



Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli

Patrick W. Hullett,^{1,2,3}  Liberty S. Hamilton,^{2,4} Nima Mesgarani,^{2,4}  Christoph E. Schreiner,^{1,2,3} and Edward F. Chang^{1,2,4}

¹University of California Berkeley and San Francisco Joint Graduate Group in Bioengineering, ²Center for Integrative Neuroscience, ³Department of Otolaryngology—Head and Neck Surgery, and ⁴Department of Neurological Surgery, School of Medicine, University of California, San Francisco, San Francisco, California 94158

The human superior temporal gyrus (STG) is critical for speech perception, yet the organization of spectrotemporal processing of speech within the STG is not well understood. Here, to characterize the spatial organization of spectrotemporal processing of speech across human STG, we use high-density cortical surface field potential recordings while participants listened to natural continuous speech. While synthetic broad-band stimuli did not yield sustained activation of the STG, spectrotemporal receptive fields could be reconstructed from vigorous responses to speech stimuli. We find that the human STG displays a robust anterior–posterior spatial distribution of spectrotemporal tuning in which the posterior STG is tuned for temporally fast varying speech sounds that have relatively constant energy across the frequency axis (low spectral modulation) while the anterior STG is tuned for temporally slow varying speech sounds that have a high degree of spectral variation across the frequency axis (high spectral modulation). This work illustrates organization of spectrotemporal processing in the human STG, and illuminates processing of ethologically relevant speech signals in a region of the brain specialized for speech perception.

Key words: functional organization; human STG; human superior temporal gyrus; modulation tuning; modulotopic; spectrotemporal processing

Significance Statement

Considerable evidence has implicated the human superior temporal gyrus (STG) in speech processing. However, the gross organization of spectrotemporal processing of speech within the STG is not well characterized. Here we use natural speech stimuli and advanced receptive field characterization methods to show that spectrotemporal features within speech are well organized along the posterior-to-anterior axis of the human STG. These findings demonstrate robust functional organization based on spectrotemporal modulation content, and illustrate that much of the encoded information in the STG represents the physical acoustic properties of speech stimuli.

Introduction

A remarkable array of imaging, electrophysiological, and functional lesion studies have implicated the human superior tempo-

ral gyrus (STG) in speech processing and perception. However, organization of basic spectrotemporal processing during speech perception is not well understood (Boatman et al., 1997; Binder et al., 2000; Boatman, 2004; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). While tonotopy is a prominent organizing principle in the ascending auditory pathway and has been investigated in STG (Talavage et al., 2004; Striem-Amit et al., 2011; Moerel et al., 2012; Nourski et al., 2014), additional characterization of spectrotemporal processing in the context of speech processing is less well defined. Here, using speech stimuli to assess broad organizational principles during natural speech

Received May 2, 2015; revised Nov. 27, 2015; accepted Jan. 5, 2016.

Author contributions: P.W.H., L.S.H., N.M., C.E.S., and E.F.C. designed research; P.W.H., L.S.H., N.M., C.E.S., and E.F.C. performed research; P.W.H. and L.S.H. analyzed data; P.W.H., L.S.H., N.M., C.E.S., and E.F.C. wrote the paper.

This work was supported by National Institutes of Health Grants DP2-OD00862 (to E.F.C.), R01-DC012379 (to E.F.C.), R01-DC02260 (to C.E.S.), and F32-DC014192 (to L.S.H.), and the McKnight Foundation (to E.F.C.). E.F.C. is a New York Stem Cell Foundation—Robertson Investigator. This research was also supported by The New York Stem Cell Foundation. The authors declare no competing financial interests. We thank Craig Atencio, Brian Malone, Bryan Seybold, and Erik Edwards for their insightful comments on the manuscript.

Correspondence should be addressed to Edward Chang, University of California, San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94158. E-mail: changed@neurosurg.ucsf.edu.

DOI:10.1523/JNEUROSCI.1779-15.2016

Copyright © 2016 the authors 0270-6474/16/362014-13\$15.00/0

perception, we examine fundamental spectrotemporal response parameters across the STG.

