). To visualize electrode coordinates in MNI space, we performed nonlinear surface registration using a spherical sulcal-based alignment in Freesurfer, aligning to the cvs_avg35_inMNI152 template[59]. This nonlinear alignment ensures that electrodes on a gyrus in the participant's native space remain on the same gyrus in the atlas space, but does not maintain the geometry of the grid. For spatial analysis, significant responses were projected to the left hemisphere in all participants, given that there were no significant differences in the proportion of responsive sites in each area or amplitude of responses between the left and right hemispheres in the suprasylvian cortex. Spatial clustering was performed using a mixture of Gaussians model with the number of clusters (cluster number = 3) identified by both Bayesian information criterion and Silhouette criterion.

Analysis

Onset latency analysis. To compute the onset of local neuronal spiking, each trial of the repeated TIMIT sentences was aligned by the onset of sound, and the mean evoked high-gamma activity for each sentence was computed. Like prior latency analysis using ECoG in human participants, the high-gamma band, as opposed to the full event-related potential (ERP), was used[60]. High-gamma is used because of its selectivity for neuronal spiking[10–12,61,62]. Additionally, high frequencies that compose the high-gamma band attenuate quickly as a function of distance making them selective for local neuronal activity under each electrode rather than pooling over large areas of cortex that span multiple gyri[63,64]. ERP waveforms from subdural grids also have extremely high waveform shape variability due to the wide pooling of local and volume conducted potentials, making them less tractable for onset latency measurements with a single standardized metric. To compute onset latency, the 500 ms before sentence onset served as the baseline for comparisons. For each 1-ms bin of the evoked high-gamma after sound onset, a Wilcoxon rank-sum test was performed to test for a significant difference from baseline ($p < 0.05$). Like prior ECoG response latency work in the human auditory cortex or primate auditory cortex, response latency was defined as the time in which the mean evoked high-gamma was significantly different from baseline and remained significant for 15 ms[19,20]. The shortest sentence onset latency for each electrode was defined as the speech onset latency for that site.

Encoding analysis. STRF and semantic encoding models were fit using normalized reverse correlation[65] with open-source code available at: http://strfpak.berkeley.edu/. Regularization was controlled by fitting a tolerance hyperparameter via cross-validation[66]. STRFs were computed on an estimation set (90% of the total data) and cross-validated on a test set, which was withheld from the estimation process (10% of the data). For STRF fitting, spectrogram representations of speech stimuli were generated using a cochlear model of auditory processing[67]. For the semantic encoding model, word vector representations of the TIMIT speech corpus were derived from the FASTTEXT data set[30] by mapping each 10 ms segment of the speech signal to the corresponding word vector representation for that word.

Modulation tuning. To characterize modulation tuning, the modulation transfer function (MTF) for each site was computed by taking the magnitude of the two-dimensional Fourier transform ($\mathscr{F}_2\{\bullet\}$) of each STRF:

$$\mathrm{MTF}(\omega_t, \omega_s) = \left| \mathscr{F}_2\{\mathrm{STRF}(t,f)\} \right| \tag{1}$$

Where (t, f) are time and frequency and (ω_t, ω_s) are temporal and spectral modulation, respectively[23].

Tractography analysis. For white matter tractography, we applied a deterministic diffusion fiber tractography algorithm[68] using the spin distribution function (SDF) template created by Yeh et al.[69] and publicly available diffusion data (http://brain.labsolver.org) with high angular and high spatial resolution from 842 individuals from the Human Connectome Project[31]. Our tractography analysis used deterministic diffusion tensor imaging (DTI) methods applied to high-resolution data from the Human Connectome Project. Seed regions were placed in the medial geniculate body (MGB) and Heschl's gyrus, with a target region of interest that spanned the lateral frontal and parietal lobe. To mitigate the impact of crossing fibers, we utilized constrained spherical deconvolution (CSD) methods[68], which improve the differentiation of multiple fiber populations within a voxel. Tractography seed and target regions, with exception of the medial geniculate body, were specified using the Human Connectome Project Multi-Modal Parcellation version 1.0 anatomical atlas and the FreeSurfer Destrieux anatomical atlas[70–72]. The medial geniculate body seed region coordinates were determined from previously reported MNI coordinates[73–77] and validated by demonstration of positive tractography from the inferior colliculus to the thalamus. The inferior colliculus is the auditory input nucleus to the medial geniculate body so this is a useful approach to increase confidence of MGB localization and a precise seed region. These inferior colliculus-to-MGB DTI tracts are shown in Fig. 5A (purple tracts). The MGB seed region was further validated by fMRI results (Fig. 5B) that are consistent with what we would expect with an accurate MGB seed region: functional connectivity primarily to planum temporale, STG, and the frontal lobe areas of interest.

