Published in final edited form as:
Cell Rep. 2018 August 21; 24(8): 2051–2062.e2. doi:10.1016/j.celrep.2018.07.076.

Joint Representation of Spatial and Phonetic Features in the Human Core Auditory Cortex

Prachi Patel[1,2], **Laura K. Long**[1,3], **Jose L. Herrero**[4,5], **Ashesh D. Mehta**[4,5], and **Nima Mesgarani**[1,2,6,*]

[1]Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA

[2]Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

[3]Doctoral Program in Neurobiology and Behavior, Columbia University, New York, NY 10027, USA

[4]Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, New York City, NY 11549, USA

[5]The Feinstein Institute for Medical Research, New York City, NY 11030, USA

[6]Lead Contact

SUMMARY

The human auditory cortex simultaneously processes speech and determines the location of a speaker in space. Neuroimaging studies in humans have implicated core auditory areas in processing the spectrotemporal and the spatial content of sound; however, how these features are represented together is unclear. We recorded directly from human subjects implanted bilaterally with depth electrodes in core auditory areas as they listened to speech from different directions. We found local and joint selectivity to spatial and spectrotemporal speech features, where the spatial and spectrotemporal features are organized independently of each other. This representation enables successful decoding of both spatial and phonetic information. Furthermore, we found that the location of the speaker does not change the spectrotemporal tuning of the electrodes but, rather, modulates their mean response level. Our findings contribute to defining the functional organization of responses in the human auditory cortex, with implications for more accurate neurophysiological models of speech processing.

[*]Correspondence: nima@ee.columbia.edu.

AUTHOR CONTRIBUTIONS

N.M. and P.P. designed the experiment. J.L.H., L.K.L., P.P., N.M., and A.D.M. recorded the data. P.P., L.K.L., and N.M. analyzed the data. P.P. and N.M. wrote the manuscript. All authors commented on the manuscript.

DATA AND SOFTWARE AVAILABILITY

The data that support the findings are available on request from the corresponding author. The codes for performing phoneme analysis, calculating high-gamma envelope, and predicting the neural responses are available at http://naplab.ee.columbia.edu/naplib.html (Khalighinejad et al., 2017b).

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.07.076.

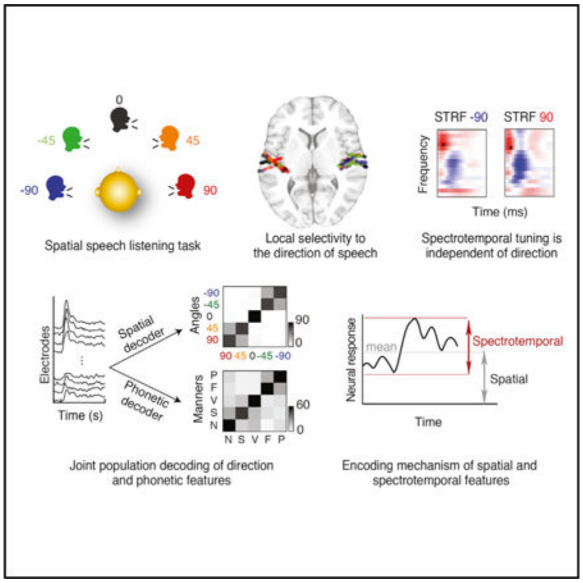
DECLARATION OF INTERESTS

The authors declare no competing interests.

In Brief

Using invasive recordings from the human core auditory cortex, Patel et al. show direct evidence for neural selectivity to the direction of the speaker relative to the listener. The direction of speech is jointly and independently represented with the speech phonetic features in the same region. Speech direction modulates the mean of the neural response to spectrotemporal features of speech.

Graphical Abstract


0
-45
45
-90
90
Spatial speech listening task
Local selectivity to the direction of speech
STRF -90
STRF 90
Frequency
Time (ms)
Spectrotemporal tuning is independent of direction
Electrodes
Time (s)
Spatial decoder
Phonetic decoder
Angles
-90
-45
0
45
90
90 45 0 -45 -90
90
0
Manners
P
F
V
S
N
N S V F P
60
0
Joint population decoding of direction and phonetic features
Neural response
mean
Spectrotemporal
Spatial
Time
Encoding mechanism of spatial and spectrotemporal features

INTRODUCTION

Speech is an important human communication signal that carries a great deal of information about the speaker, including the intended message and the location of the speaker in space relative to the listener. A major function of the human auditory cortex is to encode and link spectrotemporal and spatial aspects of the speech signal, as evident by data showing that speech comprehension improves when spatial cues are available (Kidd et al., 2005).

Neural processing of spatial sound in the mammalian auditory cortex has been studied extensively, especially in non-human mammals, where neural activity can be measured directly. Cooling (Malhotra et al., 2004) and lesion (Jenkins and Merzenich, 1984; Kavanagh and Kelly, 1987) studies of the mammalian auditory cortex have shown a necessary and selective role for various auditory cortical fields in the perception of spatial cues. Mammalian studies of response properties in auditory cortical neurons show clear spatial receptive fields, where the majority of neurons in each hemisphere of the brain are tuned preferentially to the contralateral field in space (Middlebrooks and Pettigrew, 1981). There is also strong evidence that mammalian auditory cortical neurons jointly encode multiple stimulus features, including pitch, timbre, and location (Bizley et al., 2009; Walker et al., 2011). Extending these findings to the human auditory cortex remains challenging because there is no dominant model of the anatomical and functional organization of human auditory cortical fields. The auditory cortex in humans is significantly different from non-human primates, even at a macro scale (Hackett et al., 2001). For example, compared with the monkey, the Sylvian fissure and planum temporale are larger on the left side (Geschwind and Levitsky, 1968; Yeni-Komshian and Benson, 1976), there are additional gyri that exist only in the human brain, and the human auditory cortex is specialized for speech processing (Belin et al., 2000; Norman-Haignere et al., 2015). As a result, the functional organization of the human core auditory cortex is still a matter of scientific debate (Humphries etal., 2010; Leaver and Rauschecker, 2016; Moerel etal., 2014).

Because of the difficulty in obtaining direct measurements of brain activity in humans, most of what we know about the neural basis of sound localization in the human auditory cortex is based on non-invasive neuroimaging and lesion studies (Ahveninen et al., 2006; Zatorre and Penhune, 2001; Zimmer et al., 2006). Because non-invasive methods have limited spatial and temporal resolution, the majority of studies have focused mainly on determining which brain areas are involved in processing spatial information rather than characterizing how they represent speech with spatial cues (Alain et al., 2001; Barrett and Hall, 2006). For example, the human Heschl’s gyrus has been shown to respond to the sound location (Junius et al., 2007; Zatorre and Penhune, 2001) and the spectrotemporal and phonetic features (Arsenault and Buchsbaum, 2015). Despite these findings, many questions regarding how the core auditory areas represent spatial and spectrotemporal features together remain unanswered. Specifically, how does the core human auditory cortex encode these aspects of speech to support downstream extraction of spatial “where” and phonetic “what” information (Arnott et al., 2004)? Is there local selectivity to speaker direction in human core auditory areas? And finally, which properties of the neural response correlate with spectrotemporal and spatial features of sound?

To tease apart the encoding properties of spectrotemporal and spatial cues in the human core auditory cortex (Brodmann Area 41 and 42), we directly measured neural activity during a speech perception task with spatial cues using bilateral depth electrodes implanted in core auditory areas of neurosurgical patients. We aim to shed light on the encoding of spatial features and to specify the relationship between the encoding of spatial, spectrotemporal, and phonetic features of speech at local and population levels. We use our findings to construct an encoding model to tease apart how the neural activity changes in response to spatial and spectrotemporal features. Our results further our understanding of how natural speech with spatial cues is represented and processed in the human auditory cortex, with significant implications for more complete models of speech processing in the human brain.

RESULTS

We recorded from invasive, bilateral, high-density depth electrodes (PMT Corporation) implanted in the perisylvian region of five epilepsy patients, providing partial coverage of Heschl’s gyrus (HG), the planum temporale (PT), and the superior temporal gyrus (STG) (Figure 1A). The subjects listened to speech stimuli arriving from five different directions in the horizontal plane (−90°, −45°, 0°, 45°, and 90°), simulated using the head-related transfer function (HRTF) of an average-sized head (Figure 1B; STAR Methods). To ensure that the subjects were engaged in the task and could identify the direction of the speaker, the subjects were asked to report the location of the speaker when the audio paused at periodic intervals;

subjects were attentive and could correctly report the direction of the speaker (Figure S2A). All subjects were fluent speakers of American English. Four subjects were left hemisphere language-dominant, and one subject had bilateral language biased toward the left hemisphere.

Direction Selectivity of Responses in the Core Auditory Cortex

To investigate how individual electrodes in the human core auditory cortex respond to speech arriving from different directions in space, we first identified the electrodes that had a significant response to speech compared with silence ($p < 0.001$, t test) and restricted our analysis to this subset. This test results in 201 electrodes, in areas including Heschl's gyrus and sulcus (Heschl's gyrus, 88 electrodes), the STG and sulcus (35 electrodes), and planum temporale (PT, 35 electrodes) (Figure 1A; Destrieux et al., 2010). To examine the effect of the sound direction on response amplitude, we first averaged the envelope of the high-gamma band (70–150 Hz) of each electrode to all speech sounds from the same direction. Figure 1C shows the response of five example electrodes to speech arriving from each of the five directions (−90°, −45°, 0°, 45°, 90°). These example electrodes have diverse response patterns, each tuned most selectively to one direction. Electrode E1 responds most robustly to angle −90°, E2 to −45°, E3 to 0°, E4 to 45°, and E5 to 90°. Moreover, the difference in response to different directions is most robust at sound onset but is sustained over the duration of the sound (Figures S1B and S1C).

