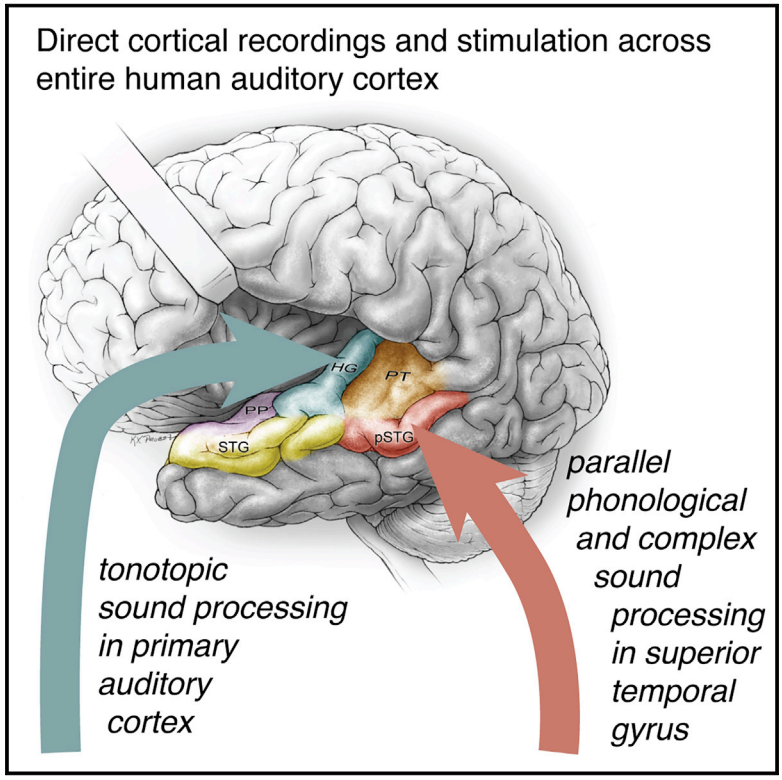
Cell

Parallel and distributed encoding of speech across human auditory cortex

Graphical abstract



Authors

Liberty S. Hamilton, Yulia Oganian, Jeffery Hall, Edward F. Chang

Correspondence

edward.chang@ucsf.edu

In brief

In the human auditory cortex, the encoding of acoustic and phonetic information of speech is distributed across areas, and the information processing is parallel.

Highlights

- We recorded intracranial signals in human primary and nonprimary auditory cortex
- A superior temporal gyrus onset zone activates parallel to primary auditory areas
- Stimulation of superior temporal gyrus impairs speech perception
- Stimulation of primary auditory cortex does not affect speech perception

Hamilton et al., 2021, Cell *184*, 1–14
September 2, 2021
https://doi.org/10.1016/j.cell.2021.07.019

CellPress

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

Cell

CellPress

Article

Parallel and distributed encoding of speech across human auditory cortex

Liberty S. Hamilton,[1,3] Yulia Oganian,[1] Jeffery Hall,[2] and Edward F. Chang[1,4,*]
[1]Department of Neurological Surgery, University of California, San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94158, USA
[2]Department of Neurology and Neurosurgery, McGill University Montreal Neurological Institute, Montreal, QC, H3A 2B4, Canada
[3]Present address: Department of Speech, Language and Hearing Sciences, Department of Neurology, The University of Texas at Austin, Austin, TX 78712, USA
[4]Lead contact
*Correspondence: edward.chang@ucsf.edu
https://doi.org/10.1016/j.cell.2021.07.019

SUMMARY

Speech perception is thought to rely on a cortical feedforward serial transformation of acoustic into linguistic representations. Using intracranial recordings across the entire human auditory cortex, electrocortical stimulation, and surgical ablation, we show that cortical processing across areas is not consistent with a serial hierarchical organization. Instead, response latency and receptive field analyses demonstrate parallel and distinct information processing in the primary and nonprimary auditory cortices. This functional dissociation was also observed where stimulation of the primary auditory cortex evokes auditory hallucination but does not distort or interfere with speech perception. Opposite effects were observed during stimulation of nonprimary cortex in superior temporal gyrus. Ablation of the primary auditory cortex does not affect speech perception. These results establish a distributed functional organization of parallel information processing throughout the human auditory cortex and demonstrate an essential independent role for nonprimary auditory cortex in speech processing.

INTRODUCTION

Speech contains myriad acoustic cues that give rise to the rich experience of spoken language. Listening to speech activates populations of neurons in multiple functional regions of the auditory cortex, including primary and nonprimary cortical fields. The computations carried out in these areas allow for extraction of meaningful linguistic content, such as consonants and vowels, or prosodic rhythm and intonation from the spectrotemporal acoustics of speech.

Many previous functional neuroimaging studies (e.g., functional magnetic resonance imaging [fMRI] and magnetoencephalography [MEG]) suggest a progressive selectivity for more complex sound types from primary toward nonprimary, higher-order areas (Brodbeck et al., 2018; Chevillet et al., 2011; de Heer et al., 2017; Rauschecker and Scott, 2009; Wessinger et al., 2001). For example, stronger responses are found in high-order auditory areas to complex, natural stimuli, such as speech and music, than simple artificial stimuli, such as pure tones (Leaver and Rauschecker, 2010; Norman-Haignere et al., 2015; Schönwiesner and Zatorre, 2009). These results are typically interpreted in the framework of an anatomical-hierarchical model, where simple acoustic features are transformed into more complex and speech-specific representations across the auditory cortex. However, the limited spatial or temporal resolution of noninvasive imaging precludes mapping of locally heterogeneous neural response types, leaving fundamental questions about speech sound representations unanswered.

Understanding sound representations in the brain has been advanced by the high spatiotemporal resolution of intracranial recordings (Berezutskaya et al., 2017; Chang, 2015; Flinker et al., 2011; Howard et al., 2000; Lachaux et al., 2012; Nourski et al., 2012; Ozker et al., 2017). The human auditory cortex lies across an expansive surface of the temporal lobe, including the superior temporal plane and the adjacent superior temporal gyrus (STG; Figure 1) on its lateral aspect. STG is exposed on the lateral temporal lobe and thus accessible by surface electrocorticography (ECoG) recording methods. In contrast, most of the temporal plane, including the tonotopic "auditory core" on the posteromedial portion of Heschl's gyrus (HG), is buried deep within the lateral fissure, which separates the temporal lobe from the frontal and parietal lobes (Brewer and Barton, 2016). These anatomical constraints make the temporal plane difficult to access for neurophysiological recordings. Therefore, functional parcellations beyond the core and their relation to anatomical landmarks are not well understood. In the current study, the high temporal resolution of ECoG allows us to probe the extent to which a serial hierarchy from primary auditory cortex on HG to the STG is consistent with data from speech and tone listening. The high spatial resolution similarly provides a more complete picture of how different functional and anatomical regions in the human auditory cortex interact.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

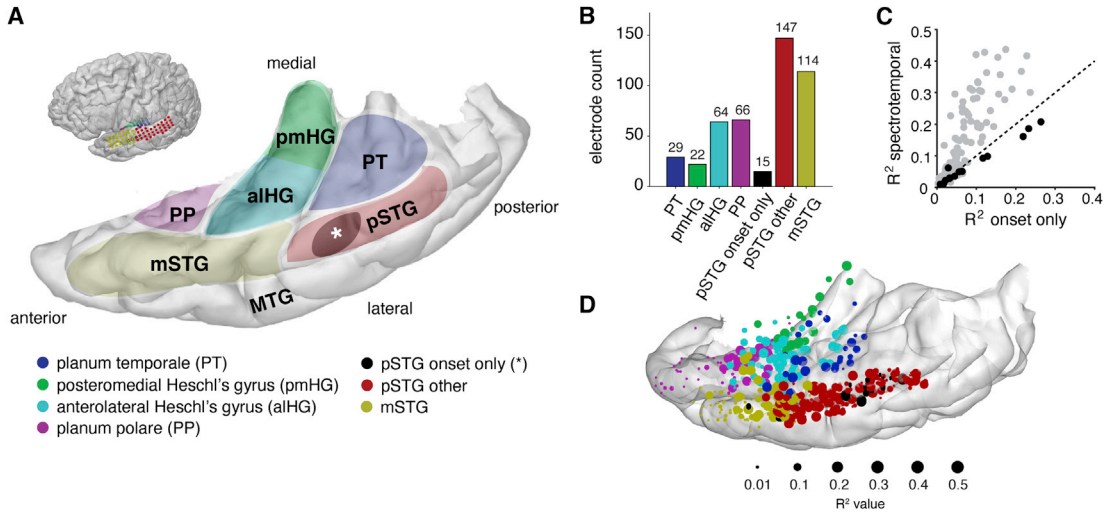


Figure 1. Anatomical parcellations of temporal lobe regions of the human auditory cortex and electrode coverage

(A) Anatomical regions of interest on the left hemisphere temporal lobe of an example participant. STG = superior temporal gyrus, MTG = middle temporal gyrus.

(B) Electrode counts across anatomical areas for all nine participants.

(C) Comparison between onset-only and spectrotemporal models shows a population described by only a singular onset feature.

(D) All participants' electrodes projected onto an Montreal Neurological Institute (MNI) atlas brain (cvs_avg35_inMNI152). Electrode size reflects the maximum amount of variance (R^2) explained by the encoding models tested in our analyses. Electrode sites are colored according to their anatomical location.

See also Figures S1 and S6 and Table S2.

While recent advances in human intracranial recordings have revealed specialization for different speech sound features in lateral STG, whether similar feature representations exist on the superior temporal surface, including core auditory cortex, and whether STG responses to non-speech reflect the same computational principles remain open questions. Such an analysis requires sampling of neural responses to natural speech and experimentally controlled simple sound stimuli across all cortical auditory areas simultaneously. Critically, a comprehensive map of feature encoding across primary and higher-order auditory areas is prerequisite to a meaningful evaluation of models of information flow and transformations of cortical representations.

Here, we simultaneously recorded neural activity from multiple subfields of the human temporal lobe auditory cortex using high-density electrode grids. Neurophysiological recordings allowed us to determine the flow of information processing and how cues in the speech signal are mapped across the auditory cortex. Instead of a simple serial hierarchy, we found evidence for distributed and parallel processing, where early latency responses were observed throughout the posterior temporal plane and STG. Furthermore, direct focal electrocortical stimulation (ECS) and an ablation case study provide evidence that HG is neither necessary nor sufficient for speech perception.

RESULTS

The anatomical divisions of the human auditory cortex include the planum temporale (PT), HG (or transverse temporal gyrus), and planum polare (PP) on the superior temporal plane (Hickok and Saberi, 2012; Moerel et al., 2014) and posterior STG (pSTG) and middle STG (mSTG) on its lateral surface (Figure 1A). Here, we define pSTG as the portion of the STG posterior to the lateral exit point of the transverse temporal sulcus (Friederici, 2015; Upadhyay et al., 2008). Recordings from temporal plane typically use penetrating depth electrodes, which may capture the long axis of HG, but do not have the uniform and widespread cortical surface coverage offered by electrode grids (Brugge et al., 2003; Griffiths et al., 2010; Nourski et al., 2014; Steinschneider et al., 2014). Few studies also record from the PT and PP, so these areas remain relatively understudied (Besle et al., 2008; Griffiths and Warren, 2002; Bidet-Caulet et al., 2007; Liégeois-Chauvel et al., 1999). Placing grid electrodes in this area requires meticulous surgical dissection of the Sylvian fissure and is only possible in cases of opercular/insular surgery (Bouthillier et al., 2012; Malak et al., 2009).

We acquired intracranial recordings from 636 electrode sites in the left temporal plane and STG in nine participants, including both grid and depth electrodes (Figures 1A, 1B, and S1), as participants listened to speech and pure tone stimuli. As in previous work (Hamilton et al., 2018), we identified an onset-specific region using non-negative matrix factorization (NMF; black electrodes in Figures 1C and 1D; n = 15 electrodes, 13 of which were located in anatomically defined pSTG area; see STAR Methods). This onset-selective region, which exhibited strong transient responses at the onset of sentences followed by relative quiescence, was observed only in a localized region of the lateral STG and not on the superior temporal plane.