A major goal of sensory neuroscience is to understand how sensory systems encode natural stimuli and, in particular, how spectrotemporal processing is organized in STG during speech processing. Traditionally sensory encoding has been studied with simple parameterized stimuli. However, recently, there has been increasing use of natural stimuli to study sensory encoding (Theunissen et al., 2000; David et al., 2004; Sharpee et al., 2006; Talebi and Baker, 2012). Part of the motivation for this is increasing evidence that sensory processing is adapted to statistics of behaviorally relevant stimuli. Data suggest receptive fields and response properties of neurons are matched to the statistics of natural inputs to maximize efficiency of information transmission (Dong and Atick, 1995; Rieke et al., 1995; Dan et al., 1996; Olshausen and Field, 1996; Hsu et al., 2004). In higher-order areas, systems become more selective for natural stimuli and less responsive to synthetic or noise stimuli (Theunissen et al., 2000; Wilkinson et al., 2000; Felsen et al., 2005; Talebi and Baker, 2012). In areas that respond broadly to both synthetic and natural stimuli, natural inputs can push neurons into different operating ranges and activate nonlinearities that evoke response properties not present with synthetic stimuli alone (David et al., 2004; Talebi and Baker, 2012). Finally, neurons adapt on a moment-by-moment basis to match their response properties to statistics of the input stimulus, thus making descriptions of sensory encoding inherently specific to the statistics and class of stimuli used for characterization (Smirnakis et al., 1997; Brenner et al., 2000; Fairhall et al., 2001; Sharpee et al., 2006). Thus, in situations where the question of interest is to understand encoding of natural stimuli, it is advantageous to use natural inputs, such as speech, to characterize the system.

Given the relevance of STG in speech processing, we are interested in characterizing spectrotemporal processing in the context of natural speech. However, characterization of spectrotemporal processing based on speech stimuli is difficult due to the statistically biased and highly correlated structure of natural signals. Recently, a number of techniques have been developed to address this bias (Theunissen et al., 2001; Paninski, 2004; Sharpee et al., 2004; David et al., 2007). Here we use maximally informative dimension (MID) analysis, an information-based method designed to lift the requirement of statistically tractable stimuli and allow the use of more complex, but ethologically relevant, natural stimuli for receptive field characterization (Sharpee et al., 2004; Atencio et al., 2008).

By using electrocorticography (ECoG) to record speech-driven activity, we compute spectrotemporal receptive fields (STRFs) using MID analysis and assess the organization of spectrotemporal processing in human STG during natural speech processing. While synthetic broad-band stimuli did not yield strong activation of the STG, STRFs could be reconstructed from vigorous responses to speech stimuli. We find that the human STG displays a robust anterior–posterior spatial distribution of spectrotemporal tuning characterized by tuning for high temporal and low spectral modulation speech features posteriorly and high spectral and low temporal modulation speech features anteriorly. This work further defines organization of spectrotemporal processing of speech in human STG, and illuminates processing of ethologically relevant speech signals in a region of the brain specialized for speech perception.

Materials and Methods

Participants and neural recordings. Subdural ECoG arrays (interelectrode distance, 4 mm) were placed unilaterally in eight patient volunteers [three right hemisphere (one female/two male), five left hemisphere (three female/two male)] undergoing a neurosurgical procedure for the treatment of medication refractory epilepsy. Seven of eight participants were native English speakers; all were fluent in English. All participants had normal hearing. None had communication deficits. All experimental protocols were approved by the University of California, San Francisco Institutional Review Board and Committee on Human Research. Location of array placement was determined by clinical criteria alone. Participants were asked to passively listen to 15–25 min of natural speech while ECoG signals were recorded simultaneously. Some participants also passively listened to dynamic moving ripple (DMR) or temporally orthogonal ripple combination (TORC) stimuli (see below). Signals were amplified and sampled at 3052 Hz. After rejection of electrodes with excessive noise or artifacts, signals were referenced to a common average and the high-gamma band (70–150 Hz) was extracted as the analytic amplitude of the Hilbert transform (Crone et al., 2001; Chang et al., 2011). Signals were subsequently downsampled to 100 Hz. The resulting signal for each electrode was z-scored based on the mean and SD of activity during the entire block.

Stimuli

Speech stimuli. Speech stimuli were delivered binaurally through free-field speakers at ~70 dB average sound pressure level. The frequency power spectrum of stimuli spanned 0–8000 Hz. The stimulus set consisted of prerecorded (2–4 s) sentences from the phonetically transcribed Texas Instruments/Massachusetts Institute of Technology (TIMIT) speech corpus with 1 s silent intervals between each sentence presentation (Garofolo et al., 1993). Each participant was presented 484–499 sentences. The speech corpus included 286 male and 116 female speakers, with 1–3 sentences spoken per speaker, and unique lexical content for each sentence. Spectrogram representations of speech stimuli were generated using a cochlear model of auditory processing (Yang et al., 1992).