To characterize the diversity of spatial selectivity shown by the examples in Figure 1C across the population of electrodes, we first defined an angle selectivity index (ASI) to measure the probabilistic preference of each electrode to the direction of the speaker in space. The ASI for each direction indicates the normalized t value of a t test between the neural responses during silence and during speech uttered from that direction. We found that a majority of speech-responsive electrodes showed selectivity to specific sound directions (Figure S1A; correlation $r = 0.4$, $p < 0.001$).

To investigate the diversity of spatial selectivity patterns across individual electrodes, we performed unsupervised hierarchical clustering of the ASIs separately across electrodes (columns) and angles (rows) (Figure 1D). The horizontal clustering shown in Figure 1D grouped electrodes by the similarity of their ASIs, revealing subgroups of electrodes with similar selectivity for each of the five directions. Most of the electrodes in each brain hemisphere were clustered together because of the similarity of their ASI patterns (blue corresponds to the left hemisphere, and red corresponds to the right hemisphere). Furthermore, we found clusters of electrodes with higher preference for each of the five angles as opposed to the preference for simply left versus right directions in space. The vertical clustering shown in Figure 1D grouped sound angles based on the similarity of the responses they elicited across the population of electrodes, demonstrating that angles closer in space elicited more similar responses.

To further quantify each electrode's most preferred angle, we defined the BA tuning of each electrode as the angle for which the electrode's ASI is maximal. Figure 1E shows the distribution of ASI values for electrodes grouped by their BA tuning in response to speech from each of the five directions. The group of electrodes that have the same BA as the direction of speech presentation responds significantly higher to this direction compared with other electrode groups, demonstrating that the electrodes have significantly preferential tuning for a particular direction. BAs are plotted on an average brain in Figure S1D. Finally, we measured the separability of responses to the five directions by computing the ratio of variance within and between responses to the five directions (f statistics). Figure 1F shows that the separability of responses to the five directions is comparable between the Heschl's gyrus, PT, and STG (Figure 1F, multiple comparisons, Wilcoxon rank-sum test, $p > 0.05$).

Contralateral Tuning of Electrodes

As shown in Figure 1D, the majority of electrodes in the right hemisphere (shown in red) respond to left-sided angles in space (shown in blue) and vice versa. To investigate the degree of this contralateral bias, we compared the average t value of speech versus silence for each electrode when speech was uttered from the right-sided angles (90° and 45°; y axis, Figure 2A) versus the left-sided angles (−90° and −45°; x axis, Figure 2A). Each point in Figure 2A corresponds to one electrode; electrodes above the line have right-sided preference, and electrodes below the line have left-sided preference. The color of the point indicates whether the electrode i located in the left (blue) or right (red) brain hemisphere. The separation of the red and blue points shows a clear contralateral preference in electrode responses to speech. This result also holds when the analysis is broken down by subject (Figure S3).

We further quantified this contralateral preference by defining a contralateral strength index (CSI) as the difference between the t value for left- and right-sided angles. A positive CSI value indicates left-sided preference, whereas a negative CSI value indicates right-sided preference, and the magnitude indicates the degree of preference. The histogram of CSI values in Figure 2B shows a significant difference between electrodes in the right and left hemispheres (absolute difference between the means = 0.251, $p < 0.001$, unpaired t test), further supporting a significant contralateral preference. The CSI values for each electrode are shown on an ICBM152 average brain in Figure 2C. Breaking down our analysis by anatomical area, the Heschl's gyrus, STG, and PT all showed similar contralateral dominance (Figure 2D).

In summary, we found direction selectivity at the level of individual electrodes, with 175 of 201 electrodes (87.06%) tuned contralaterally and 26 of 201 (12.94%) tuned ipsilaterally.

Independent Encoding of Spatial and Spectrotemporal Cues

We showed that a majority of speech-responsive electrodes in human core auditory areas are selective to speech sound from a specific direction in space. We know from previous work that the human core auditory cortex is also selective to spectrotemporal features of sound (Formisano et al., 2003; Humphries et al., 2010; Leaver and Rauschecker, 2016). To study how core areas encode spatial and spectrotemporal cues together, we calculated the spectrotemporal receptive field (STRF) (Theunissen et al., 2001a) of each electrode in response to speech from each direction. To ensure unbiased comparison, the regularization parameters of the STRFs for each electrode were optimized jointly for all directions (STAR Methods). This analysis allows us to examine the effect of the sound direction on each electrode's spectrotemporal tuning. Because we used stereo speech, for each angle and electrode we calculated a left STRF (lSTRF) from the sound presented to the left ear and a right STRF (rSTRF) from the sound presented to the right ear. Because we found a high correlation between lSTRFs and rSTRFs (average r = 0.95, p < 0.001; Figure S4A), we used the average of lSTRFs and rSTRFs to characterize the spectrotemporal tuning of each electrode. For the rest of the STRF analysis, we used the electrodes for which the STRF model could predict the neural responses with prediction correlation values higher than 0.2 (161 of 201 electrodes). Figure 3A shows the STRFs of five neighboring electrodes for all five directions (E1, most medial, to E5, most lateral electrode). We observe a gradual decrease in frequency tuning across the 5 electrodes (see a comparison of the excitatory peak of the STRFs across rows in Figure 3A), consistent with the reported tonotopy in Heschl's gyrus (Formisano et al., 2003; Humphries et al., 2010; Nourski, 2017). However, STRFs from different directions at a single electrode are very similar (see a comparison of columns in Figure 3A). These observations suggest that the tuning properties of individual electrodes do not depend on the direction of speech.

To quantify the independence of STRF tuning from the direction of speech across all electrodes and subjects, we compared the similarity of the STRFs of the same electrode from different directions with the similarity of STRFs of different electrodes from the same direction. The histogram of the correlation values in Figure 3C shows that STRFs for the same electrode from different directions (red) are significantly more similar than STRFs for different electrodes from the same direction (blue) (difference between median correlation = 0.5, p < 0.001, Wilcoxon rank-sum test). The same result was obtained using just the lSTRF and just the rSTRF (Figures S4B and S4C).

We further studied the decoupling of spectrotemporal and spatial tuning properties by measuring the best frequency (BF) and response latency (RL) of each STRF by finding its excitatory peak as the center of gravity along frequency and time dimensions, respectively (marked with a black dot in Figure 3A STRFs). The BF and RLfor electrodes are displayed on an ICBM152 brain in Figure 3B, showing a high-to-low gradient for BF and low-to- high gradient for RL along the medial-lateral axis, consistent with previous studies (Leaver and Rauschecker, 2016; Moerel et al., 2014; Nourski et al., 2014). To examine the effect of speech direction on BF and RL, we calculated STRFs from only right sided speech and STRFs from only left-sided speech (y axis in Figure 3D; blue and red show left- and right-sided angles, respectively) and compared their BF and RL parameters with STRFs calculated from center-only speech (x axis in Figure 3D). The high correlation values fortuning parameters from different angles (r = 0.94 and p < 0.001 for BF and r = 0.80 and p < 0.001 for RL) show a high similarity between the spectrotemporal tuning properties estimated from different sound directions. We observed the same result when the analysis was performed in individual subjects (Figure S4D) or brain regions (Figure S4E). This result supports the notion of independent encoding of spectrotemporal and spatial features of speech in human core auditory cortical areas.

Although the previous analysis shows that spectrotemporal tuning properties of electrodes remain constant irrespective of sound direction, it does not rule out the possibility of correlated encoding of spatial and spectrotemporal features. To test this possibility, we examined the effect of BF and RL on BA tuning. Figure 3E shows a scatter of BF and RL for each electrode, where the electrode color indicates its BA. The lack of clustering among electrodes with similar colors shows that electrodes tuning to spatial and spectrotemporal features are not correlated ($r = -0.04$, $p = 0.64$ for BF versus BA; $r = 0.08$, $p = 0.29$ for RL versus BA); therefore, the spatial and spectrotemporal feature maps appear to be organized independently of each other.

Population Decoding of Spatial and Phonetic Features

Spectrotemporal tuning properties are commonly used to characterize auditory neurons (Klein et al., 2006; Theunissen et al., 2001b). However, speech is a specialized, complex signal, constructed by concatenating distinctive units called phonemes (Ladefoged and Johnson, 2010). The reliable encoding of distinctive spectrotemporal features necessary for phonetic discrimination in the human auditory cortex is crucial for speech perception. Previous studies have shown that human core auditory areas encode phonetic features (Arsenault and Buchsbaum, 2015; Steinschneideretal., 2005). Consistent with these studies, we found that, of the 148 electrodes from this study that were presented with both speech and non-speech sounds, 84 (56.76%) responded significantly more to speech than non-speech ($p < 0.05$, unpaired t test) (STAR Methods; Figure S8). We did not find a relationship between speech specificity of electrodes and their degree of spatial tuning (Spearman $r = -0.12$, $p = 0.15$) (Figure S8B). Motivated by this observation, we extended our analysis of spectrotemporal features (Figure 3) to explicitly examine how the representation of phonetic features interacts with the encoding of spatial features. We used five manners of articulation (vowel, semivowel, plosive, fricative, and nasal) to represent the phonetic features (Khalighinejad et al., 2017a; Mesgarani et al., 2014).