Topography of response latencies for information processing

The classical hierarchical model of speech perception assumes that sound information is first received in the primary auditory cortex on HG and then transformed into more complex

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

CellPress

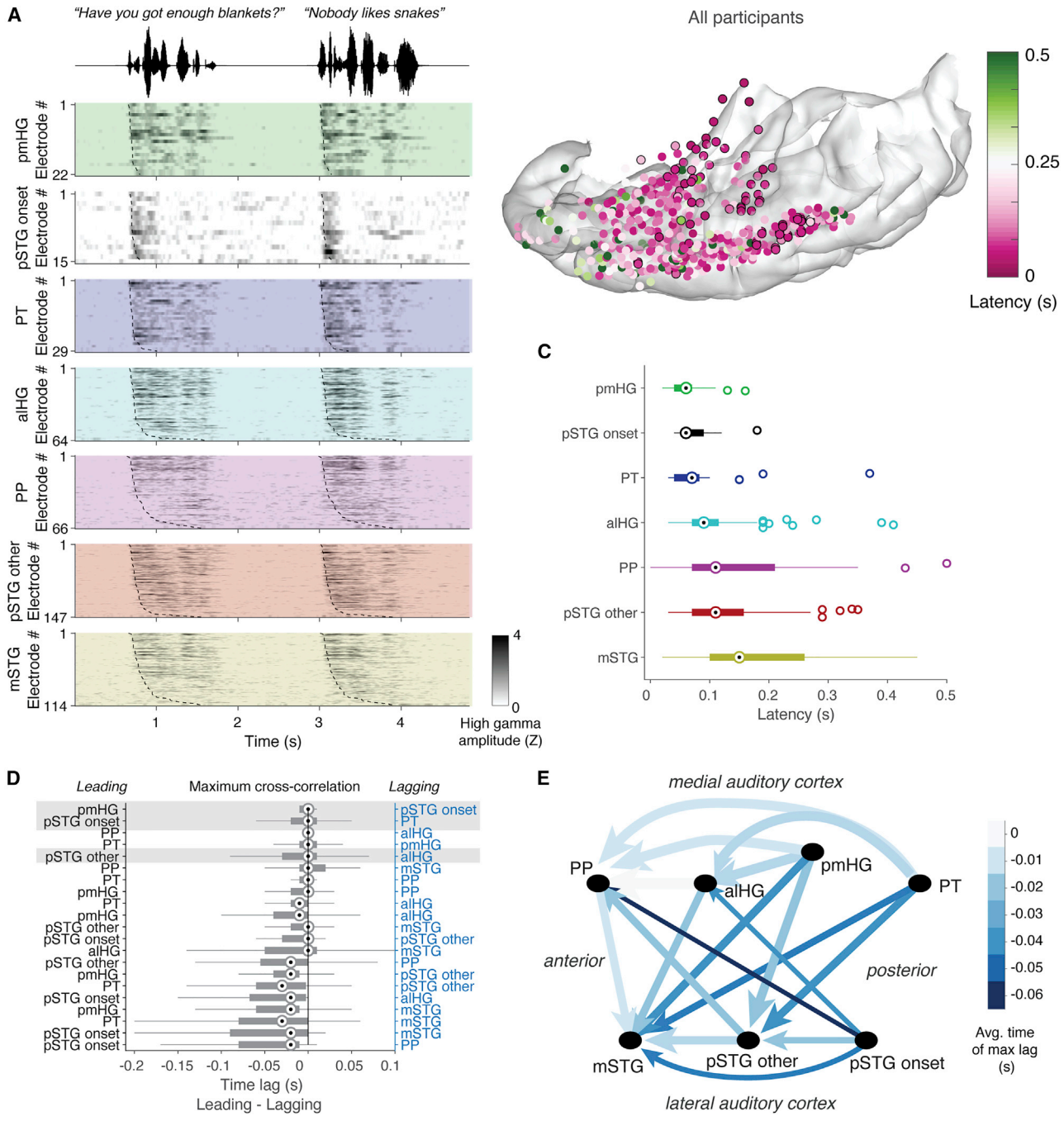


Figure 2. The onset of fast-latency responses in the pSTG is indistinguishable from the onset of responses in primary auditory areas

(A) *Z*-scored high-gamma-amplitude (HGA) responses during speech listening for two example sentences, split by single electrodes in each region of interest (ROI) and ordered by average latency. Response latencies are marked as dashed lines and were measured as the maximum derivative of the high-gamma response.

(B) The high-gamma-derived latencies for each electrode across all participants on an atlas brain. PT, pmHG, and pSTG onset electrodes are outlined in black.

(C) Comparison of onset latencies across brain regions. Only latencies <0.5 s are shown. pmHG, pSTG onset, and PT electrodes showed fast onset times that were statistically indistinguishable. Boxplots in (C) and (D) show the median, interquartile range (box), and minimum of maximum of the data (whiskers), as well as outlier values (open circles).

(D) Cross-correlation analysis between pairs of regions of interest (ROIs), ordered by mean time lag of maximum cross-correlation. Time lags to the left of 0 s indicate that the left ROI precedes the right ROI (e.g., pSTG onset precedes aIHG). Gray shading indicates lead-lag relationships that are not significantly different from zero, indicating simultaneous activation.

(legend continued on next page)

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

representations via corticocortical connections with the lateral STG. To test this assumption, we assessed the relative timing of responses to speech across the auditory cortex. Taking advantage of simultaneous recordings in multiple participants, we performed a latency analysis on trial-averaged high-gamma data within each region of interest (ROI) and then used a cross-correlation analysis to determine lead-lag relationships. We saw fast-latency responses to sentences in posteromedial HG (pmHG), pSTG onset, and PT electrodes, and longer latency responses in anterolateral HG (alHG), pSTG non-onset, mSTG, and PP electrodes (Figures 2A and 2B). Average high-gamma responses within each ROI showed short, primary-like latencies in pSTG onset, PT, and pmHG, as compared to slower responses anterolaterally both on the temporal plane (alHG, PP) and lateral STG (pSTG non-onset, mSTG) (Figure 2C, main effect of ROI: $p = 4.4 \times 10^{-16}$, degrees of freedom [df] = 6). Latencies did not significantly differ between pSTG onset, pmHG, and PT sites ($p > 0.05$, post-hoc Wilcoxon rank sum tests, Figure 2C), but were significantly different between all other areas ($p < 0.05$, Wilcoxon rank sum tests).

When comparing the lead-lag relationships between simultaneously recorded sites in each participant, we found evidence of strong coactivation of pmHG and pSTG onset electrodes, with a maximum cross-correlation at 0 lag for these areas (Figure 2D). We saw parallel lag times between PT and pSTG onset, pmHG and pSTG onset, and alHG and pSTG other. The fast responses in PT, pmHG, and pSTG onset preceded non-onset regions of the pSTG and mSTG as well as PP (Figures 2D and 2E), supporting a posterior to anterior latency gradient. The latency and lag analysis suggest that the posterior STG and posterior temporal plane may have separate, parallel inputs. While this latency analysis places constraints on information flow, it is strictly correlational, so it cannot definitively prove transfer of information from one area to another. Thus, we address causality in subsequent stimulation experiments and in a related ablation case study.

Dissociation of tone and speech encoding in medial and lateral auditory cortex

Part of the argument for serial processing hierarchies in the auditory system is that tuning for simple features such as tones may then be combined at higher stages of the pathway to generate more complex representations (e.g., phonetic) (Okada et al., 2010). To probe this, we used both simple pure tones as well as natural speech sentences as stimuli. We wanted to determine which areas were more responsive to speech and whether responses to tones could predict responses to the spectral cues in speech. Pure tones are typically used to identify tonotopic gradients associated with auditory "core" regions, as they robustly activate frequency-specific regions across the auditory pathway (Brewer and Barton, 2016; Moerel et al., 2012; Wessinger et al., 1997, 2001). Combining tone and speech stimuli within the same patient group allowed us to compare the representations of simple and complex sound stimuli throughout the auditory cortical hierarchy (see Figure 3A for an example of electrode coverage). Figures 3C and 3D shows example tone- and speech-derived receptive fields from representative electrodes on the temporal plane and lateral STG. We derived these speech receptive fields using a nonlinear maximally informative dimensions (MID) method, since this was shown to improve performance over a linear ridge spectrotemporal receptive field (STRF) model in our current data and in previous datasets (Hullett et al., 2016). Similar results were observed using a linear ridge STRF model (Figure S2A).

We found clear evidence for opposing gradients for speech and tone response magnitudes from medial to lateral areas (Figure 3B; Table S1). Overall, responses to speech increased from medial to lateral regions, whereas responses to tones decreased (Figures 3F, S3, and S4). This difference in selectivity is consistent with previous reports of selective preference for complex/speech stimuli in STG and strong tuning to pure tone frequencies on the superior temporal plane (Binder et al., 2000; Démonet et al., 1992; Leaver and Rauschecker, 2010, 2016; Nourski et al., 2012; Steinschneider et al., 2013, 2014).

Despite confirmation of simple representations in core auditory cortex and more complex representations laterally, our results in the pSTG did not align entirely with evidence for a transformation from medial to lateral areas (and from simple to complex). For example, we found significantly more pure-tone receptive fields within pSTG onset sites than in surrounding pSTG non-onset sites. However, these representations themselves were not "simple" like the strong, single-peaked, V-shaped classical receptive fields observed in PT and HG (Figures 3C and 3G). In fact, most of the pure-tone receptive fields (RFs) in pSTG onset areas were complex and multi-peaked (Figure 3G). The proportion of non-tone-responsive, single, and multi-peaked pure-tone RFs differed significantly across all seven anatomical areas ($\chi^2 = 105.7$, df = 12, $p = 4.3 \times 10^{-17}$). Despite fast response latencies, representations in the pSTG onset area differed from both narrow-frequency tuning in posteromedial temporal plane and the preference for complex speech stimuli in surrounding lateral STG.

Tuning for pure tones does not predict responses to speech outside of the core

We next asked whether frequency tuning curves were similar for speech, since previous studies have shown that responses to simple, synthesized stimuli may not predict responses to more complex, natural stimuli (Hamilton and Huth, 2020; Portfors et al., 2009; Schneider and Woolley, 2011; Theunissen et al., 2000, 2001). By directly comparing responses to pure tones and English sentences in the same electrodes, we found a significant difference between core and the lateral auditory areas. Frequency tuning for speech and tones was highly similar in PT and pmHG but diverged for other areas (Figure 3H). Electrodes in PT showed overlapping narrow-band responses to speech and tones (Figure 3E, electrodes 3 and 4), while electrodes in mSTG and pSTG show very different tuning curves depending

(E) Lag-correlation-based connections shown schematically for each of the seven ROIs. Arrows point from the leading ROI to the lagging ROI, color indicates the time delay (lag), and width of the arrow indicates the strength of the cross-correlation. Latency patterns suggest parallel information processing in pSTG onset area and posteromedial temporal plane.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

CellPress

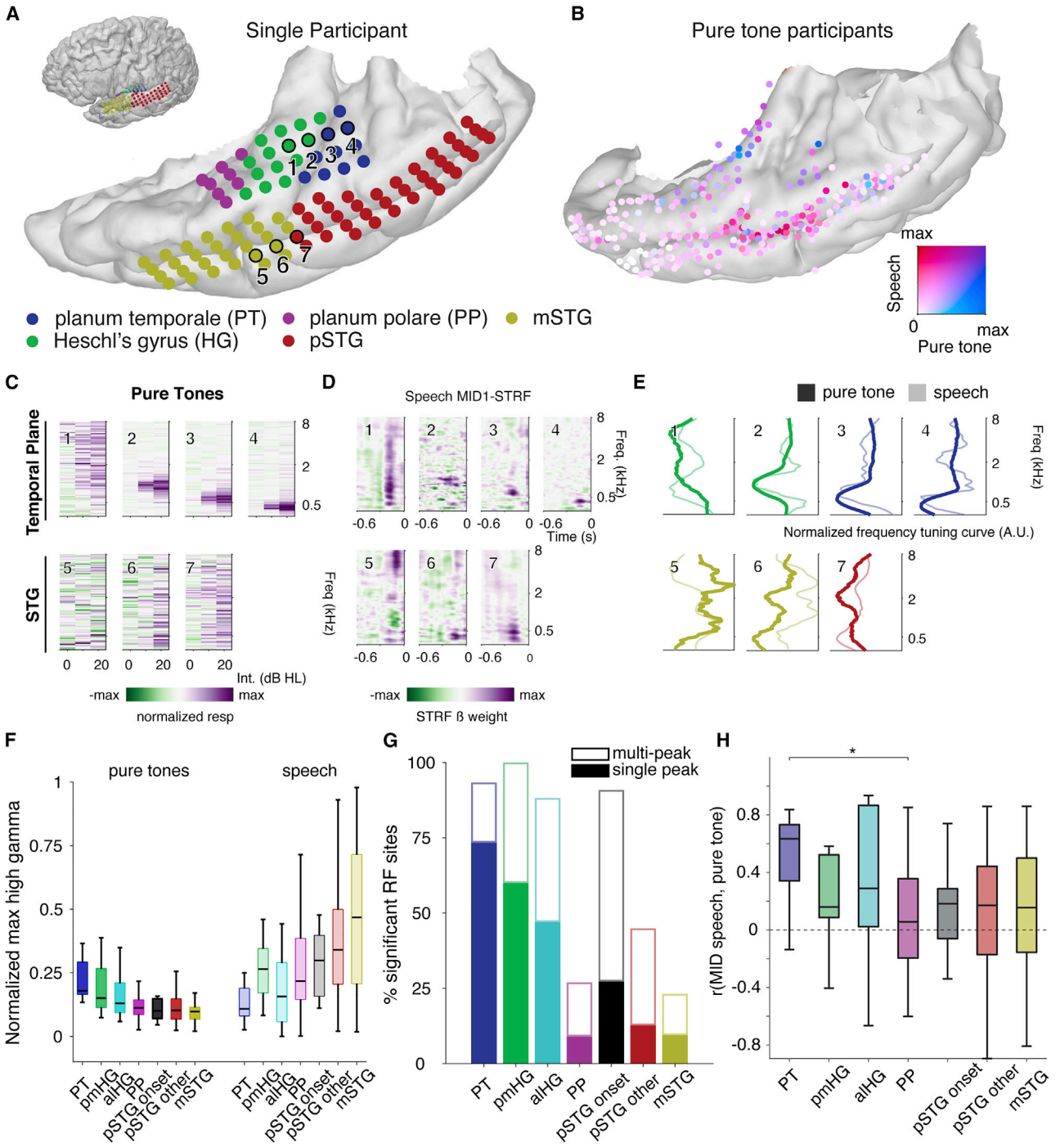


Figure 3. Regional selectivity for speech and pure tones; divergence of tuning curves to simple and complex acoustic inputs in the human auditory cortex

(A) Example temporal plane and lateral temporal cortical grids for one participant. Inset shows the whole brain.

(B) Comparison between normalized response magnitudes to speech and pure tone stimuli across all participants on atlas brain. Electrodes are colored according to the normalized magnitude of the pure-tone response (blue) and speech response during sentence listening (red). Purple indicates mixed selectivity.

(C) Pure-tone receptive fields for electrodes shown in (A).

(D) Speech spectrotemporal receptive fields using maximally informative dimensions (MID1 STRFs) for electrodes in (A).

(E) Comparison of normalized pure-tone and speech tuning curves from sites in (C) and (D).

(F) Maximum pure-tone response and speech responses by area.

(legend continued on next page)

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

on the stimulus type (Figure 3E, electrodes 5–7). The correlation between pure tone and speech frequency tuning was significantly different across anatomical regions (Kruskal-Wallis ANOVA, $\chi^2 = 12.61$, df = 6, p = 0.049). Post-hoc tests using Tukey HSD correction for multiple comparisons showed that correlations between PT speech and tone receptive fields were higher than PP (p = 0.03). A similar relationship was observed when comparing the correlation between frequency tuning from the pure-tone RF and linear model filters (Figure S2), though with these models, pmHG tuning for tones and speech was more similar than the nonlinear model.