Synthetic stimuli. DMR stimuli were presented to four subjects (EC63, GP30, GP31, and GP33) at 70 dB average sound pressure level. Of these, only GP31 was included in the functional organization analysis. The other three participants had less predictive or noncontiguously spaced MID-based STRFs along STG and thus were less useful for characterizing organization of spectrotemporal processing (but still of sufficient data quality to assess responsiveness). The DMR was composed of a bank of sinusoidal carriers whose amplitude is modulated over time with a spectrotemporal ripple envelope (Depireux et al., 2001; Escabi and Schreiner, 2002). The amplitude distribution of the envelope was Gaussian and had a maximum modulation depth of 40 dB. The DMR is broadband (500–20,000 Hz) and consisted of ~50 sinusoidal carriers per octave with randomized phase. The spectral and temporal modulation frequency parameters defined the characteristics of the ripple modulation envelope at any given point in time. The spectral modulation parameter defines the number of spectral peaks per octave. The temporal modulation parameter defines the speed and direction of the peak's change. Spectral and temporal modulation parameters were varied randomly and independently during the stimulus. The spectral modulation parameter varied between 0 and 4 cycles/octave (maximum rate of change, 1 Hz) and the temporal modulation rate parameter varied between –35 and +35 Hz (rate of change, ≤ 3 Hz). TORC stimuli were presented to two subjects (EC2 and EC28) at 70 dB average sound pressure level. TORC stimuli are generated by modulating broadband white noise (frequency content, 250–8000 Hz) with a combination of 12 temporally orthogonal ripples (Klein et al., 2000). The spectral modulation of ripples ranged between 0 and 1.4 cycles/octave and the temporal modulation of ripples ranged between 4 and 48 Hz.

Analysis

STRFs. STRFs were computed with two different methods designed specifically for use with naturalistic stimuli—MID analysis and normalized reverse correlation (Theunissen et al., 2001; Sharpee et al., 2004). To compute STRFs using MID analysis, a gradient ascent procedure was

used to search for the receptive field that maximizes the Kullback–Leibler divergence between the raw distribution of STRF–stimulus projection values and the distribution of STRF–stimulus projection values weighted by the magnitude of the response. STRF estimates based on normalized reverse correlation (normalization based on the stimulus autocorrelation matrix) were computed using ridge regression with open source code available at <http://strfpak.berkeley.edu/>. Regularization was controlled by fitting a tolerance hyperparameter via cross-validation (David et al., 2007). STRFs were computed with both methods on the same estimation set (90% of the total data) and cross-validated on the same test set, which was withheld from the estimation process (10% of the data).

Modulation tuning. To characterize modulation tuning-based organization, the modulation transfer function (MTF) for each site was computed by taking the magnitude of the two-dimensional Fourier transform ($\mathfrak{F}_2\{\cdot\}$) of each STRF according to the following equation:

$$MTF(\omega_t, \omega_s) = |\mathfrak{F}_2\{STRF(t, f)\}|$$

Where (t, f) are time and frequency and (ω_t, ω_s) are temporal and spectral modulation, respectively. The best spectrotemporal modulation (bSTM) is defined as the peak of the MTF (see Fig. 3A). The sign of the best temporal modulation determines the drift direction of spectral content within each ripple (increasing or decreasing; see Fig. 3B). Similar to previous work, we are interested in the magnitude of temporal modulation and therefore take the absolute value of best temporal modulations (Langers et al., 2003; Santoro et al., 2014). The ensemble MTF was computed by normalizing each MTF to have a sum of 1, then computing the average MTF across all sites and all participants.

Spatial analysis. Permutation tests were used to assess whether STRF-derived parameters, such as bSTM, MTF cluster type, or best frequency (BF), were locally organized. Each test determined whether the average level of similarity between a site and its neighbors would be expected if the true underlying spatial organization were random. For each site, we compute difference between a site and its neighbors (anterior, posterior, dorsal, and ventral), take the absolute value, and then compute the average. This value is computed for each site and then averaged across the whole map to compute the mean neighborhood similarity index. The map is then randomly permuted 10,000 times and the map neighborhood similarity index recomputed on each permutation to generate a distribution of randomized map neighborhood similarity indices. The true neighborhood similarity index is compared with the randomized map neighborhood similarity index distribution to assess the level of significance. To determine significance values for each spectrotemporal modulation map (two parameters at each site: one for best spectral modulation, one for best temporal modulation), the above procedure was repeated with a two-parameter neighborhood similarity metric. The two-parameter neighborhood similarity metric was generated by normalizing best temporal modulation values and best spectral modulation values by their respective maxima so both sets of data had a range of 0–1. For each parameter at each site, we compute difference between a site and its neighbors (anterior, posterior, dorsal, and ventral), take the absolute value, and then compute the average. This value for each parameter was then added together to generate the two-parameter similarity metric. This metric was averaged across sites to generate the two-parameter neighborhood similarity index for the entire map. While there are many metrics that could potentially be used to quantify organization, we choose to use neighborhood similarity because of its generality in that previously described forms of organization (linear gradients, nonlinear gradients, clusters, or modules) have neighborhood similarity as a common feature.