To study how the population of electrodes in the core auditory areas represent spatial and phonetic information, we tested how well the same population of responses could decode each type of information using a rudimentary linear classifier. For spatial decoding, we classified the direction of sound from the population responses averaged over a time window of 1.5 s and Z-scored (Figure S5A). Using all electrodes ($n = 201$), we could decode direction significantly better than chance (75.59%, $p < 0.001$, binomial test; chance = 22.16% using a permutation test) (Figure 4A). The spatial classifier’s confusion patterns show that speech uttered from the same side of space is more likely to be confused (−90° with −45° and 90° with 45°). We did not find asymmetry in decoding direction using electrodes from just the left or the right hemisphere of the brain (Figure S5C).

For phonetic decoding, we classified the manner of articulation of phonemes using the same population of neural data, averaged over a time window of 30ms and Z-scored. Like spatial decoding, manner decoding using all electrodes ($n = 201$) was significantly better than chance (57.65%, $p < 0.001$, binomial test; chance = 22.35% using a permutation test). The manner classifier’s confusion patterns reveal two separate groups of confusions: consonants (plosives and fricatives) and sonorants (vowels, semivowels, and nasals). This confusion pattern is consistent with findings in previous psychoacoustic (Miller and Nicely, 1955) and neurophysiological studies (Mesgarani et al.,2008). Additionally, angle and manner decoding were both possible when using only the electrodes located in the Heschl’s gyrus, STG, or PT (Figure S5F). Successful population decoding of both direction and phonetic features confirms that the population of neural responses in the perisylvian region is rich enough to support downstream extraction of both spatial “where” and phonetic “what” information, which may happen in separate subsequent pathways (Ahveninen et al., 2006; Barrett and Hall, 2006).

Because of the strong contralateral preference we observed in spatial selectivity (Figure 2), we also examined how well the direction of sound and phonetic features can be decoded from each hemisphere separately. We trained two additional classifiers: one using only right-hemisphere electrodes and one using only left-hemisphere electrodes. Because our limited sampling of the perisylvian area could bias the results, we first examined the variability of classification accuracy using different subsets of electrodes in each hemisphere. We performed a bootstrap analysis in which we classified the sound direction and manner of articulation from the responses of randomly selected subsets of electrodes of size N (STAR Methods), as N increased systematically (with 200 bootstraps for each N). When decoding from both hemispheres, N/2 electrodes were chosen from each brain hemisphere. The monotonic increase in classification accuracy as N increases (Figure 4B) indicates that complementary information is added by increasing coverage. We observed the same trend in individual subjects (Figure S5E). The relative accuracy improvement from single to both hemispheres was significantly higher for angle classification than for manner classification ($p < 0.001$, unpaired t test) (Figures 4B and4D), indicating that access to both brain hemispheres achieves a significantly higher accuracy in decoding the direction of sound but has no advantage in decoding phonetic features. This finding is consistent with studies showing an important role for both hemispheres in identifying the direction of sound in space (Kavanagh and Kelly, 1987; Poirier et al., 1994) and studies showing bilateral, symmetric processing of low-level speech features (Arsenault and Buchsbaum, 2015; Hickok and Poeppel, 2007).

To gain insight into how the electrodes are used to decode spatial and phonetic features of speech, we examined the weights given to each electrode by each type of classifier (Figure 4C). The weights for single-hemisphere and whole-brain classifiers are shown in Figure 4C for angle (left) and manner decoding (right). To assist visualization, the electrodes (y axis) for the spatial decoder are sorted according to ASI, and the electrodes for the manner decoder are sorted according to BF. As expected for direction decoding, the classifier with access to both hemispheres assigned positive weights to electrodes on the contralateral side and negative weights to electrodes on the ipsilateral side. For manner of articulation decoding, the classifier assigned higher weights to electrodes with a higher BF for phonetic features characterized by high-frequency acoustic features (plosives and fricatives) and higher weights to electrodes with lower BFs for manners characterized by low-frequency acoustic features (vowels, nasals, and semivowels) (acoustic features for five manners of articulation are shown in Figure S7) (Ladefoged and Johnson, 2010). The correlation between weights for single-hemisphere and both-hemisphere classifiers (Figure 4C) is higher for manner decoding ($r = 0.97$, $p < 0.001$) than for angle decoding ($r = 0.93$, $p < 0.001$). The lesser correlation between single and both hemisphere weights for angle decoding suggests that ipsilaterally selective electrodes are given higher weights for the ipsilateral direction when only one brain hemisphere is available compared with when both brain hemispheres are available (Figure S5D). In contrast, the higher correlation of manner decoding between one and both hemisphere weights shows a more symmetric representation of spectrotemporal features (Figure S5B).

To quantify the relationship between the weights assigned for manner decoding and angle decoding, we measured the correlation between the maximum absolute weight given to each electrode by manner and angle decoders. The positive correlation (Figure 4E; r = 0.37, p < 0.001) supports the notion that the same population of neurons in the core auditory cortex carries information about both spatial and phonetic features of speech, which is consistent with the notion of independent joint encoding of these parameters.

Mechanisms of Joint Encoding of Spatial and Spectrotemporal Features

The previous analyses show that spatial and spectrotemporal features of sound are jointly represented by the same group of electrodes and that spectrotemporal feature selectivity remains the same irrespective of sound direction. To shed light on the mechanism of this encoding, we test the hypothesis that, although the sound direction does not change spectrotemporal tuning properties, it can modulate the response gain and/or the mean response level (bias) of neural activity. Because the STRF does not model nonlinear dynamics such as enhanced onset responses (David et al., 2009), we restricted the analysis to only the sustained response interval (0.5 s to 3.75 s after onset of the stimulus) (Figures S1B and S1C).

To start, we calculated the relative change in the mean neural response level (bias) between speech from angle 0° and angle 90° and between speech uttered from angle 0° and angle −90°. We also calculated the relative change in the standard deviation (gain) of the neural activity for the same comparisons (Figure 5A). We observed higher separation in the mean response of right- and left-hemisphere electrodes (p < 0.001, Wilcoxon rank-sum test) than in the standard deviation (p = 0.12, Wilcoxon rank-sum test). Motivated by these findings, we constructed a model where the sound direction can modify both the bias and the gain of the spectrotemporal receptive field of the electrode:

$$r(t) = [S(t, f)*STRF\ (t, f)]\ g(angle) + b(angle), \qquad \text{(Equation 1)}$$

where * denotes convolution, $r(t)$ denotes the predicted neural response, $S(t,f$ is the spectrogram of the sound, $STRF(t,f$ is the spectrotemporal receptive field, and the direction of speaker (angle) modulates both the gain, $g\ (angle)$, and the bias, $b\ (angle)$ (Figure 5B). We used the least-squares method to fit the parameters of the model to the predicted data from non-spatial STRFs (STRFs calculated from a mono stimulus) in three scenarios: gain change, bias change, and gain and bias change. We then calculated the improvement in mean squared error (MSE) relative to the non-spatial STRF model predictions for each model. We found that the average improvement in MSE is significantly higher for the bias model compared with the gain model (Figure 5C; p < 0.001, Wilcoxon rank-sum test). Additionally, the model that modifies both gain and bias is not significantly more predictive than the model that only modifies the bias (p = 0.11, Wilcoxon rank-sum test), suggesting that sound direction modulates the bias of neural responses and not its gain. This result also holds when broken down by subject (Figure S6A) and by anatomical region (Figure S6B). Furthermore, as shown by the average bias values in Figure 5D, we found that direction of sound increases the baseline bias for contralateral directions and decreases the baseline bias for ipsilateral directions, consistent with our previous observation of strong contralateral bias (Figure 2). Bias values arranged by ASI clustering (Figure 5E) reveal a similar pattern as in Figure 1D (r = 0.57, p < 0.001).

DISCUSSION

We use direct neural recordings in the human core auditory cortex to study the representation of speech with spatial cues. We show that individual electrodes in core auditory areas respond selectively to specific directions while independently encoding spectrotemporal and phonetic features of speech. This encoding results in a representation from which both location and the phonetic features can be readily decoded. We showed that, although the location of the speaker does not change the spectrotemporal tuning of electrodes, it modulates the mean response level of high-gamma activity.

Previous electrophysiological studies of spatial hearing have used either non-invasive neuroimaging methods in humans (Ah-veninen et al., 2014) or invasive neural recordings in animals (Jenkins and Merzenich, 1984; Middlebrooks and Pettigrew, 1981). However, on one hand, non-invasive studies in humans lack the temporal and spatial resolution needed to examine the precise encoding properties of sound. As a result, studies of spatial tuning in humans have often reported inconsistent and sometimes contradictory findings. Because our electrophysiology method provides high spatial and temporal resolution, our findings provide critical evidence needed to reconcile these inconsistencies, as we elaborate below. On the other hand, the extent to which the findings from animal spatial hearing studies generalize to humans is unclear because the auditory cortex in humans, including the Heschl’s gyrus, differs significantly from non-human mammals (Morosan et al., 2001). Because the structural and functional organization of the human core auditory cortex is still a matter of scientific debate (Leaver and Rauschecker, 2016; Moerel et al., 2014), particularly regarding its response to speech (Belin et al., 2000), our findings provide critical evidence needed to better compare with animal studies, which can result in a more complete understanding of the functional organization of the human auditory cortex.

We found that the cortical representation of speech is highly selective to specific sound directions at the level of individual electrodes. Although previous non-invasive human studies have reported activation of core auditory areas in response to sound location (Johnson and Hautus, 2010; Junius et al.2007), our study provides direct evidence for direction-specific tuning in these areas. We observed a varied degree of tuning to sound directions, reflecting a diversity of neural responses with a strong contralateral bias (87.06%). This ratio is similar to previously reported spatial tuning of auditory cortical neurons in mammals (Malhotra et al., 2004; Middlebrooks and Pettigrew, 1981; Rajan et al., 1990). Our findings, however, contrast with several neuroimaging studies in humans that report little to no contralateral bias (Woldorff et al., 1999; Zimmer et al., 2006). This discrepancy could be the result of bias-coding rather than place-coding of sound direction, which would not be easily detected in fMRI (Werner-Reiss and Groh, 2008). Alternatively, the absence of contralateral tuning in sound localization studies has been attributed to the presence or absence of interaural level differences (ILDs) and interaural time differences (ITD) in the experimental design (Ortiz-Rios et al., 2017; Spierer et al.,2009). In contrast, our experimental design is naturalistic, presenting speech with all its spatial acoustic cues intact (ITD, ILD, and spectral cues).