These findings demonstrate that speech responses in core areas can be well predicted from their responses to pure tones, whereas outside the core, they cannot and are likely tuned to more complex combinations of spectral features. Crucially, the pSTG onset area, while weakly responsive to tones at fast latencies, does not share similar stimulus representations to the auditory core, again corroborating a distinct role in sound processing. That is, the pSTG onset area is not simply another A1-like area, despite its fast latency. On the contrary, there are multiple very fast response areas, some of which are tonotopic and some of which are onset driven and not tonotopic. These findings suggest that there are fundamentally different representations in the lateral STG compared to medial HG/PT. While this has traditionally been interpreted as evidence of hierarchical processing, another possibility, supported by our earlier latency analysis, is that these areas process different information and receive parallel inputs.

Anatomical separation of pitch, phonetic, and onset information in the auditory cortex

While the tone and speech comparisons showed greater selectivity for speech in the lateral STG, the features of speech that are being encoded there are not clear, especially since they were poorly predicted by pure-tone responses. We therefore performed an encoding model analysis that focused on acoustic and phonetic features of speech. Most of these features have been observed in the STG, but whether these cues are also represented on the temporal plane has been controversial.

We fit models describing tuning for spectrotemporal features (Figure 2D) as well as mixed-feature representations, as depicted in Figure 4A. This analysis revealed that neural populations captured by single electrodes selectively encode only some, but not other, features in the speech signal, with a wide diversity of responses across the human auditory cortex. Figure 4B shows receptive fields for four example electrodes in different participants. E1 encodes mainly phrase-level temporal structure cues by speech onsets, E2 encodes mainly syllable-level temporal structure (peakRate) and phonetic features. In contrast, absolute-pitch features explained significant variance in E3, while E4 encoded relative-pitch features. Across participants, this analysis revealed a functional and anatomical localization of temporal speech features and relative pitch to the STG, with onsets again confined to a zone in pSTG, whereas PT and HG were dominated by representations of spectrotemporal sound acoustics and absolute pitch (Figure 4C).

We compared the full spectrotemporal model to smaller models that contain binary speech features but no information about the precise spectrotemporal structure of the speech stimulus. Comparing a reduced model containing only an onset predictor to a full spectrotemporal model, we found that the additional spectrotemporal information did not improve model fits for a subset of electrodes, located in posterior STG (Figure 4D), in line with our unsupervised NMF analysis to uncover onset selectivity (Figures 1D and S6). Notably, none of these electrodes were located in PP or HG. A similar comparison showed that for a group of electrodes in lateral STG, a temporal and phonological feature model (including binary onsets, peakRate and phonological feature predictors) outperformed the full spectrotemporal model (Figures 4A and 4E). We combined phonological features and peakRate together in this step, as more detailed model comparisons showed these features were represented in mostly overlapping sets of electrodes.

Feature-tuned electrodes were located predominantly in middle STG and not on the temporal plane (Figures 4C and 4E), supporting the role of lateral STG in speech processing. To identify encoding of pitch features, we next tested whether the addition of pitch features to the onset, phonological feature, and peakRate model would improve model performance. We observed a clear separation between absolute pitch encoding on the medial temporal plane (specifically HG and PT) and relative pitch encoding mostly in mid to anterior STG, aHG, and PP (Figures 4C, 4F, and S4). Complex representations of onsets, phonetic features, envelope, and relative pitch are specific to the lateral STG. In contrast, simple representations, including absolute pitch, were predominantly encoded on HG and PT.

Focal ECS reveals double dissociation of regions critical for speech and non-speech processing

The neurophysiological recordings so far suggest a distributed encoding of sound properties of speech, with evidence of early independent processing in both HG and the onset zone of the STG. These results are not consistent with the classical model of simple serial cortical processing hierarchy for speech, from the core auditory cortex to the surrounding belt and parabelt auditory cortex. To causally evaluate the hypothesis of parallel processing in these areas, we used focal ECS on the medial and lateral parts of the auditory cortex in seven participants (Fenoy et al., 2006; Leonard et al., 2016; Sinai et al., 2009). We also investigated differences in artificially evoked sound patterns in HG and STG. A cortical feedforward serial model would predict that stimulation in either region would distort, modulate, or interrupt the perception of spoken words, perhaps at different levels of representation. In contrast, if lateral and medial areas are part

(G) Percentage of sites with significant receptive fields by anatomical area, as measured by significance of within receptive field (RF) responses as compared to outside RF. Percentages are split into single- versus multi-peaked RF.

(H) Correlation between frequency tuning for pure tones (as in D) to frequency tuning for speech (from MID1-STRF, as in E). Boxplot boxes show 25th and 75th percentiles and the median. Whiskers show extreme non-outlier values. * = p-value <0.05.

See also Figures S2 and S3 and Table S1.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

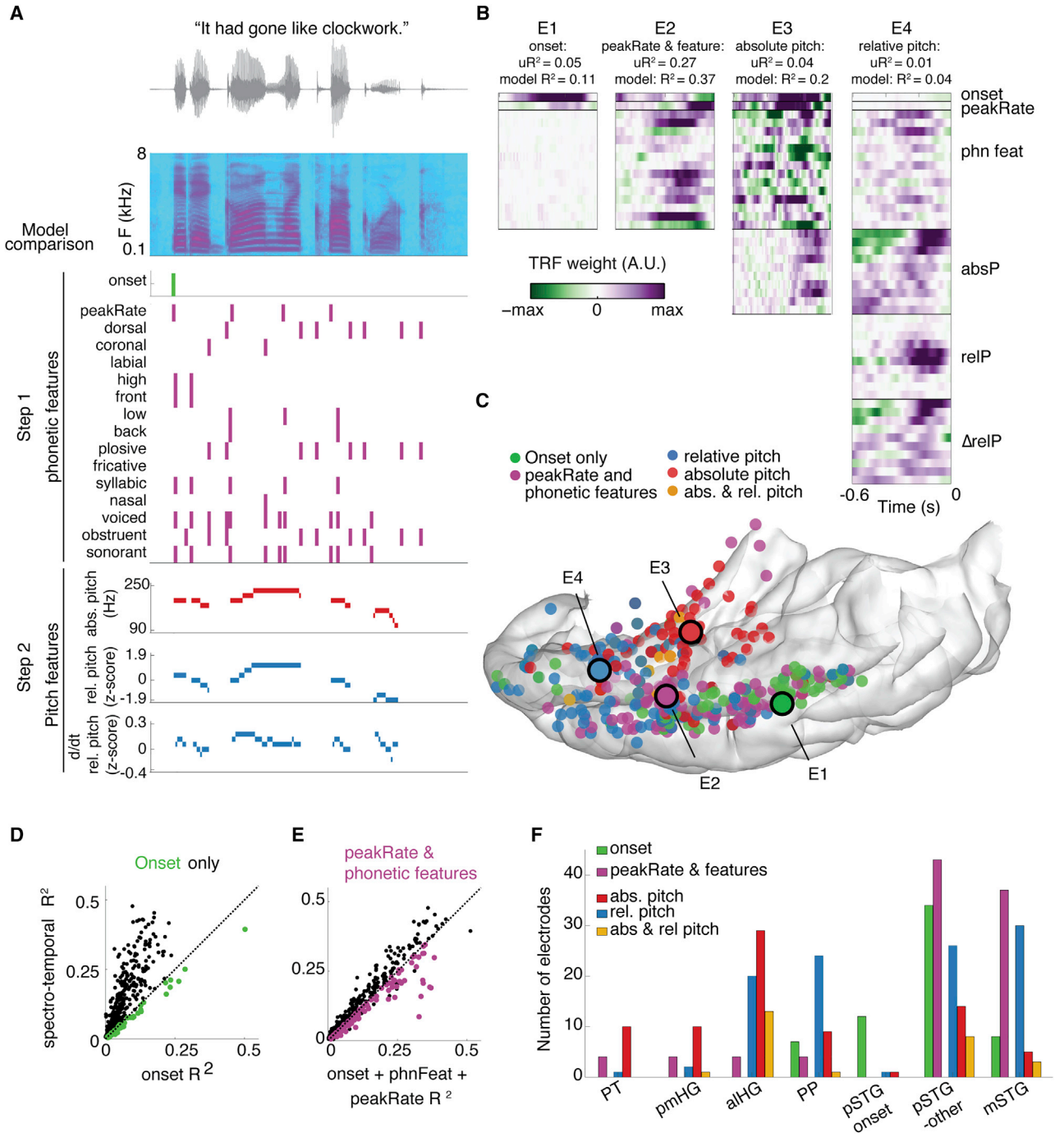


Figure 4. Regional selectivity for speech features

(A) Speech features tested in feature model comparisons for an example sentence.

(B) Receptive fields for example electrodes. The feature with maximal unique R^2 is indicated for each electrode. Electrodes were chosen for which distinct sets of features explain a large portion of variance. For example, the best model for electrode 1 includes onset, peakRate, and phonetic features, but 50% of overall explained variance is attributed to the onset predictor.

(C) Location of electrodes primarily coding for speech onsets, phonetic features and peakRate, relative pitch, and absolute pitch. Electrodes in (B) are circled.

(D and E) Onset (D) and feature-encoding electrodes (E) are defined as those for which the respective model outperforms a spectrotemporal model.

(F) Anatomical distribution of electrodes coding for different features. Absolute pitch dominates representations in PT and pmHG, and relative pitch is primarily represented in PP, aIHG, and STG ($\chi^2 = 62.5$, $p < 10^{-9}$). Orange = both relative and absolute pitch contribute unique variance, permutation $p < 0.05$. Onset-encoding electrodes are mostly located in pSTG, whereas phonological features and peakRate are represented in posterior and anterior lateral STG.

Related to Figure S5.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

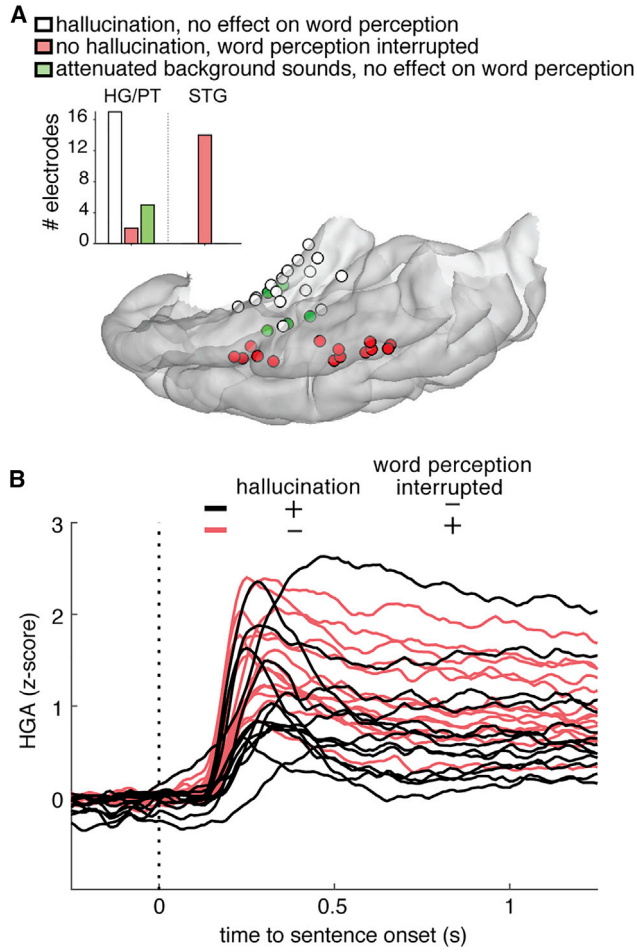


Figure 5. Electrocortical stimulation of Heschl's gyrus and superior temporal gyrus

(A) Focal electrocortical stimulation shows double dissociation between effects of stimulation on HG and lateral STG. Stimulation in HG evoked auditory hallucinations but did not interfere with word perception and repetition. Participants could not perceive words during stimulation on lateral STG, but no additional sound hallucinations were evoked.

(B) Z-scored high gamma amplitude (HGA) responses to sentences on electrodes in (A) did not differ between sites with different stimulation effects. Black traces show the evoked sentence response in electrodes where stimulation caused an auditory hallucination but no change in word perception. Red traces are sites where no hallucination occurred, but word perception was interrupted with stimulation.

Related to Figure S5 and Video S1.

of independent parallel processing streams, one would predict dissociable effects. We hypothesized that stimulating core HG would evoke simple sound percepts, whereas stimulating STG would evoke complex or speech/phonemic sounds based upon their receptive field properties. As part of a clinical mapping protocol, we stimulated each site multiple times with increasing current amplitude to carefully assess its function.

First, participants were asked to report what they perceived while cortical sites on the temporal plane and lateral STG were stimulated (Video S1; Table S2). An immediate sound hallucination was evoked by stimulation of sites on the temporal plane (20 electrode sites; Figure 5A, white electrodes), but not by stimulation on lateral STG (17 electrode sites; Figure 5A, red electrodes). The sound percept was described as occurring near the contralateral ear, starting at very low stimulation thresholds (1–2 mA, 50 Hz). During HG stimulation, participants described sounds like "running water," "moving gravel," "fan sound," "tapping," and "buzzing." Further inquiry led to descriptions like "fast modulation of sound," "like waves of sound," or "b-b-b-b." Hallucinations were reliably evoked with every trial of stimulation (applied 5–25 times at each site). We further characterized the effects of different stimulation parameters on a subset of sites, as permitted by clinical protocol (11 sites). When current amplitude was increased by 1–2 mA, participants reported "louder" sounds at every cortical site where sound could be evoked. When current frequency was increased from 50 Hz to 100 Hz, they also reported louder and "higher-pitch sounds." One participant reported hearing two tones during 100-Hz stimulation, one at the original pitch when stimulated with 50 Hz and an additional, superimposed higher-pitch tone. The converse was observed when we reduced the current amplitude by 1 mA or frequency to 10 Hz, with participants reporting "more quiet now" and "slower" sounds. When stimulation duration was changed to 5 s, participants could accurately report the start and end suggesting a sustained evoked perception (two sites tested).