Functional connectivity analysis. Brain functional images were acquired in a cohort of 50 neurologically intact participants at the Siemens 3-Tesla Prisma scanner located at UCSF. We collected 560 T2*-weighted EPI volumes for each individual with the following parameters: TR/TE = 850/32.8 ms, flip angle = 45°, voxel size = 2.2 × 2.2 × 2.2 mm³, field-of-view = 21 × 211 mm², multi-band accelerating factor = 6. Image preprocessing consists of slice-time correction, realignment to the mean functional image, assessment for rotational and translational head motion, and correction for susceptibility-induced distortions. Each participant was assessed for cognitive and language intactness and participants with excessive motor movements during the scan were excluded. We ensure alertness during the scan through continuous real-time monitoring and post-scan participant feedback. Functional images are then normalized to the EPI template in the MNI space with a combination of rigid, affine, and nonlinear warping. After smoothing the images with a 5 mm full width at half maximum (FWHM) Gaussian kernel, CSF and white matter tissue probability maps were then used to compute the mean time series used as regressors. Functional data were then bandpass filtered (0.008 Hz $<f<0.15$ Hz), and the nuisance variables were regressed out from the data, which included the six motion parameters, the first derivative and quadratic terms, as well as CSF and white matter time series. Seed ROIs were located bilaterally in the MBG (MNI coordinates: left x = −9, y = −23, z = −1; right x = 8, y = −23, z = −1) and in Heschel's gyrus. Single-subject correlation maps were generated by calculating the r−Pearson correlation coefficient between the average BOLD signal time course from the seed ROIs and the time course from all other voxels of the brain. Finally, correlation maps were converted to

z-scores, and group-level connectivity maps were calculated for each seed with a statistical threshold at $p < 5 \times 10^{-5}$ after peak-level family-wise error (FWE) correction for multiple comparisons.

Reporting summary

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

Data availability

The human patient data relevant to this study are accessible under restricted access according to our IRB protocol. The de-identified patient data for which patients have consented to public release will be made available from the corresponding author upon request. Source data are provided with this paper. The Dryad link to the Source data is specified in the GitHub source code README file.

Code availability

The analysis and data visualization code will be made available upon request. Source code are provided with this paper: https://github.com/ChangLabUcsf/Hullett_2025_Source_Code