Another difference between our findings and previous literature in humans regards the question of asymmetrical processing of sound location. Animal literature in sound localization has established hemi-field processing of spatial sounds (Jenkins and Merzenich, 1984; Middlebrooks and Pettigrew, 1981). Contrary to the expectations from animal models, several human studies reported right-hemispheric dominance for spatial sound processing, including asymmetrical responses to spatial sounds (Brunetti et al., 2005; Griffiths et al., 1998; Krumbholz et al., 2005) and spatial hemi-neglect after lesions to the right hemisphere (Spierer et al., 2009), including lesions encroaching on the Heschl’s gyrus (Zatorre and Penhune, 2001). In contrast, we did not find any hemispheric differences in spatial cue encoding in the core auditory cortex and areas middle-lateral to the core. It is worth mentioning that the reported difference in the majority of previous studies is either in cortical areas other than those we focused on (Brunetti et al., 2005; Bushara et al., 1999; Griffiths et al., 1998; Krumbholz et al., 2005) or based on only ITDs (Krumbholz et al., 2005; Spierer et al., 2009), which have been shown to have a right hemisphere processing bias (Ortiz-Rios et al., 2017). Our results suggest that naturalistic speech containing all spatial cues is processed symmetrically in core auditory areas.

Furthermore, we found no correlation between the spatial and spectrotemporal tuning properties of individual electrodes, suggesting that local tuning to these features of speech is independent in core auditory cortical areas. A similarly dissociated encoding of spatial and non-spatial features has been reported in auditory cortical neurons of ferrets (Bizley et al., 2009; Walker et al., 2011) and cats (Harrington et al., 2008; Stecker et al., 2003). In contrast, several human neuroimaging studies have reported that Heschl’s gyrus preferentially encodes spectrotemporal sound features over spatial features (Alain et al., 2001; Barrett and Hall, 2006). One possible explanation is that the use of unnatural sounds in these studies may not optimally activate auditory cortical neurons (Theunissen and Elie, 2014). Although joint encoding of spatial and spectrotemporal features has been found in mammalian studies, the human auditory cortex is specialized for speech processing (Belin et al., 2000; Norman-Haignere et al., 2015). Indeed, many of the electrodes in our study responded significantly more to speech than to non-speech sounds (STAR Methods; Figure S8). Our study therefore takes a further step by examining the organization of responses to phonetically relevant features that are crucial for distinctions of phonemes. However, although our experiment engages the specialized speech circuits, we did not find a correlation between the speech specificity of electrodes and their spatial tuning properties (Spearman $r = -0.12$, $p = 0.15$) (Figure S8). Hence, the mechanisms of spatial feature representation that we characterized in this study are likely the same for other classes of sounds.

Previous neuroimaging studies in humans have demonstrated the involvement of the auditory cortex in spatial hearing; our study extends this finding by teasing apart the encoding properties of phonetic and spatial cues in this region. We find that spatial and phonetic features are jointly represented in the core auditory cortex and areas middle-lateral from the core, providing a foundation for extraction of both phonetic information and the location of the speaker. Studies in humans (Arnott et al., 2004) and non-human primates (Rauscheckerand Tian, 2000; Romanski et al., 1999) have hypothesized separate dorsal and ventral pathways for processing spectrotemporal "what" and spatial "where" features, but how this split arises remains unexplored. We did not observe a major difference in spatial and spectrotemporal feature encoding in core auditory areas. Although we could not test the existence of these separate pathways directly because of the absence of sufficient coverage in areas anterior (planum polare [PP]) and posterior (planum temporale [PT]) to Heschl's gyrus, our results characterize the representational properties of the core auditory cortex, the area hypothesized to be the origin of the where and what pathways. Our results therefore demonstrate how a linear readout from this representation by downstream neural pathways could readily decode both spatial and phonetic features of speech.

Last, we show that, although sound direction does not change the spectrotemporal tuning of individual electrodes, it modulates the mean response level of the high-gamma activity. Because it has been shown that the high-gamma amplitude reflects the firing rate of the neural population proximal to the electrode (Buzsaki et al., 2012; Ray and Maunsell, 2011), a likely explanation for the change in mean response level is a change in the average firing rate of the underlying neural population as the speech changes direction. Alternatively, the increase in high-gamma amplitude may be due to the recruitment of a larger number of neurons with similar spectrotemporal tuning but a different response threshold (Phillips et al., 1994). Teasing apart the two scenarios requires recording from individual neurons, which is beyond the resolution of our current electrophysiology method.

Together, these findings advance our knowledge of the representational and functional organization of human auditory cortex and pave the way toward more complete models of cortical speech processing in the human brain.

Conclusion and Future Directions

We characterized the representational properties of speech with spatial cues in the human core auditory cortex. We found local selectivity to specific speaker directions at the level of individual electrodes that jointly and independently represent spatial and phonetic features. These findings raise several further questions. First, it is unclear how exactly the information in each hemisphere and across hemispheres is used to reliably estimate the location of a speaker in space. Second, we show that "what" and "where" information can be extracted from the representation in core auditory areas. Our coverage cannot address whether separate neural pathways exist for processing these cues. Moreover, it remains unclear how the cortical representation changes as a listener engages in a task that requires attending to a particular location in space. Future studies are needed to uncover the effects of top-down and adaptive mechanisms on modulating the cortical representation of speech with spatial cues in the human auditory cortex.

STAR ★ METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MATLAB	MathWorks, Natick, MA	N/A
Freesurfer parcellation	Dykstra et al., 2012; Fischl et al., 2004	N/A
Brainstorm	Tadel et al., 2011	N/A
LISTEN HRTF	J.-M. Jot et al., 1995, Audio Engineering Society, conference	N/A
NAPLib	Khalighinejad et al., 2017b	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Prof. Nima Mesgarani (nima@ee.columbia.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants and data collection

Participants were 2 male and 3 female native English speakers of age ranging from 33 to 59 years with self-reported normal hearing. As part of their clinical diagnosis of epileptic focus, each subject was implanted bilaterally with customized high-density multielectrode arrays. The number of speech responsive sites differed among subjects. Electrocorticography signals with sampling rate of 3000Hz (5 subjects) were recorded with a multichannel amplifier connected to a digital signal processor (Tucker-Davis Technologies). All data was montaged again to common average reference (Crone et al., 2001). Neural responses were first filtered using Hilbert transform to extract the high-gamma band (70Hz-150Hz) for analysis (Edwards et al., 2009) and were then down-sampled to 100Hz. The research protocol was approved by the Institutional Review Board of Columbia University, and signed consent forms were obtained from all patients for all evaluation procedures.

METHOD DETAILS

Stimulus

Natural American English sentences were presented using standard Panasonic stereo earphones (Panasonic RP-HJE120) at a sampling rate of 44.1KHz. We presented 123 speech trials varying in duration from 2–13 s each, randomly divided among five angles. We compared 30 individual head-related transfer functions (HRTFs) (J.-M. Jot et al., 1995, Audio Engineering Society, conference) from LISTEN HRTF database, and chose the HRTF of an average size head to render binaural speech in space. Each speech trial was then perceived as if it was uttered from 0 elevation and azimuths randomly chosen from 5 angles: −90, −45, 0, 45, 90. To confirm that subjects attended to the speech, they were asked to point in the direction of the last sound at random pauses during the task and behavior was recorded (Figure S2).

QUANTIFICATION AND STATISTICAL ANALYSIS

Angle Selectivity; Estimation of ASI, CSI, BA

Speech-responsive sites were determined by calculating the maximum t-value of each electrode's response between silence and speech from any of the 5 angles. Electrodes with a maximum t-value greater than 17 ($p < 0.001$) were selected (Figure S1A), resulting in 61, 29, 31, 24, 56 electrodes from subjects 1–5 respectively for a total of 201 electrodes used in all further analysis.

We observed that t-values calculated between silence and onset, and that between silence and sustained part of response were highly correlated (Figure S1C). For each site, we calculated t-values for the neural response between silence and single angle resulting in 5 t-values, one for each angle. This t-value vector was then z-scored to obtain relative angle preference for each electrode. The new vector was normalized by subtracting the minimum and dividing by the sum to restrict the ASI values between 0 and 1. This normalization leads to a probabilistic interpretation of ASI, so that for a given electrode, ASI for an angle is the probability that the electrode prefers that angle. We performed unsupervised hierarchical clustering on rows (angles) and columns (electrodes) of these ASI vectors based on correlation distance to generate a local selectivity matrix. BA was defined as the angle to which an electrode has maximum ASI.