To assess how stimulation affected speech perception, ECS was synchronized with single-word presentation, and participants were asked to repeat what they heard. Interestingly, participants did not have any difficulty in hearing words during stimulation on temporal plane sites, including HG. Stimulation on sites that evoked a sound hallucination had no effect on speech intelligibility, nor did stimulation affect the sound quality of perception of the spoken words (0 out of 20 sites tested; see also Video S1). Importantly, participants frequently reported hearing both the evoked sound percepts and the spoken words independently and without distortion of either (12 sites). Two participants also reported bilateral attenuation of sounds from stimulation of sites in the anterolateral HG (Figure 5A, green electrodes). In these cases, they reported that environmental sounds such as the fan noise and hospital equipment beeping were "muffled," but speech was clear (two sites).

Electrical stimulation of the lateral STG had a completely different effect (14 sites). Participants were unable to detect when electrical stimulation was applied. That is, there was no evoked sound hallucination with stimulation despite the fact the same sites exhibited robust and clear tuning to speech features. However, when stimulation was applied during spoken words, participants experienced significant impairments in speech perception. They commonly reported: "I can't hear" or "I could hear you speaking but can't make out the words." One participant reported that syllables in the word seemed "swapped." They did not report that it was quieter or muffled. This is consistent with a previous report that demonstrated impairments in phonetic discrimination for speech sounds but no effect on tone discrimination (Sinai et al., 2009). We observed occasional paraphasic errors when repeating words, with phonemic substitutions or deletions. These effects could not be predicted by evoked responses alone, as sites with induced

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

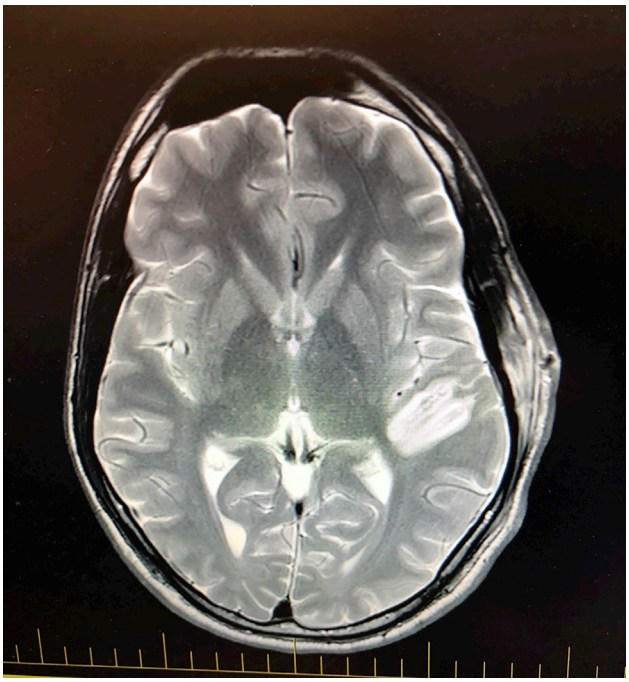


Figure 6. Focal ablation of the left HG without damage to the pSTG has no effect on speech perception or language comprehension

Magnetic resonance (MR) image shows the extent of the surgical ablation in the axial plane along the axis of the HG, sparing pSTG. Image is shown in radiological orientation.

hallucination or word perception effects showed no difference in high-gamma responses (Figure 5B).

We observed an unexpected and striking double dissociation of effects of stimulation on medial HG and lateral STG. Stimulation of HG evoked clear contralateral sound percepts, without interruption or distortion of word perception. In contrast, stimulation of the lateral STG did not evoke any sounds, but interfered significantly with speech processing. These results suggest that medial and lateral auditory areas may be part of parallel sound analysis pathways, rather than part of a commonly assumed single serial, hierarchical pathway. HG does not appear to be required for speech perception, whereas STG does.

Focal ablation of the left HG does not affect speech comprehension

Additional causal evidence was sought for evaluating the role of HG in speech perception. Injury to the posterior STG is strongly associated with impairments in speech perception and comprehension, as well as language production (Hillis et al., 2017). The consequences of lesioning HG are less clear, as selective stroke or surgical resections there are extraordinarily rare. We present a case study of a patient that underwent a selective ablation of HG. The patient is a 33-year-old right-handed man with a history of refractory seizures with auditory auras. The seizure semiology was described as a "ringing, high-pitched sound from around the right ear." As the seizure progressed, the intensity of the sound would increase in loudness and could last tens of seconds or minutes. At the start, he could hear and comprehend speech from others while concurrently experiencing the high-pitched sound aura. For example, he could speak to his mother, who would then verbally instruct him to lie down before the seizure became worse. Per his report, hearing words spoken by others was normal and did not sound distorted during the aura. In most episodes, the seizure self-resolved. Occasionally, the ongoing seizure propagation spread was associated with inability to comprehend language. His speech became "incoherent" with "mixed up words," and afterward, it evolved into a secondarily generalized convulsive seizure. High-resolution MRI of his brain was normal.

To localize his seizure onset zone, he underwent stereotactic implantation of two parallel multi-electrode depth leads, placed longitudinally along the long axis of the left HG. Ten additional leads were placed in other brain areas. His seizures were found to originate from HG electrodes. Bedside stimulation mapping was performed, and electrical stimulation mapping of the HG electrodes reproduced his auditory auras with sound hallucination. Furthermore, he was able to comprehend speech and had fluent language during stimulation mapping of HG.

Because he had preserved language functions during stimulation mapping, the decision was made to proceed with thermocoagulation of HG (see Figure 6). This was carried out through the same indwelling electrode leads (Bourdillon et al., 2017) while the patient was fully awake and carefully assessed throughout ablations at each site along the lead. He tolerated this procedure well, without changes in speech or language comprehension. Post-treatment MRI showed excellent ablation of the left HG with preservation of the adjacent lateral STG. Formal audiometry after the procedure demonstrated normal bilateral audiograms, as well as normal speech comprehension and production. He was seizure free for 1 year after the HG thermocoagulation ablation. This case study provides additional causal evidence that sound processing in left HG is not necessary for speech comprehension.

DISCUSSION

The human auditory system decomposes the speech signal into components that are relevant for perception. In this study, we provide a characterization of speech responses across the human auditory cortex, including the HG, surrounding areas of PT and PP, and the posterior to middle extent of lateral STG. Microsurgical access to the Sylvian fissure provided dense simultaneous recordings of the highly heterogeneous responses to speech from all regions of the human auditory cortex, in contrast to previous intracranial approaches that relied on only piecemeal sampling from one of these regions at a time. We evaluated the role of each area in processing of speech sounds using converging experimental approaches probing the timing and order of activation, the nature of simple and complex sound representations in each area, and their causal role in speech comprehension using functional and surgical ablation.

Our findings on the functional organization of the human primary and parabelt auditory cortex are in line with previous studies that employed depth recording electrodes in HG to show that core auditory cortical areas show tonotopic organization and fast-latency responses to click trains and can track pitch changes in pure tones (Brugge et al., 2009; Griffiths et al., 2010;

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

Howard et al., 1996; Steinschneider et al., 2014), as well as with noninvasive tonotopic mapping using fMRI (Barton et al., 2012; Da Costa et al., 2011; Dick et al., 2017; Humphries et al., 2010; Leaver and Rauschecker, 2016; Saenz and Langers, 2014; Schönwiesner et al., 2015; Talavage et al., 2004; Wessinger et al., 1997; Woods and Alain, 2009; Woods et al., 2010). In contrast, responses in PP were slower and more similar to higher-order areas of mid- to anterior STG, with different receptive fields for tones and speech, little frequency selectivity, and encoding of relative rather than absolute pitch in speech.

Neural responses to speech were strongest in lateral STG, where selectivity was greater for acoustic-phonetic and prosodic features than to pure tones. In addition to replicating the existence of an onset zone in pSTG (Hamilton et al., 2018), we found functionally distinct, but anatomically interleaved, populations in mid-STG representing different linguistically relevant features in speech. These included phonological features, acoustic onset edges that cue syllables (Oganian and Chang, 2019), and relative pitch, which is the main cue to intonational prosody (Tang et al., 2017). While relative pitch was also represented in PP, peakRate and phonetic features were represented predominantly in middle STG, supporting its role for speech processing. Of note, this is consistent with a spatial population code for speech cues that are both short (phoneme segment length; e.g., consonant features) and relatively long (suprasegmental; e.g., prosodic cues) in duration.

Processing for absolute versus relative pitch was distinctly regionalized. In our previous work in the STG, absolute-pitch responses were rare compared to relative-pitch and phonological representations (Tang et al., 2017). Here, we found that absolute-pitch selectivity dominated in the temporal plane (HG and PT). Absolute-pitch sensitivity has been observed in nonhuman primates at the anterolateral border of the auditory core (Bendor and Wang, 2005, 2010) and fits well with the narrow, low spectral tuning for pure tones and speech vocal pitch in these areas. In contrast, relative-pitch representations dominated in mid-anterior lateral STG and PP. This is consistent with previous human studies, where sounds with pitch activate more of lateral HG than sounds without pitch and sounds with pitch variation activate regions of PP and anterior STG (Patterson et al., 2002). Such selectivity may be analogous to voice-selective areas on the anterior temporal plane in macaque (Perrodin et al., 2011; Petkov et al., 2008, 2009).

On the other hand, our response latency analysis challenges current models of information flow from primary to parabelt auditory cortex (Brodbeck et al., 2018; Hickok and Poeppel, 2007; Jasmin et al., 2019; Rauschecker and Tian, 2000; Saenz and Langers, 2014). For example, the comparably short response latencies in posteromedial HG, PT, and the pSTG onset area and the differences in representational content of these areas do not support simple serial processing. The pSTG onset area responded to onsets exclusively, with a broad spectral but narrow temporal response, a pattern that was not seen in temporal plane regions. This region is not purely speech selective and also responds to non-speech and synthetic sound onsets (Hamilton et al., 2018). In contrast, HG/PT STRFs were more selective spectrally. The similar response timescales in these areas indicate that the onset zone in pSTG reflects early processing occurring in parallel with the computations performed by circuits on the temporal plane itself (Nourski et al., 2014), with a general pattern that the fastest and earliest activations occur across the entire posterior aspect of the human auditory cortex.

Despite the similarities in latency, ECS of contacts in lateral and superior temporal areas revealed a striking functional and anatomical double dissociation. Stimulation of sites in posteromedial temporal plane induced vivid sound hallucinations on the contralateral side but no impairment of speech perception, whereas stimulation of lateral STG had the opposite effect. This aligns with clinical studies showing that resection of HG does not result in speech comprehension deficits (Russell and Golfinos, 2003; Sakurada et al., 2007; Silbergeld, 1997). In contrast, damage to left lateral STG results in severe speech comprehension (and language production) deficits (Butler et al., 2014; Wernicke, 1874). Together, this suggests that the primary auditory cortex on HG is not the main source of input to the entire STG. Rather, we propose that pSTG receives direct input from outside the auditory core.

"Core" auditory cortex is defined as the heavily myelinated, tonotopic region that receives thalamic projections from the ventral medial geniculate body (vMGB) via the lemniscal auditory pathway (Bartlett, 2013; Dick et al., 2012; Galaburda and Sanides, 1980; Hackett, 2011; Hackett et al., 2001, 2007; Scott et al., 2017a). However, it is largely underappreciated that the "parabelt" auditory cortex in the STG receives direct and distinct thalamic projections via the nonlemniscal pathway, from the medial and dorsal divisions of the MGB, and pulvinar (Bartlett, 2013; Hackett et al., 1998, Scott et al., 2017a). Of note, the medial and dorsal divisions of MGB are significantly larger relative to the ventral division in humans compared to nonhuman primates (Brugge and Howard, 2002; Winer, 1984). Functionally, sound representation in core and non-core areas is dramatically different; core receptive fields have narrow frequency tuning and are tuned to contralateral sounds (Bitterman et al., 2008; Khalighinejad et al., 2021), whereas lateral STG has complex spectrotemporal but poor spatial selectivity. A parsimonious interpretation is that distinct parallel thalamocortical pathways process different aspects of the speech signal. We speculate that the non-lemniscal pathway plays an essential role for speech perception and therefore may not require core A1 processing. This is an alternative to the mainstream model of core-belt-parabelt cortical pathway that is thought to underlie a hierarchical processing of progressively complex, abstract features. However, it is likely that some relevant aspects of speech (i.e., localization or acoustic quality) occur via cortico-cortical interactions between areas with different thalamocortical inputs.

Thus, the auditory cortical system may have parallel organization to a greater extent than the visual system (Rauschecker et al., 1997). For example, in the ventral stream for object recognition, feedforward connections relay information hierarchically from the lateral geniculate nucleus (LGN) to V1 and onward to V4, through successively larger retinotopically organized receptive fields (Felleman and Van Essen, 1991; Sereno et al., 1995). While increasing evidence shows additional parallel processing within the visual system, injury to V1 still has

immediate and long-standing effects on visual processing. In contrast, the auditory system is heavily parallel throughout, A1 (but not nonprimary STG) is organized tonotopically, and isolated A1 injury has no clear consequences for speech perception or intelligibility.

Our results demonstrate a comprehensive cortical map of the acoustic and phonetic representations underlying human speech perception. Speech representations are distributed across the human auditory cortex, with clear parallel processing as well as potential serial processing at longer latencies in the anterior and middle STG. Overall, our findings speak to a distributed mosaic of specialized processing units, each representing different acoustic and phonetic cues in the speech signal, the combination of which creates the rich experience of natural speech comprehension.