References

1. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
2. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
3. Rauschecker, J. P., Tian, B., Pons, T., Mishkin, M. & Carolina, N. Serial and parallel processing in rhesus monkey auditory cortex. *J. Comp. Neurol.* **103**, 89–103 (1997).
4. Hamilton, L. S., Oganian, Y., Hall, J. & Chang, E. F. Parallel and distributed encoding of speech across human auditory cortex. *Cell* **184**, 4626–4693 (2021).
5. Marrouch, N., Slawinska, J., Giannakis, D. & Read, H. L. Data-driven Koopman operator approach for computational neuroscience. *Ann. Math. Artif. Intell.* **88**, 1155–1173 (2020).
6. Romanski, L. M. & Hwang, J. Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience* **214**, 36–48 (2012).
7. Romanski, L. M., Averbeck, B. B. & Diltz, M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* **93**, 734–747 (2004).
8. Garofolo, J. S. et al. TIMIT acoustic-phonetic continuous speech corpus. (Linguistic Data Consortium, Philadelphia, 1993).
9. Crone, N. E., Boatman, D., Gordon, B. & Hao, L. Induced electrocorticographic gamma activity during auditory perception. *Clin. Neurophysiol.* **112**, 565–582 (2001).
10. Ray, S. & Maunsell, J. H. R. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
11. Steinschneider, M., Fishman, Y. I. & Arezzo, J. C. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb. Cortex* **18**, 610–625 (2008).
12. Leonard, M. K. et al. Large-scale single-neuron speech sound encoding across the depth of human cortex. *Nature* **626**, 593–602 (2023).
13. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, Inc., 1990).
14. Yi, H. G., Leonard, M. K. & Chang, E. F. The encoding of speech sounds in the superior temporal gyrus. *Neuron* **102**, 1096–1110 (2019).
15. Bhaya-grossman, I. & Chang, E. F. Speech computations of the human superior temporal gyrus. *Annu. Rev. Psychol.* **73**, 79–102 (2022).
16. Turkeltaub, P. E. & Coslett, H. B. Localization of sublexical speech perception components. *Brain Lang.* **114**, 1–15 (2010).
17. Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816–847 (2012).
18. Chinedu-Eneh, E. O. et al. Influences of electrode density on intracranial seizure localisation: a single-blinded randomised crossover study. *eBioMedicine* **113**, 105606 (2025).
19. Camalier, C. R., D'Angelo, W. R., Sterbing-D'Angelo, S. J., De La Mothe, L. A. & Hackett, T. A. Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proc. Natl. Acad. Sci. USA* **109**, 18168–18173 (2012).
20. Nourski, K. V. et al. Functional organization of human auditory cortex: investigation of response latencies through direct recordings. *Neuroimage* **101**, 598–609 (2014).
21. Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–1010 (2014).
22. Polley, D. B., Read, H. L., Storace, D. A. & Merzenich, M. M. Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J. Neurophysiol.* **97**, 3621–3638 (2007).
23. Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E. & Chang, E. F. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *J. Neurosci.* **36**, 2014–2026 (2016).
24. Linden, J. F., Liu, R. C., Sahani, M., Schreiner, C. E. & Merzenich, M. M. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J. Neurophysiol.* **90**, 2660–2675 (2003).
25. Nagel, K. I. & Doupe, A. J. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* **51**, 845–859 (2006).
26. Schaeffer, D. J. et al. Divergence of rodent and primate medial frontal cortex functional connectivity. *Proc. Natl. Acad. Sci. USA* **117**, 21681–21689 (2020).
27. Miller, L. M., Escabí, M. A., Read, H. L. & Schreiner, C. E. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J. Neurophysiol.* **87**, 516–527 (2002).
28. Winer, J. A., Miller, L. M., Lee, C. C. & Schreiner, C. E. Auditory thalamocortical transformation: structure and function. *Trends Neurosci.* **28**, 255–263 (2005).
29. Klein, D. J., Simon, J. Z., Depireux, D. A. & Shamma, S. A. Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J. Comput. Neurosci.* **20**, 111–136 (2006).
30. Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C. & Joulin, A. Advances in pre-training distributed word representations. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (eds Calzolari, N. et al.) 937–944 (European Language Resources Association, 2018).
31. Van Essen, D. C. et al. The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013).
32. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
33. Thomas, C. et al. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proc. Natl. Acad. Sci. USA* **111**, 16574–16579 (2014).
34. Sitek, K. R. et al. Mapping the human subcortical auditory system using histology, postmortem MRI and in vivo MRI at 7T. *Elife* **8**, 1–36 (2019).
35. Pestilli, F., Yeatman, J. D., Rokem, A., Kay, K. N. & Wandell, B. A. Evaluation and statistical inference for human connectomes. *Nat. Methods* **11**, 1058–1063 (2014).
36. Keller, C. J. et al. Mapping human brain networks with cortico-cortical evoked potentials. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130528 (2014).

37. Cope, T. E. et al. Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nat. Commun.* **8**, 2154 (2017).

38. Leonard, M. K., Baud, M. O., Sjerps, M. J. & Chang, E. F. Perceptual restoration of masked speech in human cortex. *Nat. Commun.* **7**, 1–9 (2016).

39. Sohoglu, E., Peelle, J. E., Carlyon, R. P. & Davis, M. H. Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* **32**, 8443–8453 (2012).

40. Di Liberto, G. M., Lalor, E. C. & Millman, R. E. Causal cortical dynamics of a predictive enhancement of speech intelligibility. *Neuroimage* **166**, 247–258 (2018).

41. Houde, J. F. & Chang, E. F. The cortical computations underlying feedback control in vocal production. *Curr. Opin. Neurobiol.* **33**, 174–181 (2015).

42. Guenther, F. H. & Hickok, G. Neural models of motor speech control. in *Neurobiology of Language* (Elsevier Inc., 2016).

43. de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L. & Theunissen, F. E. The hierarchical cortical organization of human speech processing. *J. Neurosci.* **37**, 6539–6557 (2017).

44. Rocchi, F. et al. Common fronto-temporal effective connectivity in humans and monkeys ll common fronto-temporal effective connectivity in humans and monkeys. *Neuron* **109**, 852–868.e8 (2021).