We measured the amount of contralateral preference by defining a Contralateral Selectivity Index (CSI):

$$\text{CSI} = \frac{\left(\sum \text{t} - \ \text{values for angles} \ -45 \ \& \ - \ 90\right) - \left(\sum \text{t} - \ \text{values for angles } 45 \ \& \ 90\right)}{\left(\sum \text{t} - \ \text{values for angles} \ -45 \ \& -90\right) \ + \ \left(\sum \text{t} - \ \text{values for angles } 45 \ \& \ 90\right)}.$$

Spectrotemporal Receptive Fields; Estimation of BF and RL

We calculated the spectrotemporal receptive fields (STRF) of each electrode using a normalized reverse correlation algorithm (STRFLab software package available at http://www.strflab.berkeley.edu) (Theunissen et al., 2001a). Regularization and cross-validation techniques were used to prevent over-fitting of the STRF (David et al., 2007). STRFs were calculated using both mono-stimulus (stimulus without HRTF filters) and stereo-stimulus (separately for left (/STRF) and right (rSTRF) channel inputs). The time-frequency auditory spectrogram was generated using a model of the peripheral auditory system (Chi et al., 2005) using mono, stereo-left and stereo-right stimuli.

To determine the relationship between direction and tuning, we used the stereo stimulus to compute the STRF (separately for left and right inputs) for all five angles. The following ranges of sparseness and tolerance values were used to calculate STRFs: tolerance 0.1, 0.05, 0.5; sparseness 8,16,32. However, maximizing the prediction score on cross-validation subset resulted in the same tolerance value of 0.05 and sparseness value of 8 for all electrodes and all angles.

Decoding

We used a regularized least square (RLS) linear classifier to decode azimuth and manner of articulation from the neural data (training on 90% data, testing on 10% over 10 cross-validations). To select the window size for azimuth decoding, we made a plot of decoding accuracy versus time window (Figure S5A) and found a steep rise in accuracy with increase in time window which plateaued at 1.5 s window and therefore chose this time duration for analysis. For manner decoding, we averaged the neural response in time window of 30 ms centered at the maximum phonetic separability ie. peak f-statistic (Patel et al., 1976).

To check the effect of sample size on accuracy, we performed bootstrap analysis for left-hemisphere-only, right-hemisphere-only, and both hemisphere electrodes. For N ranging from 2 to 90, we selected N different electrodes randomly from left brain sites, N from right brain sites, and Nfrom both brain sites (N/2 left and N/2 right) and calculated the decoding accuracy of 5 angles for each case by training on 90% data, testing on 10% data, and cross-validating 10 times. For each N, we bootstrapped 200 times for single hemispheres (100 for Left and 100 for Right) and 200 for both brain hemispheres. We constructed a plot of mean decoding accuracy and standard deviation across bootstraps versus number of electrodes.

Model

To determine a model that explains the encoding of spatial information, we used the mono-STRF computed by combining all angles as absence of spatial information was imperative to determine the model improvement. Using linear regression with least-squares algorithm, we fitted the predicted neural response to three models specified by: $y = ax; y=x + b; y = ax +$ b, where y is the actual brain response, x is the STRF-predicted response, and a (gain) and b (mean response level or bias) are variable parameters. We then calculated MSE between predicted and actual responses before and after fitting the data for all three cases. We estimated percentage decrease in the error by comparing the error after to before model fitting.

Speech Specificity

Speech specificity of electrodes was measured using a separate listening experiment which included 69 consisting of speech, environmental noises, music genres, coughing, laughter, and tones. Because of time constraint, this task was recorded in 3 of the 5 subjects. We calculated the average response of the electrodes to all trials, and performed a t test between the responses to speech and non-speech classes (Figure S8) to determine speech specificity.

Generation of Brain Figures

The electrodes were mapped onto the brain of each subject using co-registration by iELVis (Groppe et al., 2017) followed by their identification on the post-implantation CT scan using Biolmage Suite(Papademetris et al., 2006). Anatomical locations of these electrodes were obtained using Freesurfer’s automated cortical parcellation (Dykstra et al., 2012; Fischl et al., 2004) by Destrieux brain atlas (Destrieux et al., 2010). These labels were closely inspected by neurosurgeon using subject’s co-registered post-implant MRI. The electrodes were plotted on the average brain template ICBM152 (Fonov et al., 2011) using Brainstorm (Tadel et al., 2011).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was funded by a grant from NIH, NIDCD (DC014279) and the Pew Charitable Trusts Pew Biomedical Scholars Program.

REFERENCES

Ahveninen J, Jȁȁskelȁinen IP, Raij T, Bonmassar G, Devore S, Hȁmȁlȁinen M, Levȁnen S, Lin F-H, Sams M, Shinn-Cunningham BG, et al. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. Proc. Natl. Acad. Sci. USA 703, 14608–14613.

Ahveninen J, Kopco N, and Jȁȁskelȁinen IP (2014). Psychophysics and neuronal bases of sound localization in humans. Hear. Res. 307, 86–97. [PubMed: 23886698]

Alain C, Arnott SR, Hevenor S, Graham S, and Grady CL (2001). “What” and “where” in the human auditory system. Proc. Natl. Acad. Sci. USA 98, 12301–12306. [PubMed: 11572938]

Arnott SR, Binns MA, Grady CL, and Alain C (2004). Assessing the auditory dual-pathway model in humans. Neuroimage 22, 401–408. [PubMed: 15110033]

Arsenault JS, and Buchsbaum BR (2015). Distributed neural representations of phonological features during speech perception. J. Neurosci. 35, 634–642. [PubMed: 25589757]

Barrett DJK, and Hall DA (2006). Response preferences for “what” and “where” in human non-primary auditory cortex. Neuroimage 32, 968–977. [PubMed: 16733092]

Belin P, Zatorre RJ, Lafaille P, Ahad P, and Pike B (2000). Voice-selective areas in human auditory cortex. Nature 403, 309–312. [PubMed: 10659849]

Bizley JK, Walker KMM, Silverman BW, King AJ, and Schnupp JWH (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J. Neurosci. 29, 2064–2075. [PubMed: 19228960]

Brunetti M, Belardinelli P, Caulo M, Del Gratta C, Della Penna S, Ferretti A, Lucci G, Moretti A, Pizzella V, Tartaro A, et al. (2005). Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study. Hum. Brain Mapp. 26, 251–261. [PubMed: 15954141]

Bushara KO, Weeks RA, Ishii K, Catalan M-J, Tian B, Rauschecker JP, and Hallett M (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. Nat. Neurosci. 2, 759–766. [PubMed: 10412067]

Buzsaki G, Anastassiou CA, and Koch C (2012). The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 73, 407–420.

Chi T, Ru P, and Shamma SA (2005). Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. 118, 887–906. [PubMed: 16158645]

Crone NE, Boatman D, Gordon B, and Hao L (2001). Induced electroc orticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. Clin. Neurophysiol. 112, 565–582. [PubMed: 11275528]

David SV, Mesgarani N, and Shamma SA (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. Network 18, 191–212. [PubMed: 17852750]

David SV, Mesgarani N, Fritz JB, and Shamma SA (2009). Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. J. Neurosci. 29, 3374–3386. [PubMed: 19295144]

Destrieux C, Fischl B, Dale A, and Halgren E (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15. [PubMed: 20547229]

Dykstra AR, Chan AM, Quinn BT, Zepeda R, Keller CJ, Cormier J, Madsen JR, Eskandar EN, and Cash SS (2012). Individualized localization and cortical surface-based registration of intracranial electrodes. Neuroimage 59, 3563–3570. [PubMed: 22155045]

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, and Knight RT (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. J. Neurophysiol. 102, 377–386. [PubMed: 19439673]

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, et al. (2004). Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22. [PubMed: 14654453]

Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, and Collins DL; Brain Development Cooperative Group (2011). Unbiased average age- appropriate atlases for pediatric studies. Neuroimage 54, 313–327. [PubMed: 20656036]

Formisano E, Kim D-S, Di Salle F, van de Moortele P-F, Ugurbil K, and Goebel R (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40, 859–869. [PubMed: 14622588]

Geschwind N, and Levitsky W (1968). Human brain: left-right asymmetries in temporal speech region. Science 161, 186–187. [PubMed: 5657070]

Griffiths TD, Rees G, Rees A, Green GGR, Witton C, Rowe D, Buchel C, Turner R, and Frackowiak RSJ (1998). Right parietal cortex is involved in the perception of sound movement in humans. Nat. Neurosci. 1, 74–79. [PubMed: 10195113]

Groppe DM, Bickel S, Dykstra AR, Wang X, Megevand P, Mercier MR, Lado FA, Mehta AD, and Honey CJ (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. J. Neurosci. Methods 281, 40–48. [PubMed: 28192130]

Hackett TA, Preuss TM, and Kaas JH (2001). Architectonic identification ofthe core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. 441, 197–222. [PubMed: 11745645]

Harrington IA, Stecker GC, Macpherson EA, and Middlebrooks JC(2008). Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. Hear. Res. 240, 22–41. [PubMed: 18359176]

Hickok G, and Poeppel D (2007). The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402. [PubMed: 17431404]

Humphries C, Liebenthal E, and Binder JR (2010).Tonotopicorganization of human auditory cortex. Neuroimage 50, 1202–1211. [PubMed: 20096790]

Jenkins WM, and Merzenich MM (1984). Role of cat primary auditory cortex for sound-localization behavior. J. Neurophysiol. 52, 819–847. [PubMed: 6512590]

Johnson BW, and Hautus MJ (2010). Processing of binaural spatial information in human auditory cortex: neuromagnetic responses to interaural timing and level differences. Neuropsychologia 48, 2610–2619. [PubMed: 20466010]

Junius D, Riedel H, and Kollmeier B (2007). The influenceofexternalization and spatial cues on the generation of auditory brainstem responses and middle latency responses. Hear. Res. 225, 91–104. [PubMed: 17270375]

Kavanagh GL, and Kelly JB (1987). Contribution of auditory cortex to sound localization by the ferret (Mustela putorius). J. Neurophysiol. 57, 1746–1766. [PubMed: 3598629]

Khalighinejad B, Cruzatto da Silva G, and Mesgarani N (2017a). Dynamic encoding of acoustic features in neural responses to continuous speech. J. Neurosci. 37, 2176–2185. [PubMed: 28119400]

Khalighinejad B, Nagamine T, Mehta A, and Mesgarani N (2017b). NAPLib: An open source toolbox for real-time and offline Neural Acoustic Processing. Proc. IEEE Int. Conf. Acoust. Speech Signal Process 2017, 846–850. [PubMed: 29430213]

Kidd G, Jr., Arbogast TL, Mason CR, and Gallun FJ (2005). The advantage of knowing where to listen. J. Acoust. Soc. Am. 118, 3804–3815. [PubMed: 16419825]

Klein DJ, Simon JZ, Depireux DA, and Shamma SA (2006). Stimulusinvariant processing and spectrotemporal reverse correlation in primary auditory cortex. J. Comput. Neurosci. 20, 111–136. [PubMed: 16518572]

Krumbholz K, Schonwiesner M, von Cramon DY, Rubsamen R, Shah NJ, Zilles K, and Fink GR (2005). Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. Cereb. Cortex 15, 317–324. [PubMed: 15297367]

Ladefoged P, and Johnson K (2010). A Course in Phonetics, Sixth Edition (Cengage Learning).