Limitations of the study

A limitation of the current study is that our stimulation and ablation results involved unilateral and not bilateral HG. While neural activity during speech processing is largely bilateral (Cogan et al., 2014), in clinical language mapping, unilateral stimulation of STG alone on the language dominant side results in comprehension deficits. Thus, the other hemisphere is unable to compensate for this disruption. Previous work has shown that the ipsilateral primary auditory cortex connects to contralateral primary auditory cortex, but not to contralateral STG (and vice versa) (Hackett et al., 1999; Kaas and Hackett, 2000). Thus, it is not possible for pSTG to receive inputs through contralateral A1 directly. Information might instead travel from contralateral A1 to contralateral STG and then to ipsilateral STG. It is also possible that cross-hemispheric connectivity could be present between STG and alHG, but based on our latency analysis, this seems unlikely. Future studies incorporating bilateral stimulation may be able to uncover to what extent pSTG and HG are truly functionally independent.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.07.019.

ACKNOWLEDGMENTS

The authors would like to thank Matthew Leonard and Brian Malone for helpful comments on the manuscript. The authors also thank Michael T. Lawton, Neal Fox, Matthew Leonard, Matthias Sjerps, Kunal Raygor, and Leah Muller for assistance with intraoperative recordings and Patrick Hullett for assistance with MID analysis and comments on the manuscript. This work was supported by grants from the NIH (F32 DC014192-01 to L.S.H. and R01-DC012379 and U01-NS117765 to E.F.C.). This research was also supported by Bill and Susan Oberndorf, the Joan and Sandy Weill Foundation, and the William K. Bowes Foundation. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

AUTHOR CONTRIBUTIONS

L.S.H. and E.F.C. conceived the neurophysiology recording experiments. L.S.H., Y.O., E.F.C., and others collected the data. L.S.H. and Y.O. analyzed the data. J.H. contributed data from the thermoablation case study. L.S.H., Y.O., and E.F.C. wrote and revised the paper. E.F.C. performed the surgeries, designed the stimulation experiments, and supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 15, 2020
Revised: February 11, 2021
Accepted: July 19, 2021
Published: August 18, 2021

REFERENCES

Aertsen, A.M., and Johannesma, P.I.M. (1981). The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol. Cybern. *42*, 133–143.

Atencio, C.A., Sharpee, T.O., and Schreiner, C.E. (2008). Cooperative nonlinearities in auditory cortical neurons. Neuron *58*, 956–966.

Bartlett, E.L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. Brain Lang. *126*, 29–48.

Barton, B., Venezia, J.H., Saberi, K., Hickok, G., and Brewer, A.A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. Proc. Natl. Acad. Sci. USA *109*, 20738–20743.

Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. Nature *436*, 1161–1165.

Bendor, D., and Wang, X. (2010). Neural coding of periodicity in marmoset auditory cortex. J. Neurophysiol. *103*, 1809–1822.

Berezutskaya, J., Freudenburg, Z.V., Güçlü, U., van Gerven, M.A.J., and Ramsey, N.F. (2017). Neural tuning to low-level features of speech throughout the perisylvian cortex. J. Neurosci. *37*, 7906–7920.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., and Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J. Neurosci. *28*, 14301–14310.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., and Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. J. Neurosci. *27*, 9252–9261.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., and Possing, E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex *10*, 512–528.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., and Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature *451*, 197–201.

Bourdillon, P., Isnard, J., Catenoix, H., Montavont, A., Rheims, S., Ryvlin, P., Ostrowsky-Coste, K., Mauguiere, F., and Guénot, M. (2017). Stereo electroencephalography-guided radiofrequency thermocoagulation (SEEG-guided RF-TC) in drug-resistant focal epilepsy: Results from a 10-year experience. Epilepsia *58*, 85–93.

Bouthillier, A., Surbeck, W., Weil, A.G., Tayah, T., and Nguyen, D.K. (2012). The hybrid operculo-insular electrode: a new electrode for intracranial investigation of perisylvian/insular refractory epilepsy. Neurosurgery *70*, 1574–1580, discussion 1580.

Brewer, A.A., and Barton, B. (2016). Maps of the Auditory Cortex. Annu. Rev. Neurosci. *39*, 385–407.

Brodbeck, C., Presacco, A., and Simon, J.Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. Neuroimage *172*, 162–174.

Brugge, J.F., and Howard, M.A. (2002). Hearing. In Encyclopedia of the Human Brain, V.S. Ramachandran, ed. (Academic Press), pp. 429–448.

Brugge, J.F., Volkov, I.O., Garell, P.C., Reale, R.A., and Howard, M.A., 3rd. (2003). Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. J. Neurophysiol. *90*, 3750–3763.

Brugge, J.F., Nourski, K.V., Oya, H., Reale, R.A., Kawasaki, H., Steinschneider, M., and Howard, M.A., 3rd. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. J. Neurophysiol. *102*, 2358–2374.

Butler, R.A., Lambon Ralph, M.A., and Woollams, A.M. (2014). Capturing multidimensionality in stroke aphasia: mapping principal behavioural components to neural structures. Brain *137*, 3248–3266.

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

Cheung, C., Hamilton, L.S., Johnson, K., and Chang, E.F. (2016). The auditory representation of speech sounds in human motor cortex. eLife *5*, 1–19.

Chevillet, M., Riesenhuber, M., and Rauschecker, J.P. (2011). Functional correlates of the anterolateral processing hierarchy in human auditory cortex. J. Neurosci. *31*, 9345–9352.

Cogan, G.B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., and Pesaran, B. (2014). Sensory–motor transformations for speech occur bilaterally. Nature *507* (7490), 94–98.

Da Costa, S., van der Zwaag, W., Marques, J.P., Frackowiak, R.S.J., Clarke, S., and Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl's gyrus. J. Neurosci. *31*, 14067–14075.

Dalca, A.V., Danagoulian, G., Kikinis, R., Schmidt, E., and Golland, P. (2011). Segmentation of nerve bundles and ganglia in spine MRI using particle filters. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 537–545.

de Heer, W.A., Huth, A.G., Griffiths, T.L., Gallant, J.L., and Theunissen, F.E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. J. Neurosci. *37*, 6539–6557.

Démonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., and Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. Brain *115*, 1753–1768.

Dick, F., Tierney, A.T., Lutti, A., Josephs, O., Sereno, M.I., and Weiskopf, N. (2012). In vivo functional and myeloarchitectonic mapping of human primary auditory areas. J. Neurosci. *32*, 16095–16105.

Dick, F.K., Lehet, M.I., Callaghan, M.F., Keller, T.A., Sereno, M.I., and Holt, L.L. (2017). Extensive Tonotopic Mapping across Auditory Cortex Is Recapitulated by Spectrally Directed Attention and Systematically Related to Cortical Myeloarchitecture. J. Neurosci. *37*, 12187–12201.

Edwards, E., Soltani, M., Kim, W., Dalal, S.S., Nagarajan, S.S., Berger, M.S., and Knight, R.T. (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. J. Neurophysiol. *102*, 377–386.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Fenoy, A.J., Severson, M.A., Volkov, I.O., Brugge, J.F., and Howard, M.A., 3rd. (2006). Hearing suppression induced by electrical stimulation of human auditory cortex. Brain Res. *1118*, 75–83.

Fischl, B., Sereno, M.I., Tootell, R.B.H., and Dale, A.M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. *8*, 272–284.

Flinker, A., Chang, E.F., Barbaro, N.M., Berger, M.S., and Knight, R.T. (2011). Sub-centimeter language organization in the human temporal lobe. Brain Lang. *117*, 103–109.

Friederici, A.D. (2015). White-matter pathways for speech and language processing. In Handbook of Clinical Neurology, Chapter 10, M.J. Aminoff, F. Boller, and D.F. Swaab, eds. (Elsevier), pp. 177–186.

Galaburda, A., and Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. J. Comp. Neurol. *190*, 597–610.

Garofolo, J.S., Lamel, L.F., Fisher, W.F., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L., and Zue, V. (1993). TIMIT Acoustic-Phonetic Continuous Speech Corpus (Linguistic Data Consortium).

Griffiths, T.D., and Warren, J.D. (2002). The planum temporale as a computational hub. Trends Neurosci. *25*, 348–353.

Griffiths, T.D., Kumar, S., Sedley, W., Nourski, K.V., Kawasaki, H., Oya, H., Patterson, R.D., Brugge, J.F., and Howard, M.A. (2010). Direct recordings of pitch responses from human auditory cortex. Curr. Biol. *20*, 1128–1132.

Hackett, T.A. (2011). Information flow in the auditory cortical network. Hear. Res. *271*, 133–146.

Hackett, T.A., Stepniewska, I., and Kaas, J.H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. J. Comp. Neurol. *394*, 475–495.

Hackett, T.A., Stepniewska, I., and Kaas, J.H. (1999). Callosal connections of the parabelt auditory cortex in macaque monkeys. Eur. J. Neurosci. *11*, 856–866.

Hackett, T.A., Preuss, T.M., and Kaas, J.H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. *441*, 197–222.

Hackett, T.A., De La Mothe, L.A., Ulbert, I., Karmos, G., Smiley, J., and Schroeder, C.E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. J. Comp. Neurol. *502*, 924–952.

Hamilton, L.S., Chang, D.L., Lee, M.B., and Chang, E.F. (2017). Semi-automated anatomical labeling and inter-subject warping of high-density intracranial recording electrodes in electrocorticography. Frontiers in Neuroinformatics *11*, 62.

Hamilton, L.S., Edwards, E., and Chang, E.F. (2018). A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. Curr. Biol. *28*, 1860–1871.e4.

Hamilton, L.S., and Huth, A.G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. Lang. Cogn. Neurosci. *35* (5), 573–582.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. Nat. Rev. Neurosci. *8*, 393–402.

Hickok, G., and Saberi, K. (2012). Redefining the Functional Organization of the Planum Temporale Region: Space, Objects, and Sensory–Motor Integration. In The Human Auditory Cortex, D. Poeppel, ed. (Springer), pp. 333–350.

Hillis, A.E., Rorden, C., and Fridriksson, J. (2017). Brain regions essential for word comprehension: Drawing inferences from patients. Ann. Neurol. *81*, 759–768.

Howard, M.A., 3rd, Volkov, I.O., Abbas, P.J., Damasio, H., Ollendieck, M.C., and Granner, M.A. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. Brain Res. *724*, 260–264.

Howard, M.A., Volkov, I.O., Mirsky, R., Garell, P.C., Noh, M.D., Granner, M., Damasio, H., Steinschneider, M., Reale, R.A., Hind, J.E., and Brugge, J.F. (2000). Auditory cortex on the human posterior superior temporal gyrus. J. Comp. Neurol. *416*, 79–92.

Hullett, P.W., Hamilton, L.S., Mesgarani, N., Schreiner, C.E., and Chang, E.F. (2016). Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. J. Neurosci. *36*, 2014–2026.

Humphries, C., Liebenthal, E., and Binder, J.R. (2010). Tonotopic organization of human auditory cortex. Neuroimage *50*, 1202–1211.

Jasmin, K., Lima, C.F., and Scott, S.K. (2019). Understanding rostral-caudal auditory cortex contributions to auditory perception. Nat. Rev. Neurosci. *20*, 425–434.

Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. USA *97*, 11793–11799.

Khalighinejad, B., Herrero, J.L., Bickel, S., Mehta, A.D., and Mesgarani, N. (2021). Functional characterization of human Heschl's gyrus in response to natural speech. Neuroimage *235*, 118003.

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

Leaver, A.M., and Rauschecker, J.P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. *30*, 7604–7612.

Leaver, A.M., and Rauschecker, J.P. (2016). Functional Topography of Human Auditory Cortex. J. Neurosci. *36*, 1416–1428.

Leonard, M.K., Cai, R., Babiak, M.C., Ren, A., and Chang, E.F. (2016). The peri-Sylvian cortical network underlying single word repetition revealed by electrocortical stimulation and direct neural recordings. Brain Lang. *193*, 58–72.

Liégeois-Chauvel, C., de Graaf, J.B., Laguitton, V., and Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb. Cortex *9*, 484–496.

Malak, R., Bouthillier, A., Carmant, L., Cossette, P., Giard, N., Saint-Hilaire, J.-M., Nguyen, D.B., and Nguyen, D.K. (2009). Microsurgery of epileptic foci in the insular region. J. Neurosurg. *110*, 1153–1163.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E.F. (2014). Phonetic feature encoding in human superior temporal gyrus. Science *343*, 1006–1010.

Moerel, M., De Martino, F., and Formisano, E. (2012). Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J. Neurosci. *32*, 14205–14216.

Moerel, M., De Martino, F., and Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. Front. Neurosci. *8*, 225.

Moses, D.A., Mesgarani, N., Leonard, M.K., and Chang, E.F. (2016). Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. J. Neural Eng. *13*, 056004.

Norman-Haignere, S., Kanwisher, N.G., and McDermott, J.H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. Neuron *88*, 1281–1296.

Nourski, K.V., Steinschneider, M., Oya, H., Kawasaki, H., Jones, R.D., and Howard, M.A., III. (2012). Spectral organization of the human lateral superior temporal gyrus revealed by intracranial recordings. Cereb. Cortex.

Nourski, K.V., Steinschneider, M., McMurray, B., Kovach, C.K., Oya, H., Kawasaki, H., and Howard, M.A., 3rd. (2014). Functional organization of human auditory cortex: investigation of response latencies through direct recordings. Neuroimage *101*, 598–609.

Oganian, Y., and Chang, E.F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci. Adv. *5*, eaay6279.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., and Hickok, G. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb. Cortex *20*, 2486–2495.

Ozker, M., Schepers, I.M., Magnotti, J.F., Yoshor, D., and Beauchamp, M.S. (2017). A Double Dissociation between Anterior and Posterior Superior Temporal Gyrus for Processing Audiovisual Speech Demonstrated by Electrocorticography. J. Cogn. Neurosci. *29*, 1044–1060.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., and Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. Neuron *36*, 767–776.

Perrodin, C., Kayser, C., Logothetis, N.K., and Petkov, C.I. (2011). Voice cells in the primate temporal lobe. Curr. Biol. *21*, 1408–1415.

Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N.K. (2008). A voice region in the monkey brain. Nat. Neurosci. *11*, 367–374.

Petkov, C.I., Logothetis, N.K., and Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? Neuroscientist *15*, 419–429.

Portfors, C.V., Roberts, P.D., and Jonson, K. (2009). Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. Neuroscience *162*, 486–500.

Rauschecker, J.P., and Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. *12*, 718–724.

Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. USA *97*, 11800–11806.

Rauschecker, J.P., Tian, B., Pons, T., Mishkin, M., and Carolina, N. (1997). Serial and parallel processing in rhesus monkey auditory cortex. J. Comp. Neurol. *382*, 89–103.

Ray, S., and Maunsell, J.H.R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. *9*, e1000610.

Russell, S.M., and Golfinos, J.G. (2003). Amusia following resection of a Heschl gyrus glioma. Case report. J. Neurosurg. *98*, 1109–1112.

Saenz, M., and Langers, D.R.M. (2014). Tonotopic mapping of human auditory cortex. Hear. Res. *307*, 42–52.

Sakurada, K., Sato, S., Sonoda, Y., Kokubo, Y., Saito, S., and Kayama, T. (2007). Surgical resection of tumors located in subcortex of language area. Acta Neurochir. (Wien) *149*, 123–129, discussion 129–130.

Schneider, D.M., and Woolley, S.M.N. (2011). Extra-classical tuning predicts stimulus-dependent receptive fields in auditory neurons. J. Neurosci. *31*, 11867–11878.

Schönwiesner, M., and Zatorre, R.J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc. Natl. Acad. Sci. USA *106*, 14611–14616.

Schönwiesner, M., Dechent, P., Voit, D., Petkov, C.I., and Krumbholz, K. (2015). Parcellation of Human and Monkey Core Auditory Cortex with fMRI Pattern Classification and Objective Detection of Tonotopic Gradient Reversals. Cereb. Cortex *25*, 3278–3289.

Schotola, T. (1984). On the use of demisyllables in automatic word recognition. Speech Commun. *3*, 63–87.

Scott, B.H., Saleem, K.S., Kikuchi, Y., Fukushima, M., Mishkin, M., and Saunders, R.C. (2017a). Thalamic connections of the core auditory cortex and rostral supratemporal plane in the macaque monkey. J. Comp. Neurol. *525*, 3488–3513.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

Sharpee, T., Rust, N.C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput. *16*, 223–250.

Silbergeld, D.L. (1997). Tumors of Heschl's gyrus: report of two cases. Neurosurgery *40*, 389–392.

Sinai, A., Crone, N.E., Wied, H.M., Franaszczuk, P.J., Miglioretti, D., and Boatman-Reich, D. (2009). Intracranial mapping of auditory perception: event-related responses and electrocortical stimulation. Clin. Neurophysiol. *120*, 140–149.

Slaney, M. (1998). Auditory toolbox. Interval Research Corporation, Tech. Rep *10* (1998), 1194.

Steinschneider, M., Nourski, K.V., and Fishman, Y.I. (2013). Representation of speech in human auditory cortex: is it special? Hear. Res. *305*, 57–73.

Steinschneider, M., Nourski, K.V., Rhone, A.E., Kawasaki, H., Oya, H., and Howard, M.A., 3rd. (2014). Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. Front. Neurosci. *8*, 240.

Talavage, T.M., Sereno, M.I., Melcher, J.R., Ledden, P.J., Rosen, B.R., and Dale, A.M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. J. Neurophysiol. *91*, 1282–1296.

Tang, C., Hamilton, L.S., and Chang, E.F. (2017). Intonational speech prosody encoding in the human auditory cortex. Science *357*, 797–801.

Theunissen, F.E., Sen, K., and Doupe, A.J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J. Neurosci. *20*, 2315–2331.

Theunissen, F.E., David, S.V., Singh, N.C., Hsu, A., Vinje, W.E., and Gallant, J.L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network *12*, 289–316.

Upadhyay, J., Silver, A., Knaus, T.A., Lindgren, K.A., Ducros, M., Kim, D.-S., and Tager-Flusberg, H. (2008). Effective and structural connectivity in the human auditory cortex. J. Neurosci. *28*, 3341–3349.

Wernicke, C. (1874). Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis (Cohn).

Wessinger, C.M., Buonocore, M.H., Kussmaul, C.L., and Mangun, G.R. (1997). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. Hum. Brain Mapp. *5*, 18–25.

Wessinger, C.M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J.P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J. Cogn. Neurosci. *13*, 1–7.

Winer, J.A. (1984). The human medial geniculate body. Hear. Res. *15*, 225–247.

Woods, D.L., and Alain, C. (2009). Functional imaging of human auditory cortex. Curr. Opin. Otolaryngol. Head Neck Surg. *17*, 407–411.

Woods, D.L., Herron, T.J., Cate, A.D., Yund, E.W., Stecker, G.C., Rinne, T., and Kang, X. (2010). Functional properties of human auditory cortical fields. Front. Syst. Neurosci. *4*, 155.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Code for reproducing figures and analysis	This paper	Zenodo Data: https://doi.org/10.5281/zenodo.4994665
Bezier Curves	Dalca et al., 2011	https://github.com/adalca/bezier
Other		
Human patient participants recruited from local area (see Table S2).	This paper	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Edward Chang (edward.chang@ucsf.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data that support the findings of this study are available on request from the lead contact. The data are not publicly available because they could compromise research participant privacy and consent.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the Key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The University of California, San Francisco Institutional Review Board approved all procedures, and all patients provided written informed consent to participate.

We acquired electrophysiological data from 9 patients (8 M/1 F, mean ± stdev. age: 32 ± 12 years) undergoing left hemisphere insular or opercular tumor resection (N = 5, acute intraoperative setting) or phase II monitoring for intractable epilepsy (N = 4). In patients with tumors, the location of their tumors near eloquent cortex necessitated dissection of the Sylvian fissure, which thereby provided access to the temporal plane. In epilepsy patients, temporal plane access was clinically indicated for seizure monitoring (seizure onset zone thought to originate from within Sylvian fissure). In 9 patients, we acquired simultaneous neural recordings directly from the temporal plane and the lateral surface of the brain (including the superior temporal gyrus and middle temporal gyrus). In 7 patients, 32 channel (8 × 4) or 64 channel (8 × 8) grids with 4mm center-to-center spacing and 1.17 mm diameter exposed contact lateral grids (Integra or AdTech) were placed on the temporal plane. In 8 patients, recordings from the lateral surface were acquired using grids with identical specifications (4-mm spacing and 1.17 mm diameter) but either 256 channels, 64 channels, or 32 channels total. In one patient, bilateral stereo EEG depth electrodes provided coverage of sites in Heschl's gyrus. Patient demographics and the specifics of grid versus depth recording coverage are detailed in Table S2. In addition to the 9 participants included in functional analyses, we also performed electrocortical stimulation in an additional 4 participants (see Electrocortical stimulation).

METHOD DETAILS

Neural recordings

We acquired electrophysiological recordings at a sampling rate of 3051.8 Hz using a 256-channel PZ2 amplifier or 512-channel PZ5 amplifier connected to an RZ2 digital acquisition system (Tucker-Davis Technologies, Alachua, FL, USA). The local field potential was recorded from each electrode, notch-filtered at 60 Hz and harmonics (120 Hz and 180 Hz) to reduce line-noise related artifacts, and re-referenced to the common average across channels sharing the same connector to the preamplifier (Cheung et al., 2016). For

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

STRF and clustering analyses, signals were bandpass filtered in the high gamma range (70-150 Hz) using the log-analytic amplitude of the Hilbert transform at 8 logarithmically-spaced center frequency bands within this range. We then took the first principal component across these 8 bands to extract stimulus-related neural activity (Edwards et al., 2009; Moses et al., 2016; Ray and Maunsell, 2011). Signals were subsequently downsampled to 100 Hz, then z-scored relative to the mean and standard deviation of activity across a recording block.

Electrode localization

For recordings that were performed intraoperatively (N = 5 participants), no CT scan was available, so we localized electrodes on each individual's brain using intraoperative photographs of grid placement. For recordings performed in a chronic in-patient setting (N = 4 participants), we co-registered the preoperative T1 to a postoperative CT scan and localized electrodes using in house software (Hamilton et al., 2017). Pial surface reconstructions were created from preoperative T1 MRI scans using Freesurfer. For visualization of 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 (Fischl et al., 1999). 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.

Stimuli

Speech stimuli

Participants listened passively to 499 sentences taken from the Texas Instruments Massachussetts Institute of Technology (TIMIT) acoustic-phonetic corpus (Garofolo et al., 1993) (286 males/116 female talkers from different regions of the United States of America). Each sentence was repeated once with 0.4 s of silence in between each sentence. In addition, a subset of 10 sentences (5 male, 5 female talkers) was repeated 10 times. Presentation of these stimuli was controlled using custom MATLAB software on a Windows laptop, and played through free-field speakers (Logitech). Sentences were presented in pseudo-random order. All N = 9 participants heard these stimuli.

Tone stimuli

A subset of participants (N = 5) also listened to pure tone stimuli, generated as 80 mel-spaced frequencies from 74.5 Hz to 8 kHz (to match the spectrogram representations of the sentence stimuli). Each sine wave pure tone was 50 ms in duration with a 5-ms cosine ramp at the beginning and end of the tone. Pure tones were played at 3 intensity levels at 10 dB spacing, with the lowest intensity calibrated to be minimally audible in the hospital room. Each pure tone frequency/intensity pair was repeated 3 times, with jittered inter-stimulus intervals to minimize predictability of the stimulus (range 0.28 s minimum ISI – 0.5 s maximum ISI).

Latency analysis

To calculate the latencies of responses to speech, we took the high gamma response to individual sentences, aligned them by first phoneme, then calculated the time of the maximum derivative of the average sentence response. Thus, response latencies in Figure 2 were not reliant on the receptive field or linear modeling analyses, but instead took into account the rise time of responses to TIMIT sentences.

Pure tone receptive fields

To calculate the pure tone receptive field, we first took the high gamma signals time-locked to the onset of each tone, and constructed a post-stimulus time histogram (PSTH) from 0 (tone onset) to 500 ms, collapsing across repetitions. Classical receptive fields were constructed by calculating the average high gamma response to each frequency-intensity pair in a window defined by the peak of the PSTH. The magnitude of the pure tone response was calculated by collapsing across all frequency-intensity pairs to get the maximum Z-scored high gamma response for each electrode. As a proxy for ''clean''/significant receptive field tuning, we created a binary mask (3 intensity bins x 80 frequencies) for each receptive field using the normalized tuning curve that was then rescaled from 0 to 3 (for the 3 intensity bins), and rounded to the nearest integer value. These amplitudes were then used to create a mask of 1 s inside the receptive field (for all frequencies where the tuning curve amplitude met a given intensity value bin). This procedure effectively creates a matrix of NaNs for frequencies in the ''background'' (outside of the classical receptive field) and a (typically V-shaped) set of 1 s for frequencies inside the receptive field. To determine whether sites showed significant tuning in this way, we calculated the amplitude of responses to tones of each frequency and intensity outside the receptive field and compared them to responses to tones at each specific frequency and intensity inside the receptive field. We then used a Wilcoxon rank sum test to calculate the difference in mean response for tones inside versus outside the receptive field. Receptive fields with ''significant'' tuning are shown with red axes in Figures S3 and S4, where it is clear that the response inside the tuning curve has a higher amplitude (stronger purple values) compared to tones outside the receptive field. For non-significant sites, overall amplitudes to tones may be greater than silence, but show no strong tuning to a given frequency range.

To calculate whether responses to tones were significantly greater than silence, we computed a Wilcoxon rank sum test to assess the difference in activity for 50 ms of pre-trial silence and tone-evoked activity from 0-250 ms following tone onset. P values were

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

corrected for multiple comparisons using the Bonferroni method, where the number of comparisons was equal to the number of electrodes being evaluated.

Nonlinear maximally informative dimensions

Because of potential nonlinear relationship between neural and natural stimuli, we fit nonlinear spectrotemporal models using maximally informative dimensions and an analysis incorporating two dimensions. This analysis has been previously reported by others (Sharpee et al., 2004; Atencio et al., 2008) for different sensory stimuli and by our group specifically for spectro-temporal receptive fields estimated from STG ECoG recordings (Hullett et al., 2016). In brief, we use the 80-band mel frequency spectrogram as our stimulus input (identical for the ridge STRF procedure). MIDs were calculated using a gradient ascent procedure in which we maximize the Kullback-Leibler divergence between the distribution of projection values of the stimulus onto the MID filters and the distribution of the projection values weighted by the magnitude of the response. We fit the MID filters using the same training set and test set used for the ridge STRF models. MID filters included the same number of time delays as our original STRF analysis (60 total, up to 0.6 s).

To evaluate the performance of these models and incorporating additional contributions from nonlinearities in the stimulus-response relationship, we calculated the mutual information between stimulus and response for each MID model (MID1, MID2, and MID1+2 jointly) and compared it to a spike triggered average and our ridge regression STRF. This mutual information calculation was performed for each filter v (MID1, MID2, or ridge STRF) as follows:

$$I(v) = \int dx P(x|resp) \log_2 \left[\frac{P(x|resp)}{P(x)} \right]$$

Where x is the projection of the stimulus onto the MID1, MID2, or STRF filter, and *resp* is the high gamma response. To calculate the mutual information for the joint MID1+2, we used the following equation:

$$I(MID_1, MID_2) = \iint dx_1 dx_2 P(x_1, x_2|resp) \log_2 \left[\frac{P(x_1, x_2|resp)}{P(x_1, x_2)} \right]$$

Where x_1 is the projection onto MID1 and x_2 is the projection onto MID2. Each of these mutual information calculations was performed for all electrodes separately. We computed MID estimates using 4 jack-knife estimates. To compute the information values, we then averaged across these estimates while taking into account the fact that MIDs are defined only up to the sign. That is, each information value was estimated as follows:

$$\bar{I}(MID1) = \left(I(v_1) + sign(I(v_1)I(v_3)) * I(v_2) + sign(I(v_1)I(v_3)) * I(v_3) + sign(I(v_1)I(v_4)) * v_4 \right) \Big/ 4$$

with the same procedure for MID2 and the joint MID12.