Group analysis was achieved by combining data across subjects in a common coordinate system defined by anatomical landmarks. The x -axis was defined as a line that runs parallel to the long axis of STG along its dorsal–ventral midpoint. The y -axis was defined as a line orthogonal to the x -axis with the origin aligned with the anterior temporal pole (see Fig. 6D, red). After coordinates of ECoG sites were defined, data were binned at 4 mm resolution consistent with the interelectrode distance of the ECoG arrays (4 mm interelectrode distance). Tonal and spectrotemporal modulation maps represent topographic distributions of BF and preferred spectrotemporal modulation across the two-dimensional

surface of the STG. To compute the gradient of such maps, a two-dimensional plane was fit to the data using linear (planar) regression. The direction of steepest angle of the plane is taken to be the gradient of the topographic map (Baumann et al., 2011). The gradient direction is specified as an angle counterclockwise from the x -axis of the coordinate system that runs in the anterior–posterior direction along the long axis of STG. For spectrotemporal modulation maps, which represent a map of two parameters, spectral and temporal modulation values were normalized by their group maxima so each set of data has a range of 0–1. Gradients were calculated for the spectral modulation map and temporal modulation map independently and then averaged to determine the gradient for the joint spectrotemporal modulation map.

Results

STRF maps

Given the role of STG in speech processing, we are interested in characterizing functional organization in STG based on responses to natural speech stimuli. After ECoG array placement, participants were asked to passively listen to 15–25 min of natural speech, which consisted of prerecorded (2–4 s) sentences from the phonetically transcribed TIMIT speech corpus (Garofolo et al., 1993). After data collection, an STRF was computed off-line from the local field potential signal at each cortical site to generate an STG STRF map in each subject (Fig. 1A). STRFs were computed using the high-gamma (70–150 Hz) band of ECoG recordings (Crone et al., 2001), which correlates with spiking activity (Ray and Maunsell, 2011) and spike-based tuning properties in the midlaminar auditory cortex (Steinschneider et al., 2008). To compute STRFs, two different methods designed for use with natural signals were used: MID analysis (Sharpee et al., 2004) and normalized reverse correlation (Theunissen et al., 2001). Using these two methods, we computed MID and normalized reverse correlation-based STRFs using responses to speech stimuli and evaluated their performance through cross-validation. While MID-based STRFs were generally similar to normalized reverse correlation-based STRFs, MID-based STRFs produced higher prediction values and were therefore used in the remainder of the analysis (Fig. 1C; $p < 0.001$, Wilcoxon signed-rank test, mean percentage increase in prediction: $19.0 \pm 1.9\%$ SEM).

For purposes of studying organization of spectrotemporal processing, it is necessary to restrict analysis to STRFs that characterize the underlying spectrotemporal processing. Similar to previous work, only sites that predict $\geq 5\%$ of the variance in the response were included in the analysis (Fig. 1D,E; Kim and Doupe, 2011). This set of STRFs showed relatively high prediction performance (mean $r = 0.48 \pm 0.12$ SD; Fig. 1D), comparable to STRF prediction values from lower-order areas (Calabrese et al., 2011; Kim and Doupe, 2011). These highly predictive STRFs are based on the acoustic properties of speech alone and the model does not take into account other aspects of speech, such as semantic meaning. This indicates that the physical acoustic properties of speech are a major component of the encoded information in STG and the computed speech-based STRFs provide a good characterization of the underlying spectrotemporal processing of speech. Similar to STRFs found in lower-order auditory areas, STRFs within STG showed clear contiguous excitatory and inhibitory regions and structure characteristic of “temporal” and “spectral” STRFs (Fig. 2; Nagel and Doupe, 2008; Atencio et al., 2008). Temporal STRF aspects exhibit short excitatory regions followed by short inhibitory regions (Fig. 2, ●). These types of STRFs are characteristic of sites tuned for rapid temporal modulations in sound energy that occur at the onset or offset of sound and within many consonants. Spectral STRF aspects are characterized by temporally long excitatory regions