Leaver AM, and Rauschecker JP (2016). Functional topography of human auditory cortex. J. Neurosci. 36, 1416–1428. [PubMed: 26818527]

Malhotra S, Hall AJ, and Lomber SG (2004). Cortical control of sound localization in the cat: unilateral cooling deactivation of 19 cerebral areas. J. Neurophysiol. 92, 1625–1643. [PubMed: 15331649]

Mesgarani N, David SV, Fritz JB, and Shamma SA (2008). Phoneme representation and classification in primary auditory cortex. J. Acoust. Soc. Am. 123, 899–909. [PubMed: 18247893]

Mesgarani N, Cheung C, Johnson K, and Chang EF (2014). Phonetic feature encoding in human superiortemporal gyrus. Science 343,1006–1010. [PubMed: 24482117]

Middlebrooks JC, and Pettigrew JD (1981). Functional classes of neurons in primary auditory cortexofthe cat distinguished by sensitivity to sound location. J. Neurosci. 1, 107–120. [PubMed: 7346555]

Miller GA, and Nicely PE (1955). An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27, 338.

Moerel M, De Martino F, and Formisano E (2014). An anatomical and functional topography of human auditory cortical areas. Front. Neurosci. 8, 225. [PubMed: 25120426]

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, and Zilles K (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13, 684–701. [PubMed: 11305897]

Norman-Haignere S, Kanwisher NG, and McDermott JH (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. Neuron 88, 1281–1296. [PubMed: 26687225]

Nourski KV (2017). Auditory processing in the human cortex: An intracranial electrophysiology perspective. Laryngoscope Investig. Otolaryngol. 2, 147–156.

Nourski KV, Steinschneider M, McMurray B, Kovach CK, Oya H, Kawasaki H, and Howard MA, 3rd. (2014). Functional organization of human auditory cortex: investigation of response latencies through direct recordings. Neuroimage 101, 598–609. [PubMed: 25019680]

Ortiz-Rios M, Azevedo FAC, Kusmierek P, Balla DZ, Munk MH, Keliris GA, Logothetis NK, and Rauschecker JP (2017). Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. Neuron 93, 971–983.e4. [PubMed: 28190642]

Papademetris X, Jackowski MP, Rajeevan N, DiStasio M, Okuda H, Constable RT, and Staib LH (2006). BioImage Suite: An integrated medical image analysis suite: An update. Insight J. 2006, 209. [PubMed: 25364771]

Patel JK, Kapadia CH, and Owen DB (1976). Handbookofstatistical distributions (M. Dekker).

Phillips DP, Semple MN, Calford MB, and Kitzes LM (1994). Level- dependent representation of stimulus frequency in cat primary auditory cortex. Exp. Brain Res. 102, 210–226. [PubMed: 7705501]

Poirier P, Lassonde M, Villemure J-G, Geoffroy G, and Lepore F (1994). Sound localization in hemispherectomized patients. Neuropsychologia 32, 541–553. [PubMed: 8084413]

Rajan R, Aitkin LM, Irvine DR, and McKay J (1990). Azimuthal sensitivity of neurons in primary auditory cortex of cats. I. Types of sensitivity and the effects of variations in stimulus parameters. J. Neurophysiol. 64, 872–887. [PubMed: 2230931]

Rauschecker JP, and Tian B (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proc. Natl. Acad. Sci. USA 97, 11800–11806. [PubMed: 11050212]

Ray S, and Maunsell JHR (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. 9, e1000610. [PubMed: 21532743]

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, and Rauschecker JP (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2, 1131–1136. [PubMed: 10570492]

Spierer L, Bellmann-Thiran A, Maeder P, Murray MM, and Clarke S(2009). Hemispheric competence for auditory spatial representation. Brain 132, 1953–1966. [PubMed: 19477962]

Stecker GC, Mickey BJ, Macpherson EA, and Middlebrooks JC (2003). Spatial sensitivity in field PAF of cat auditory cortex. J. Neurophysiol. 89,2889–2903. [PubMed: 12611946]

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, and Howard MA, 3rd. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb. Cortex 15, 170–186. [PubMed: 15238437]

Tadel F, Baillet S, Mosher JC, Pantazis D, and Leahy RM (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 879716. [PubMed: 21584256]

Theunissen FE, and Elie JE (2014). Neural processing of natural sounds. Nat. Rev. Neurosci. 15, 355–366. [PubMed: 24840800]

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, and Gallant JL (2001a). Estimating spatio-temporal receptive fields ofauditory and visual neurons from their responses to natural stimuli. Network 12, 289–316. [PubMed: 11563531]

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, and Gallant JL (2001b). Estimating spatio-temporal receptive fields ofauditory and visual neurons from their responses to natural stimuli. Netw. Comput. Neural Syst. 12, 289–316.

Walker KMM, Bizley JK, King AJ, and Schnupp JWH (2011). Multiplexed and robust representations of sound features in auditory cortex. J. Neurosci. 31, 14565–14576. [PubMed: 21994373]

Werner-Reiss U, and Groh JM (2008). A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. J. Neurosci. 28, 3747–3758. [PubMed: 18385333]

Woldorff MG, Tempelmann C, Fell J, Tegeler C, Gaschler-Markefski B, Hinrichs H, Heinz HJ, and Scheich H (1999). Lateralized auditory spatial perception and the contralaterality of cortical processing as studied with functional magnetic resonance imaging and magnetoencephalography. Hum. Brain Mapp. 7, 49–66. [PubMed: 9882090]

Yeni-Komshian GH, and Benson DA (1976). Anatomical study of cerebral asymmetry in the temporal lobe of humans, chimpanzees, and rhesus monkeys. Science 192, 387–389. [PubMed: 816005]

Zatorre RJ, and Penhune VB (2001). Spatial localization after excision of human auditory cortex. J. Neurosci. 21, 6321–6328. [PubMed: 11487655]

Zimmer U, Lewald J, Erb M, and Karnath H-O (2006). Processing of auditory spatial cues in human cortex: an fMRI study. Neuropsychologia 44, 454–461. [PubMed: 16038950]

Highlights

- Direct recording from human auditory cortex reveals selectivity to speech direction
- Spatial and spectrotemporal speech features are independently and jointly encoded
- Neural population responses enable successful decoding of spatial and phonetic features
- Speech direction modulates the mean neural response to spectrotemporal speech features