Linear receptive field analysis

To uncover tuning for spectrotemporal and acoustic-phonetic features in individual electrode sites, we also fit linear receptive field models (Aertsen and Johannesma, 1981; Theunissen et al., 2001) of the form:

$$\hat{x}(t) = x_0 + \sum_f \sum_{\tau=0}^{T} \beta(\tau, f) S(f, t-\tau)$$

Where x is the neural activity recorded at a single electrode, $\beta(\tau, f)$ contains the regression weights for each feature f at time lag τ, and S is the stimulus representation for feature f at time $t - \tau$. We also fit an intercept x_0 for each electrode to allow for differences in baseline activity. We included delays of up to 600 ms to model longer latency responses that may be observed in mid- to anterior STG sites (Hamilton et al., 2018). β weights were fit using ridge regression. The ridge parameter was estimated using a bootstrap procedure in which the training set was randomly divided into 80% prediction and 20% ridge testing sets. The ridge parameter was chosen from a range of 30 log-spaced values from 10^{-2} to 10^{7}, as well as a ridge parameter of 0 (no regularization). The final value was chosen as the parameter that gave the best average performance across electrodes as assessed by correlation between the predicted and ridge test set performance. Once an optimal regularization parameter was chosen, the model was then trained on the complete training set. The final performance of the model was computed on a final held out set not included in the ridge parameter selection. Performance was measured as the correlation between the predicted response on the model and the actual high gamma measured for sentences in the test set.

We estimated models using a mixture of stimulus representations, including the same mel-band spectrogram used for MID models, sentence onset features (Hamilton et al., 2018), phonological features (Mesgarani et al., 2014), absolute and relative pitch features (Tang et al., 2017), and peakRate, calculated as the maximum change in the derivative of the acoustic envelope (Oganian and Chang, 2019). These features are described below and were used in combination to estimate the unique amount of variance explained by each feature (for example, a full feature model included sentence onset, peakRate, phonological features, absolute and relative pitch features).

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

Mel-band spectrogram

The spectrograms of each sentence were calculated using a mel-band auditory filterbank of 80 filters with center frequencies from approximately 75 Hz to 8 kHz. This frequency decomposition is thought to reflect the filtering performed by the human auditory system (Slaney, 1998), and has been used extensively in our previous work to describe spectrotemporal tuning within the STG (Hamilton et al., 2018; Mesgarani et al., 2014).

Sentence onset feature

The sentence onset feature consisted of a binary vector of values, with a 1 at the onset of the first sample of the first phoneme of each sentence, and 0 elsewhere.

Peak rate features

Peak rate was calculated as local peaks in the derivative of the amplitude envelope of speech (Oganian and Chang, 2019). First, we extracted the amplitude envelope of speech using the specific loudness method by Schotola (1984). This method first decomposes the speech signal into critical bands based on the Bark scale. Signals were square-rectified within each filter bank, bandpass filtered between 1 and 10 Hz, downsampled to 100 Hz, and then averaged across frequency bands to get the envelope. We then calculated the derivative of this envelope, and extracted local peaks in this derivative to create a sparse time series of "peakRate" features.

Phonological features

Phonological features consisted of binary phonological features used in our previous work (Hamilton et al., 2018; Mesgarani et al., 2014). These features describe single phonemes as a combination of voicing, place and manner of articulation features. They are a reduced representation of the speech sound signal that better captures responses to speech in non-primary auditory cortex (Mesgarani et al., 2014). As with sentence onset features, these matrices include a 1 at the onset of each phonological feature, and a 0 elsewhere. Features included sonorant, obstruent, voiced, back, front, low, high, dorsal, coronal, labial, syllabic, plosive, fricative, and nasal.

Absolute pitch features

Absolute pitch was calculated using procedures identical to Tang et al. (2017). In brief, the fundamental frequency (F0) was calculated using an automated autocorrelation method in Praat, and manually corrected for doubling or halving errors. Absolute pitch was calculated as the natural logarithm of pitch values in Hz. We then created a binary feature matrix by discretizing these pitch values into 10 bins, equally spaced from the 2.5 to the 97.5 percentile values.

Relative pitch features

Relative pitch was also calculated using procedures identical to Tang et al. (2017). The fundamental frequency (F0) was extracted as described above for absolute pitch. Relative pitch was calculated by z-scoring the log-F0 absolute pitch values within each sentence (within speaker), such that values were high or low relative to the average pitch of the speaker of that sentence. These normalized values were then discretized into 10 bins, as above. To calculate relative pitch change, we took the derivative of the z-scored log-F0 relative pitch values, and then discretized this pitch derivative curve.

Variance partitioning to determine unique variance explained

To calculate the unique variance explained by a given set of features, we calculated R^2 values for a reduced model that did not include these features of interest but did include other confounding features. We then compared these to R^2 values to an extended model including the features of interest and the features from the reduced model. For example, to calculate the unique variance explained by phonetic features and peakRate together, we compared a model including onsets, phonetic features and peakRate to a model including onsets only. We used the following combinations of models to determine unique R^2 values:

$$R^2_{\text{unique onset}} = R^2_{\text{onset + phnfeat + peakRate}} - R^2_{\textit{phnfeat + peakRate}}$$

$$R^2_{\text{unique peak rate}} = R^2_{\text{onset + phnfeat + peakRate}} - R^2_{\textit{onset + phnfeat}}$$

$$R^2_{\text{unique phnfeat}} = R^2_{\text{onset + phnfeat + peakRate}} - R^2_{\textit{onset + peakRate}}$$

$$R^2_{\text{unique absF0}} = R^2_{\text{onset + phnfeat + peakRate + absF0 + relF0 + }\Delta\text{relF0}} - R^2_{\textit{onset + phnfeat + peakRate + relF0 + }\Delta\textit{relF0}}$$

$$R^2_{\text{unique relF0 + }\Delta\text{relF0}} = R^2_{\text{onset + phnfeat + peakRate + absF0 + relF0 + }\Delta\text{relF0}} - R^2_{\textit{onset + phnfeat + peakRate + absF0}}$$

To determine whether the addition of a feature or set of features resulted in a statistically significant increase in R^2, we used permutation testing. In these tests, the added feature labels were shuffled in 2.4 s chunks over time to create a shuffled distribution of feature values. 2.4 s was chosen as 4 times the length of the delay period of 0.6 s. We then calculated the change in R^2 that was

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

observed by adding these (dummy) variables to the full model. Each feature shuffle was performed 1000 times, allowing us to determine significant differences up to a threshold of $p < 0.001$.

Calculating frequency tuning curves for speech

To calculate frequency tuning curves for speech and compare them to pure tone responses, we took the speech-derived STRF (using the mel-band spectrogram), summed across all time delays, and smoothed the resulting frequency tuning curve using a 3rd-order Savitzky-Golay filter with a 31-point window. Pure tone tuning curves were similarly collapsed across intensities and smoothed using the same procedure. Example frequency tuning curves for speech and pure tones are shown in Figure 3E. The number of peaks in each tuning curve was calculated using the findpeaks function of the MATLAB signal processing toolbox on each tuning curve normalized to its maximum, with a minimum peak prominence of 0.5, and a minimum peak height of 0.5.

To calculate the correlation between pure tone and frequency tuning (Figure 3H), we used the frequency tuning curves defined above, and calculated the Pearson correlation between these tuning curves for each electrode. In Figure 3H, only correlations where at least one stimulus type (pure tones or speech) elicited a significant response were included.

Unsupervised clustering of LFP time series

We performed unsupervised clustering of electrodes in the superior temporal gyrus, middle temporal gyrus, and temporal plane (planum temporale, Heschl's gyrus, and planum polare) across all participants using convex non-negative matrix factorization (cNMF), similar to Hamilton et al. (2018). In brief, cNMF uses an iterative decomposition to estimate the neural high gamma time series X [n time points x p electrodes], according to the following equation:

$$X \approx \widehat{X} = FG^T$$

Where

$$F = XW$$

The G matrix [p electrodes x k clusters] represents the spatial weight of an electrode in a given cluster, and W [p electrodes x k clusters] represents the weights applied to the electrode time series. In order to look for canonical response types across the whole auditory cortex, we performed NMF on the average responses across the sentence stimuli that were heard by all participants. To choose the number of basis functions, we calculated the percent variance explained by each number of basis functions from k = 1 to k = 16, and chose the number of clusters at the elbow of the percent variance curve (Figure S6), resulting in a final number of k = 4 clusters. These clusters represented approximately 89% of the variance in auditory cortical responses.

Although the NMF weights *G* are continuous and allow electrodes to belong to more than one cluster, for some analyses we assigned electrodes to their "best" NMF cluster. This was performed by sorting the NMF spatial weights within each electrode, and finding electrodes where the maximum cluster weight on the 4 clusters was greater than two times the next highest weight, and where the highest spatial weight was at least 0.1 (to avoid assigning electrodes to a cluster where no particular temporal response profile was a good match).

Electrocortical stimulation (ECS)

Patients underwent electrocortical stimulation as part of a clinical protocol for identification of areas critical to sensory and language processing. In order to investigate the causal contribution of the temporal plane and STG to sound and speech perception, we analyzed the effects of ECS on passive perception of environmental sounds and during word repetition in 7 patients (S06, S08, S09, S10, S11, S12, S13), 3 of which overlapped with the participants in previous analyses (S06, S08, S09). Detailed descriptions of sensory perception induced by ECS and effects of ECS on word repetition have been described previously (Leonard et al., 2016).

Stimulation was carried out using a Natus clinical stimulator, with typical settings: bipolar, biphasic stimulation at 50Hz for 1-2 s. The stimulation current was determined individually at a level between 2 and 5mA, and set at a level that induced a sound percept or interfered with speech perception. Stimulation of each site started at 2mA and was gradually increased to 5 mA until a perceptual effect was elicited. If stimulation at 5mA did not induce any effect, the site was marked as negative. For each combination of stimulation parameters and each task, if stimulation at a given cortical site caused an effect, the site was tested at least two more times, some up to 25 times depending on individual patients. If greater than 50% of the stimulation trials caused errors, that site was demarcated. Overall, each site was stimulated 5-25 times.

To further investigate the effects of stimulation parameters on perceived sounds, an extended stimulation protocol was used in a subset of patients and sites that included variation in stimulation frequency (down to 10 Hz, up to 100 Hz) and stimulation duration (up to 5 s).

To test for induced sound perception, participants were asked to report if they heard any new sounds and to describe the sound quality. For repetition, stimulation was timed with the auditory presentation of single words and participants were asked to report which word they heard. Importantly, in both tasks, participants were unable to detect when electrical stimulation was applied. Subjective reports for each stimulation site are given in Table S3. Individual stimulation sites are shown in Figure S5.

Please cite this article in press as: Hamilton et al., Parallel and distributed encoding of speech across human auditory cortex, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.019

Thermoablation case study

One patient (33-year-old male) underwent thermocoagulation to treat intractable auditory seizures that were localized to the HG. Procedures were carried out at the Montreal Neurological Institute according to procedures described by Bourdillon et al. (2017), where ablation was carried out along indwelling electrodes along the axis of HG. The patient's speech and language were assessed throughout the surgery using standard clinical procedures.

QUANTIFICATION AND STATISTICAL ANALYSIS

We used nonparametric statistical tests including Kruskal Wallis nonparametric ANOVAs followed by signed rank tests for paired data, or rank sum tests for unpaired data as our data were not normally distributed. In some cases, bootstrap or permutation tests were used.

CellPress

Supplemental figures

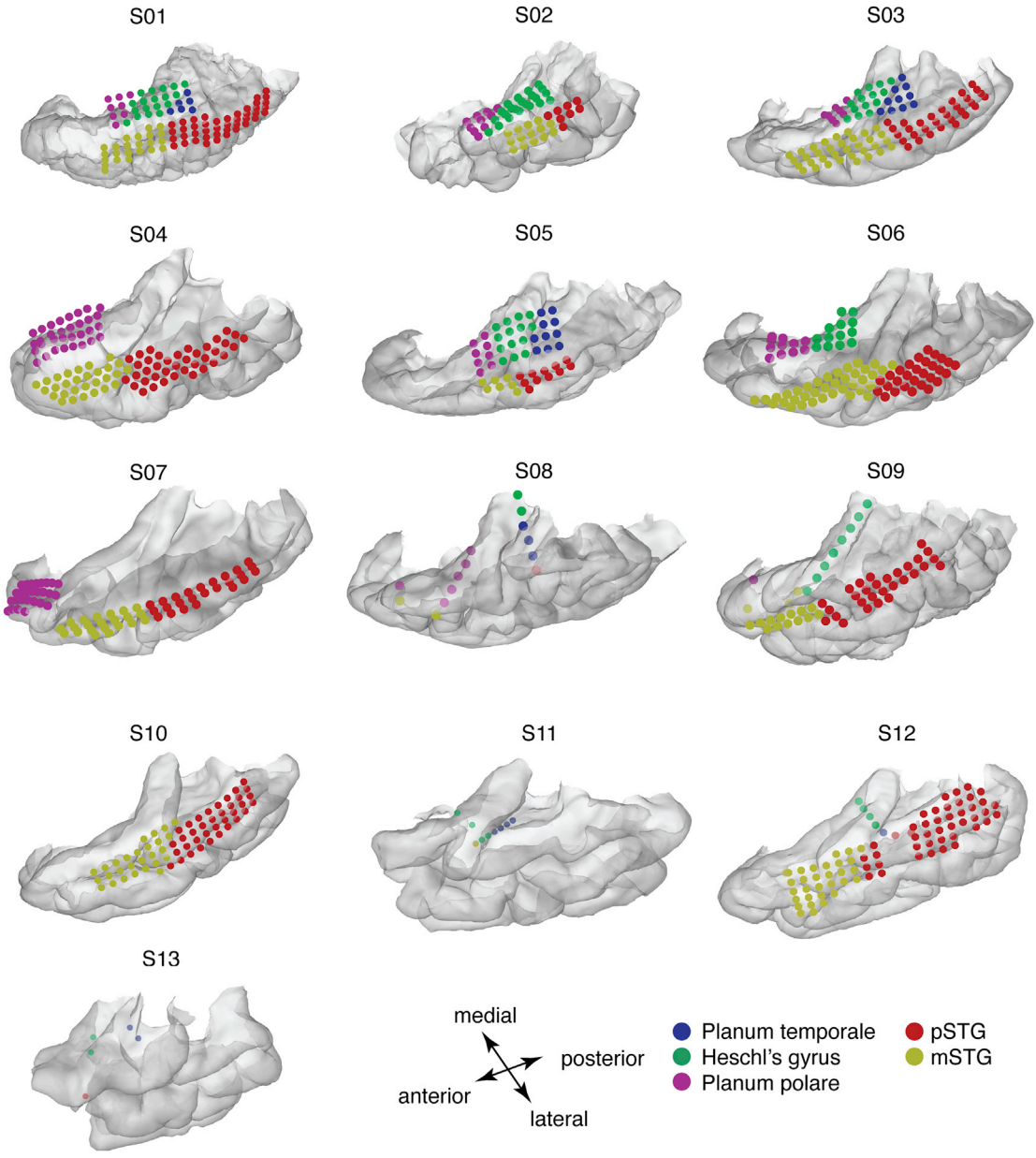


Figure S1. Electrodes from all participants across all auditory areas, shown on each participant's left hemisphere temporal lobe reconstruction, related to Figure 1

Electrodes are colored according to anatomical region. S01 - S09 were included in receptive field mapping and NMF analyses. S10 - S13 were included in stimulation mapping only.