45. Romanski, L. M., Tian, B., Fritz, J., Mishkin, M. & Rauschecker, J. P. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* **2**, 1131–1136 (1999).

46. Dick, X. F. K. et al. Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally directed attention and systematically related to cortical myeloarchitecture. *J. Neurosci.* **37**, 12187–12201 (2017).

47. Warren, R. M. Perceptual restoration of missing speech sounds. *Science* **167**, 392–393 (1970).

48. Samuel, A. G. Phonemic restoration: insights from a new methodology. *J. Exp. Psychol.* **110**, 474–494 (1981).

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

50. Schulz, G. M., Varga, M., Jeffires, K. & Ludlow, C. L. Functional neuroanatomy of human vocalization: an H215O PET study. *Cereb. Cortex* **15**, 1835–1847 (2005).

51. Guenther, F. H., Ghosh, S. S. & Tourville, J. A. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* **96**, 280–301 (2006).

52. Brown, S. et al. The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain Cogn.* **70**, 31–41 (2009).

53. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–331 (2013).

54. Wilson, S. M. & Iacoboni, M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* **33**, 316–325 (2006).

55. Plakke, B. & Romanski, L. M. Auditory connections and functions of prefrontal cortex. *Front. Neurosci.* **8**, 1–13 (2014).

56. Fernández-Miranda, J. C. et al. Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Struct. Funct.* **220**, 1665–1680 (2015).

57. Geschwind, N. The organization of language and the brain. *Science* **170**, 940–944 (1970).

58. Dick, A. S. & Tremblay, P. Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. *Brain* **135**, 3529–3550 (2012).

59. Hamilton, L. S., Chang, D. L., Lee, M. B. & Chang, E. F. Semi-automated anatomical labeling and inter-subject warping of high-density intracranial recording electrodes in electrocorticography. *Front. Neuroinform.* **11**, 1–19 (2017).

60. Nourski, K. V. et al. Processing of auditory novelty across the cortical hierarchy: an intracranial electrophysiology study. *Neuroimage* **183**, 412–424 (2018).

61. Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* **29**, 13613–13620 (2009).

62. Crone, N. E., Miglioretti, D. L., Gordon, B. & Lesser, R. P. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* **121**, 2301–2315 (1998).

63. Kajikawa, Y. & Schroeder, C. E. How local is the local field potential? *Neuron* **72**, 847–858 (2011).

64. Chang, E. F. Towards large-scale, human-based, mesoscopic neurotechnologies. *Neuron* **86**, 68–78 (2015).

65. Theunissen, F. E. et al. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Netw. Comput. Neural Syst.* **12**, 289–316 (2001).

66. David, S. V., Mesgarani, N. & Shamma, S. A. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* **18**, 191–212 (2007).

67. Yang, X., Wang, K. & Shamma, S. A. Auditory representations of acoustic signals. *IEEE Trans. Inf. Theory* **38**, 824–839 (1992).

68. Yeh, F., Verstynen, T. D., Wang, Y., Fernández-miranda, J. C. & Isaac, W. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS ONE* **8**, 1–16 (2013).

69. Yeh, F. C. et al. Population-averaged atlas of the macroscale human structural connectome and its network topology. *NeuroImage* **178**, 57–68 (2018).

70. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

71. Fischl, B. et al. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**, 11–22 (2004).

72. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. NeuroImage automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).

73. Devlin, J. T. et al. Reliable identification of the auditory thalamus using multi-modal structural analyses. *Neuroimage* **30**, 1112–1120 (2006).

74. Mihai, P. G. & Tschentscher, N. Modulation of the primary auditory thalamus when recognizing speech with background noise. *J. Neurosci.* **41**, 7136–7147 (2021).

75. Mu, M. et al. Structural brain changes in tinnitus. *Cereb. Cortex* **16**, 1283–1288 (2006).

76. Díaz, B., Hintz, F., Kiebel, S. J. & Von Kriegstein, K. Dysfunction of the auditory thalamus in developmental dyslexia. *Proc. Natl. Acad. Sci. USA* **109**, 13841–13846 (2012).

77. Glad Mihai, P. et al. Modulation of tonotopic ventral medial geniculate body is behaviorally relevant for speech recognition. *Elife* **8**, 1–28 (2019).

Acknowledgements

Author contributions

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/s41467-025-67517-7.

Correspondence and requests for materials should be addressed to Edward F. Chang.

Peer review information *Nature Communications* thanks Heather Read and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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