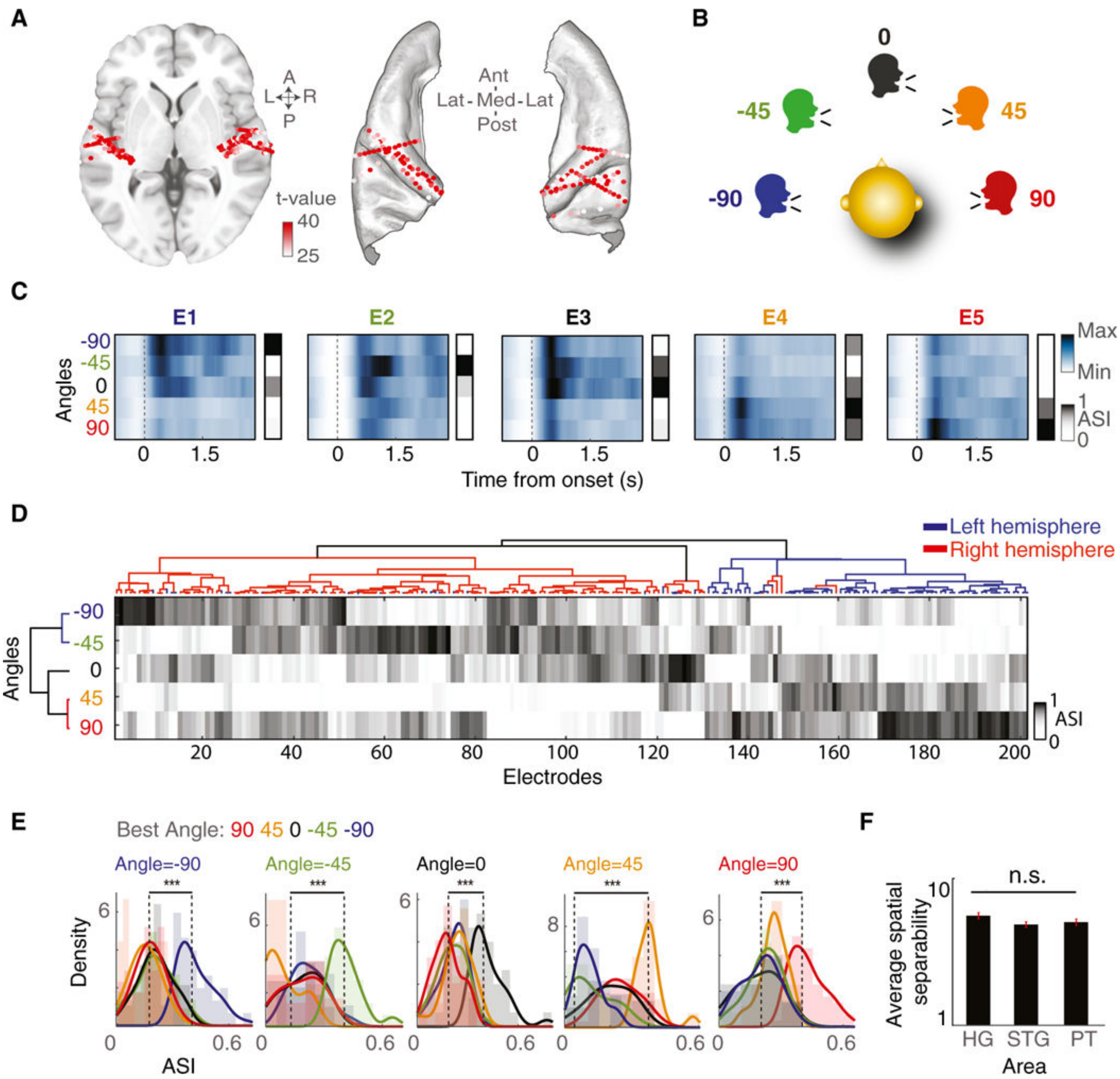


Figure 1. Direction Selectivity of Responses in the Human Auditory Cortex

(A) Speech-responsive electrodes from all subjects shown on an ICBM152 average brain on axial MRI (left) and on core auditory areas (right). Color saturation indicates the speech versus silence t value for each electrode.

(B) Task schematic. Subjects are presented with speech uttered from five color-coded angles in the horizontal plane.

(C) Average high-gamma responses and ASIs of five representative electrodes to −90°, −45°, 0°, 45°, and 90° angles.

(D) Hierarchical clustering of angle selectivity indices (ASIs) for all electrodes (columns) and angles (rows). Electrode clusters are shown by the dendrogram at the top, whereas angle clusters are shown by the dendrogram on the left; electrode clusters in red and blue indicate electrode locations in right and left brain hemispheres, respectively.

(E) Histograms of ASI for a given sound angle; each electrode group is colored by its BA.

(F) Average separability of direction (f statistic) in the Heschl’s gyrus, STG, and PT. The error bars indicate SE.

***p < 0.001. See also Figure S1.

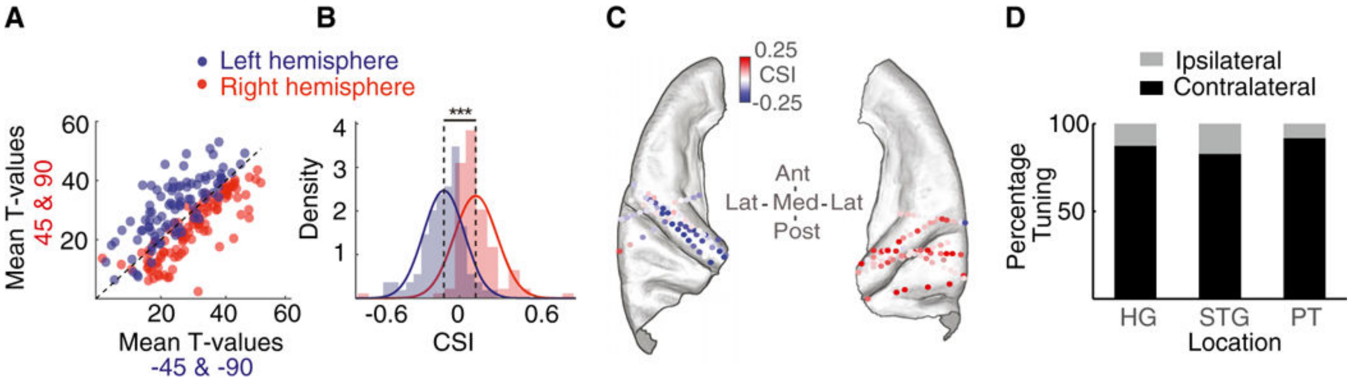


Figure 2. Contralateral Tuning of Electrodes

(A) Average t values for left- versus right-sided angles for each electrode in the left and right brain hemispheres.

(B) Histogram of the contralateral strength index (CSI) for left and right brain hemisphere electrodes.

(C) CSI plotted on an ICBM152 average brain. Each electrode is colored according to its CSI value (red, positive CSI, indicating left angle preference; blue, negative CSI, indicating right angle preference).

(D) Percentages of ipsilateral and contralateral electrodes found in the Heschl’s gyrus, STG, and PT.

***$p < 0.001$. See also Figure S3.

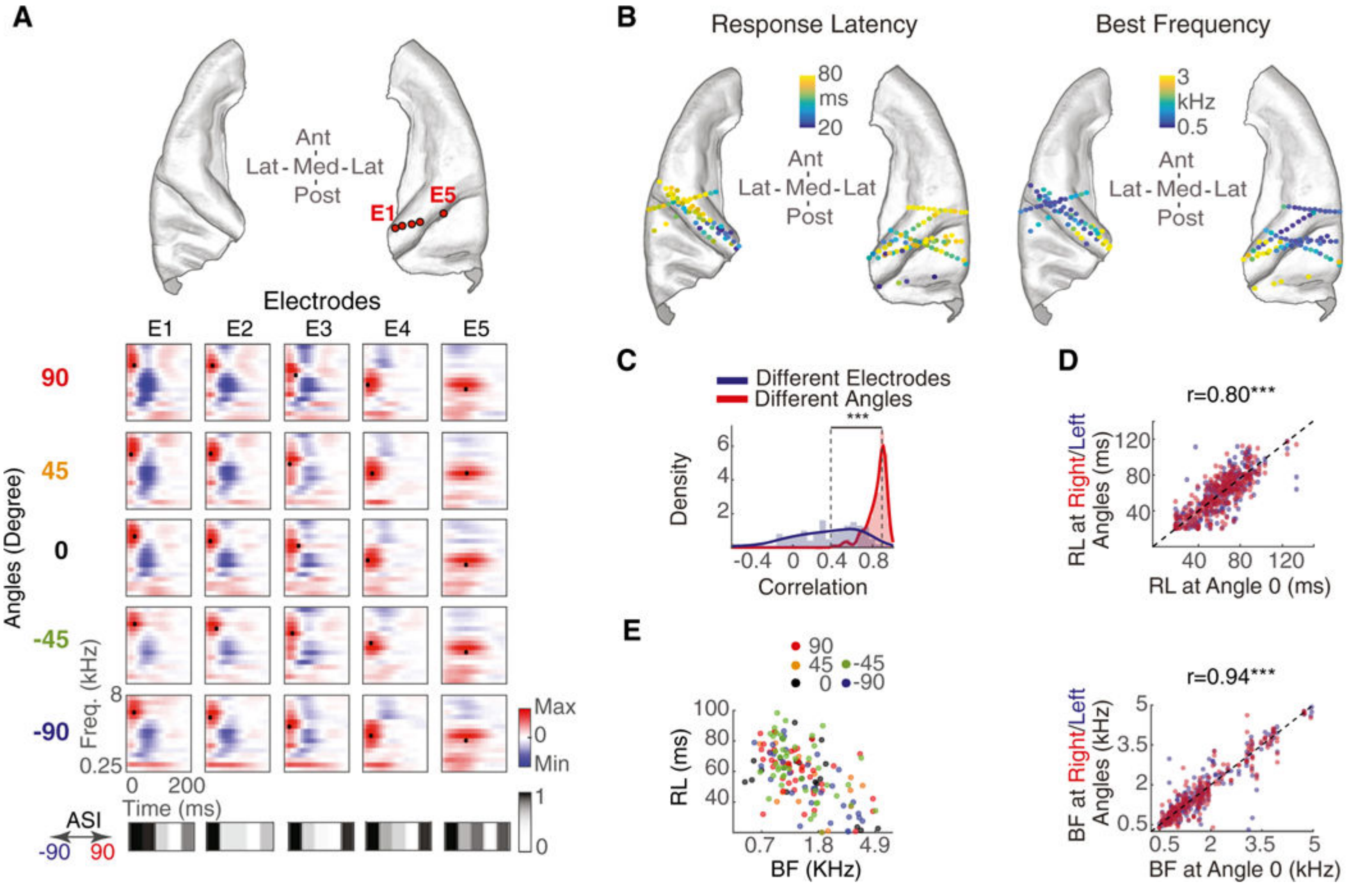


Figure 3. Independent Encoding of Spatial and Spectrotemporal Features

(A) Top: five example electrodes from one subject. Bottom: spectrotemporal receptive fields (STRFs) of these electrodes. Each column is one electrode, and each row is a different speech direction from which the STRF is calculated. The best frequency (BF) and response latency (RL) for each STRF are marked with a black dot. ASI vectors of each electrode are shown below.

(B) Electrodes plotted on the core auditory cortex of ICBM152, color-coded by BF (right) and RL (left).

(C) Histograms of correlation between STRFs from the same angle but different electrodes (blue) compared with STRFs from the same electrode but different angles (red).

(D) RL and BF for right or left angles versus RL and BF for angle 0°. Electrodes are color-coded by angle (red, right side; blue, left side).