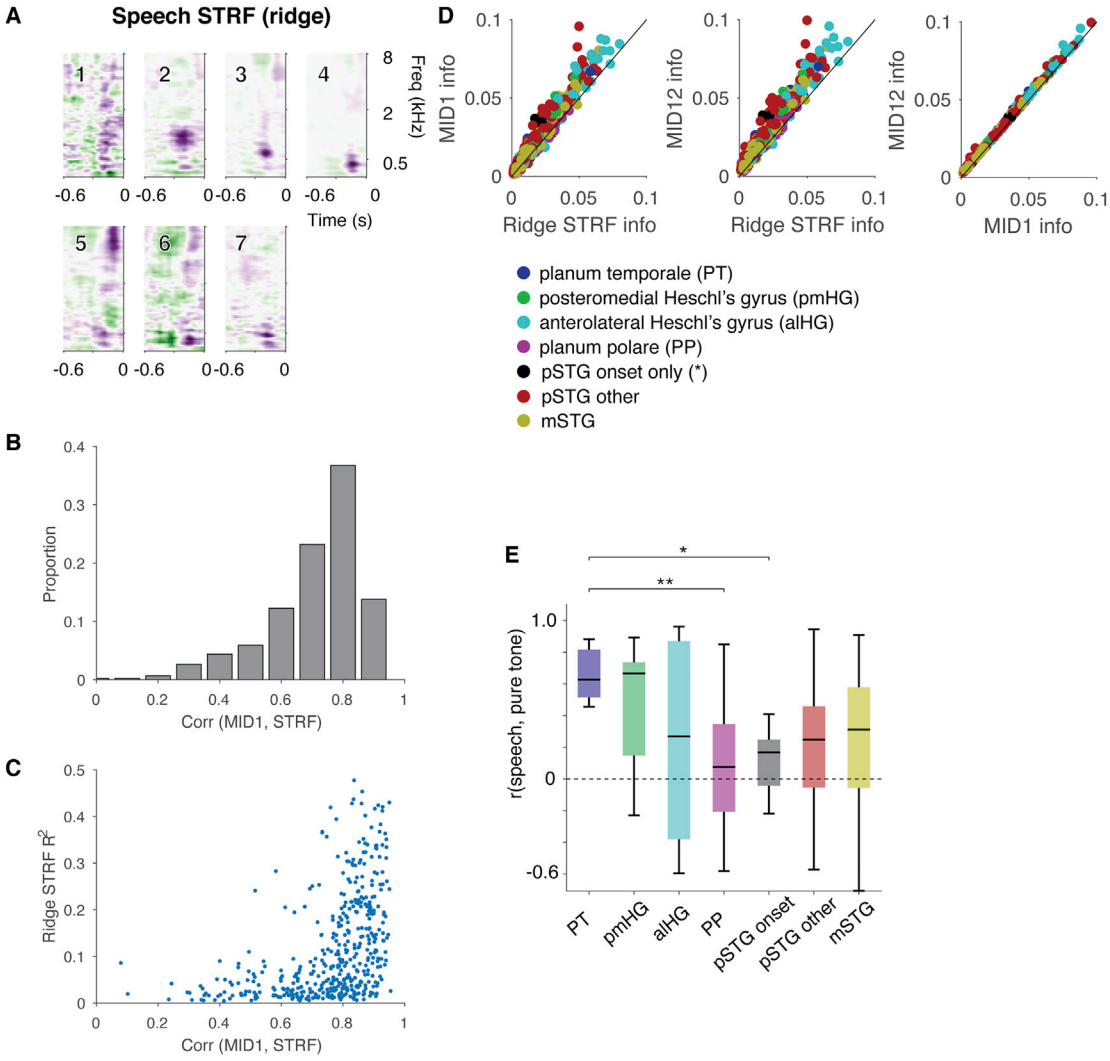


Figure S2. Comparison between ridge STRF and nonlinear MID models, related to Figure 3

A. Example ridge STRFs for same electrodes shown in Figure 2D. Overall, spectrotemporal tuning was highly similar for these and the MID1 STRFs. B. Correlation (r) between MID1 and ridge STRF filters showing degree of similarity between the learned filters. C. Correlations between MID1 and STRF are higher for STRFs with greater explained variance (STRF R^2). D. Mutual information comparisons between the ridge STRF model and MID models with 1 and 2 dimensions. Overall, adding a second MID dimension adds very little additional information. The MID model significantly outperforms the ridge STRF, but results on tuning were qualitatively and quantitatively similar. E. Correlation between MID1 filter spectral tuning and pure tone tuning (compare to Figure 2H). The pattern of results was similar for ridge STRFs as compared to MIDs.

(legend on next page)

Figure S3. Pure-tone receptive fields recorded from individual electrodes in PT, HG, PP, posterior STG (pSTG), and mid-to-anterior STG (mSTG), related to Figure 3

Blue trace shows computed tuning curve based on smoothed receptive field data. Red axes indicate electrodes for which within receptive field responses were significantly larger (Bonferroni corrected $p < 0.05$) than outside receptive field responses (i.e., inside or outside the blue tuning curve). This provides a proxy for "clean" receptive fields. Overall, most sites in PT and HG showed strong, classical V-shaped tuning for pure tones. PP sites tended to show multiple peaks or weak tuning. Sites in pSTG showed some pure tone tuning, but the magnitude of responses was overall lower compared to HG and PT. While some narrow-band, v-shaped tuning curves were observed in pSTG, many responses tended to be broader or multi-peaked. In mid-to-anterior STG, fewer sites with classical RFs were observed.

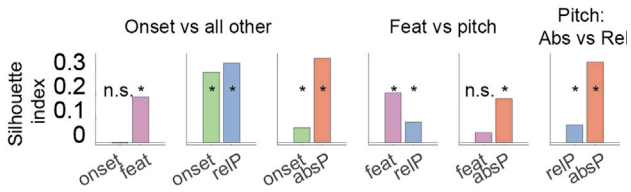


Figure S4. Anatomical clustering of electrodes with different feature-encoding profiles, related to Figure 4

(A) Silhouette index for clustering of different encoding populations. Onsets cluster in pSTG, Phonetic features and relative pitch in mid-STG, absolute pitch on the temporal plane (p = 0.1, *p < 0.01, median permutation test). Onset-encoding electrodes were confined to pSTG, with a clear separation from pitch encoding electrodes but not feature -encoding electrodes, which were also present in pSTG and extended anteriorly into midSTG (onset cluster pairwise permutation: $p_{relPitch}$ = 0.001, $p_{absPitch}$ = 0.001, p_{feat} = 0.14). Absolute pitch encoding populations were concentrated on the temporal plane and were anatomically separate from all other features (absolute pitch cluster pairwise permutation p_{ons} = 0.001 p_{feat} = 0.001, $p_{relPitch}$ = 0.01), whereas feature and relative pitch encoding were both localized to middle STG with relative pitch encoding extending more anteriorly than feature encoding (pairwise permutation p_{feat} = 0.001, $p_{relPitch}$ = 0.008). Notably, even though phonological features and relative pitch encoding were located within the same broad anatomical area, individual electrodes tended to encode one set of features at a time rather than the combination of the two. For example, 83% of electrodes that had significant feature encoding did not have significant relative pitch encoding (out of total 302 electrodes with significant feature encoding). For relative pitch electrodes, the same trend held, though less strongly–51% of relative pitch encoding electrodes did not significantly encode features (out of total 105 electrodes with significant relative pitch encoding).

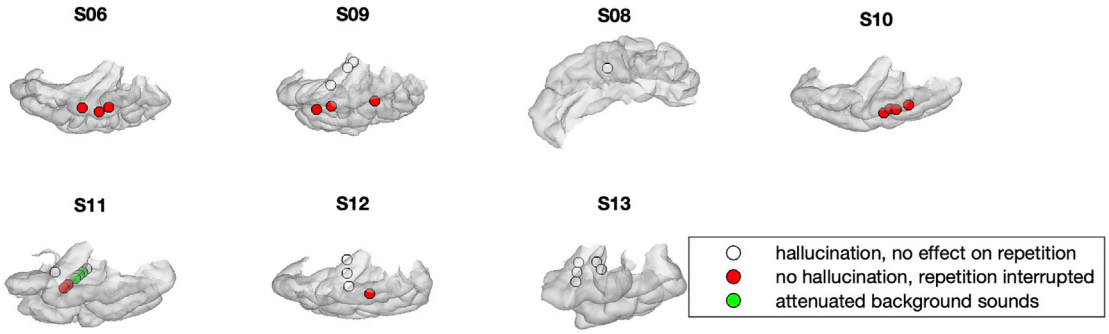


Figure S5. Single-subject reconstructions of stimulation electrodes, related to Figure 5

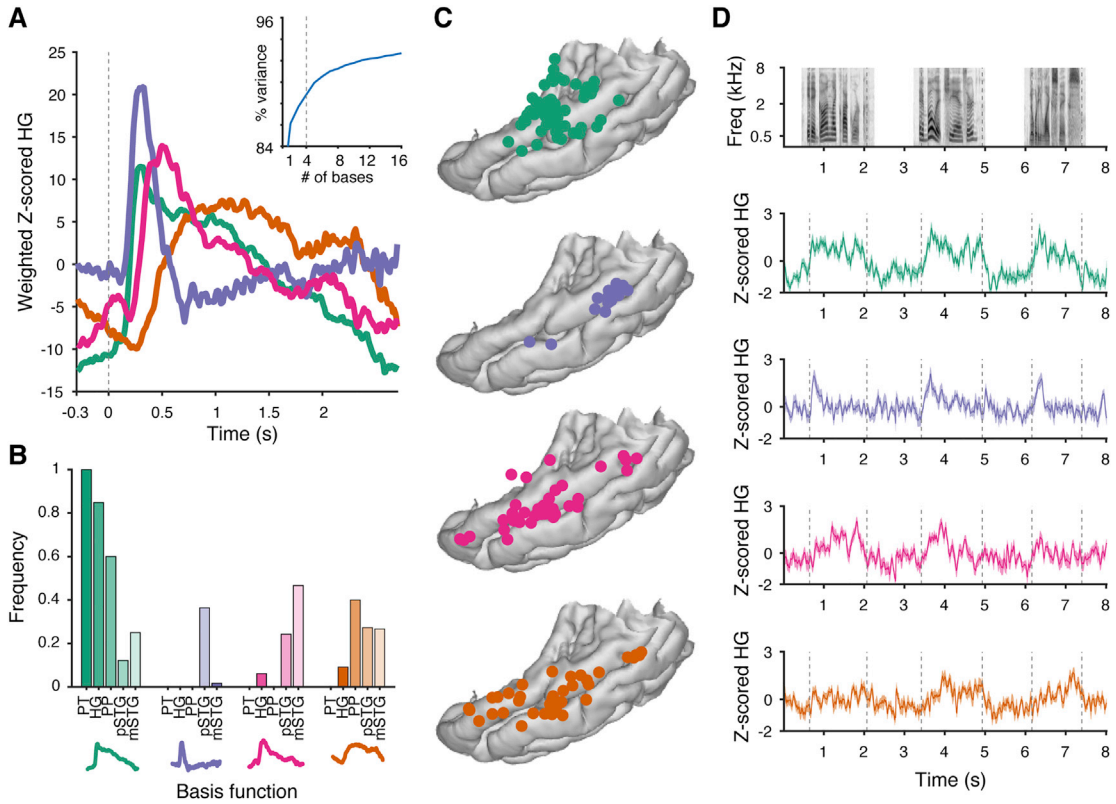


Figure S6. Identification of onset zone of the pSTG via non-negative matrix factorization, related to Figure 1

(A) Canonical temporal response profiles to speech as measured through unsupervised cNMF. Auditory cortical responses to speech could be decomposed into four classes that explain 89% of the variance in average responses to sentences (inset). These temporal response profiles included (1) a fast (short-latency), mixed onset+sustained response (green), (2) a fast (short-latency), onset-only response (purple), (3) a slower, mixed onset+sustained response, and (4) a very slow sustained response. (B) Proportion of sites with each temporal response type within each area, e.g., sum across all bars for each area sum to 1 or less. (C) Anatomical distribution of sites in each of the four clusters. (D) Example single electrode traces from clusters 1 – 4. The purple cluster was identified as the onset-only zone shown in other figures in the manuscript.