(E) BF versus RL plot for all speech-responsive electrodes, colored by their best angle (BA) tuning.

***p < 0.001. See also Figure S4.

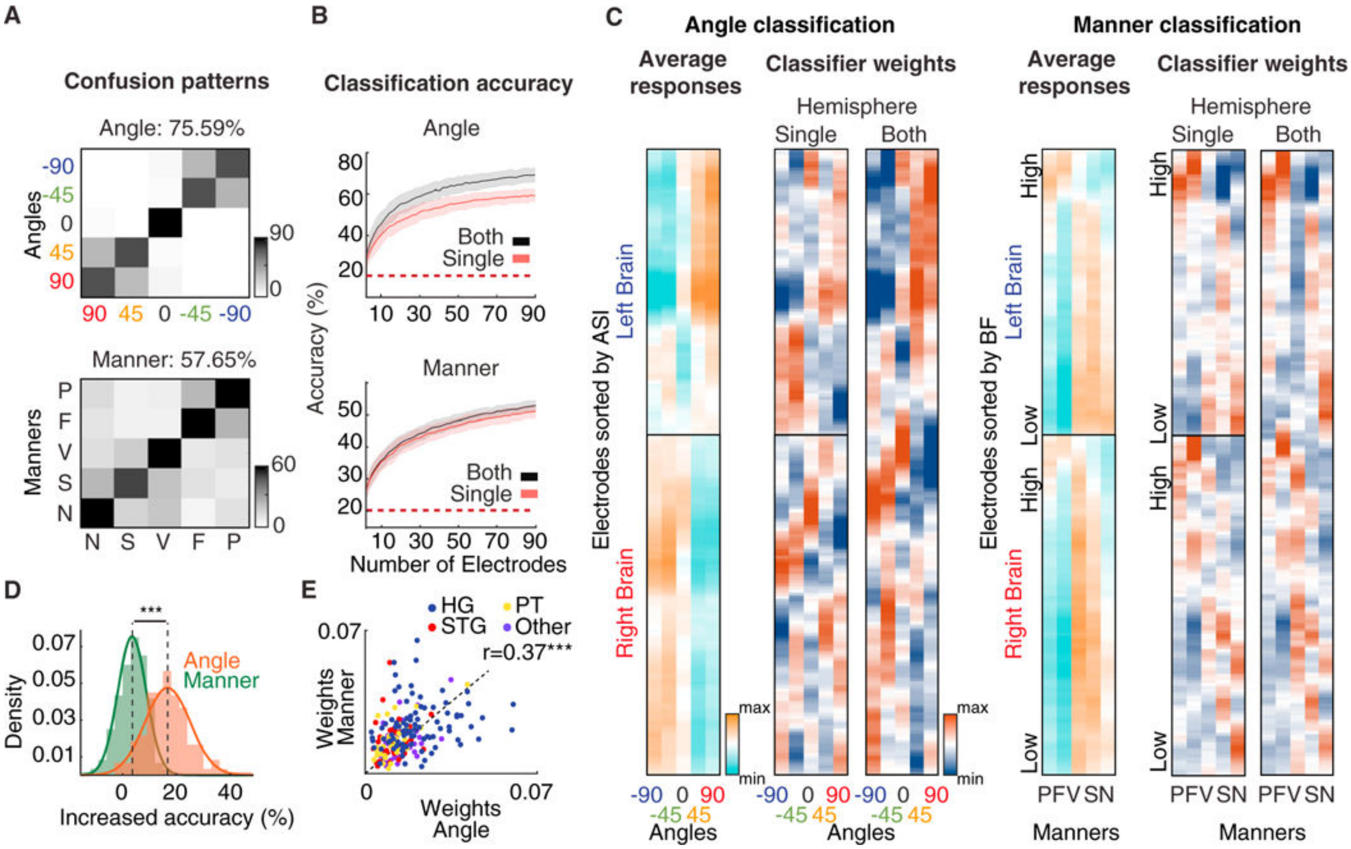


Figure 4. Joint Population Decoding of Spatial and Phonetic Features

(A) Confusion patterns for classifying angles (top) and manners of articulation (bottom) from all electrodes (n = 201).

(B) Mean classification accuracy for varying numbers of electrodes for angle classification (top) and manner classification (bottom). Error bars denote standard deviation, and colors denote whether electrodes from single (red) or both (black) brain hemispheres were used for classification.

(C) Average of *Z*-scored electrode responses (far left) separated by hemisphere of the brain and the weights assigned to electrodes by spatial (left) and manner (right) classifiers. Angle classifier weights are sorted by the ASI of electrodes, and manner classifier weights are sorted by the BF of electrodes.

(D) Percentage increase in classification accuracy from a single to both brain hemispheres for manner and angle classification.

(E) Scatterplot of the maximum weight given to each electrode by manner and angle classifiers, colored by electrode location.

***p < 0.001. See also Figure S5.

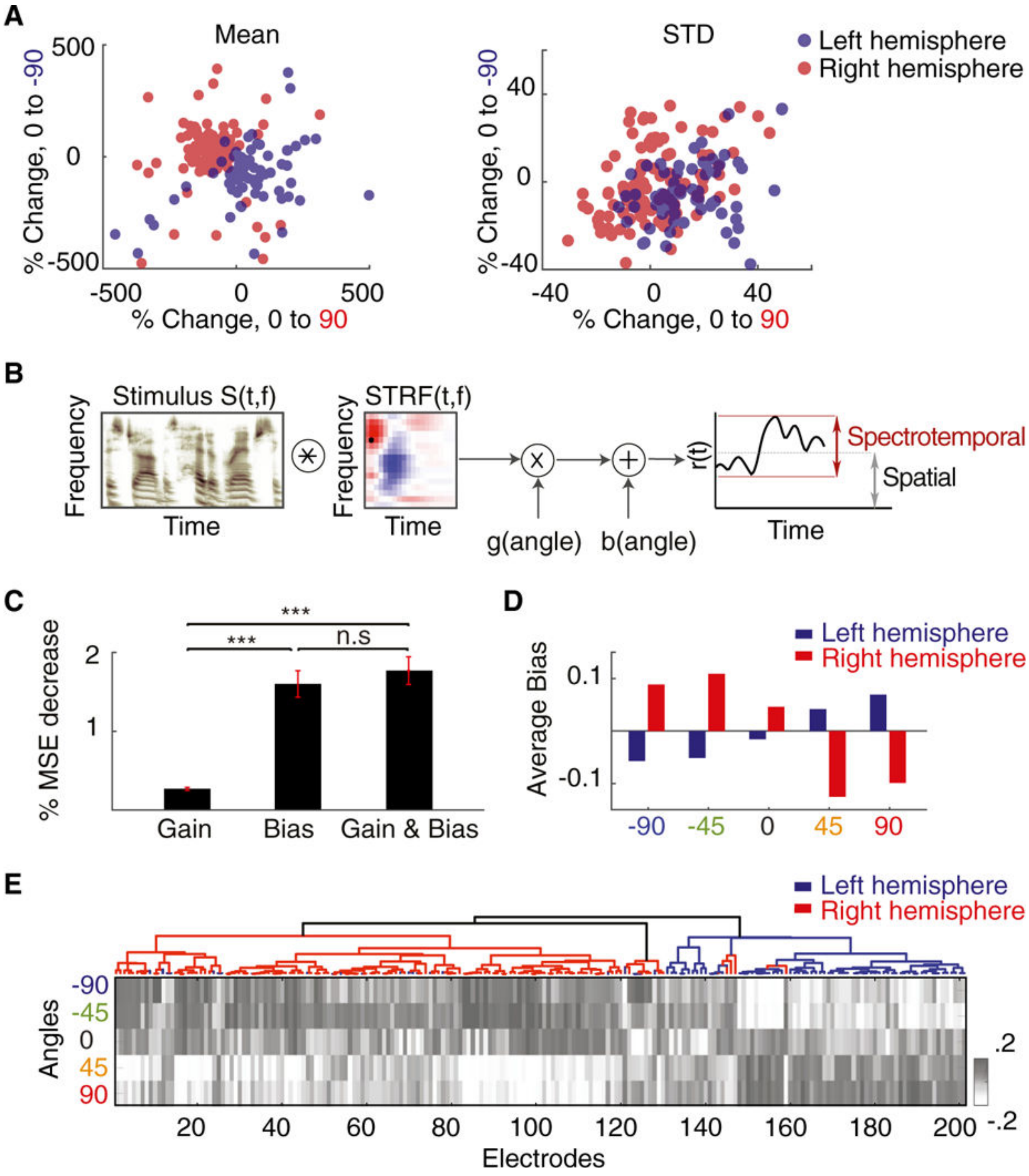


Figure 5. Mechanism of Joint Encoding of Spatial and Spectrotemporal Features at Individual Electrodes

(A)Scatterplot of percentage change of the mean (left) and standard deviation (right) of neural responses relative to the baseline (angle 0°) for angle 90° (x axis) and angle —90° (y axis) forall electrodes.

(B) Proposed computational model. The auditory spectrogram of speech, *S(t,f),* is convolved with the electrode’s STRF and then modulated by a gain and a bias factor that depend on the direction of sound.

(C) Mean reduced prediction error of the neural responses relative to baseline (non-spatial STRF) when modulating the gain, the bias, or both in the model. The error bars indicate SE.
(D) Average bias values for five angles from all speech-responsive electrodes colored by right (red) and left (blue) brain hemispheres.
(E) Mean response level (bias) values for five angles from each speech-responsive electrode arranged by ASI and colored by right (red) and left (blue) brain hemispheres.
***p < 0.001. See also Figure S6.