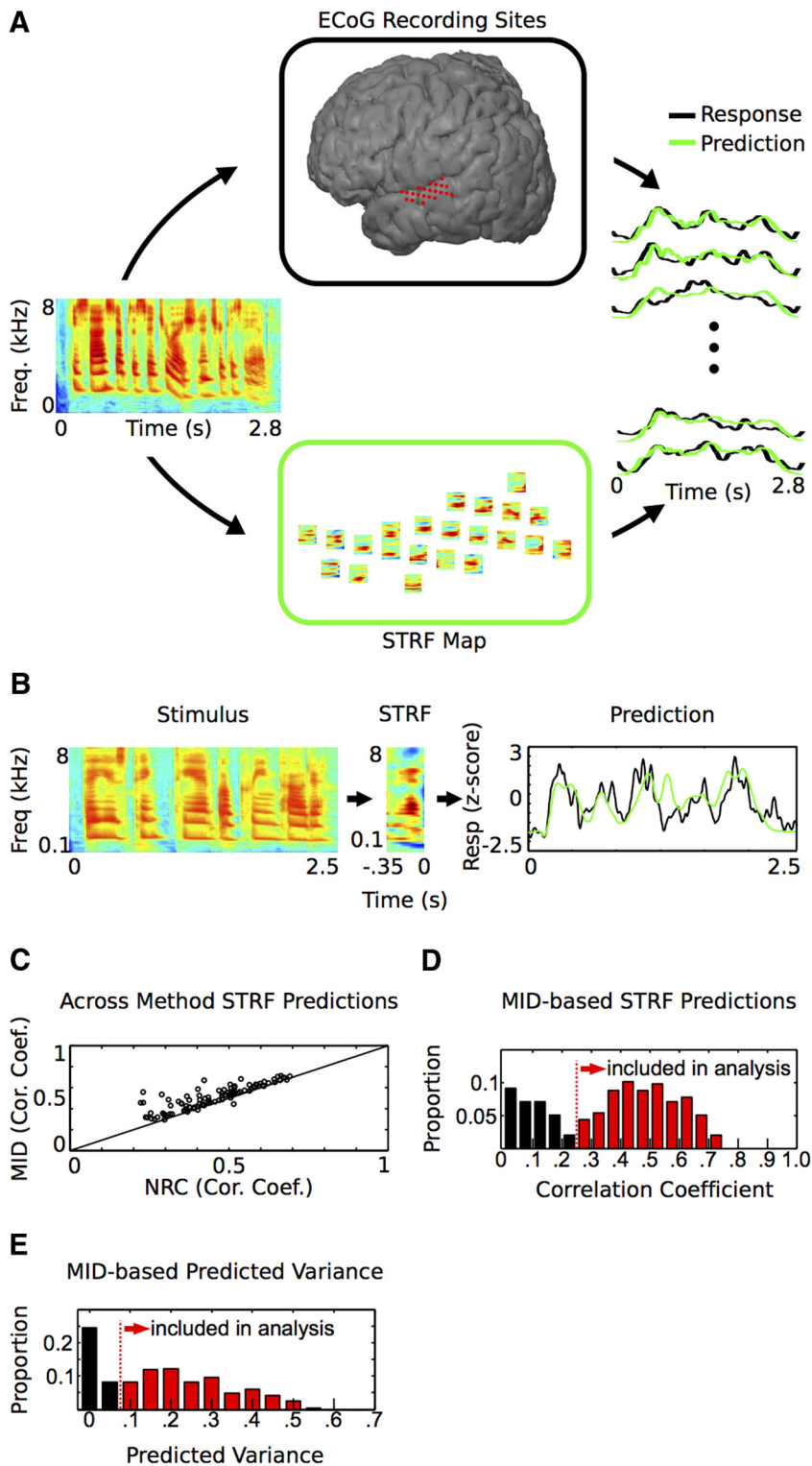


Figure 1. Experimental approach and the STRF. **A**, Experimental approach. An STRF was computed off-line for each ECoG electrode site (top, center) to generate a corresponding STRF map (bottom, center). The STRF describes the spectrotemporal structure in the stimulus that drives activity at a particular site. On the right is a subset of measured and predicted responses for the sentence “He sized up the situation and shook his head” (spectrogram at left). **B**, An STRF and the predicted and measured response for a single sentence. Predicted responses are obtained by convolving the stimulus with the STRF and are proportional to the similarity between the spectrotemporal content in the stimulus and the receptive field. **C**, Comparison of two methods used to compute STRFs. MID-based STRFs show higher predictive performance compared with normalized reverse correlation (NRC)-based STRFs (mean percentage increase in prediction: $19.0 \pm 1.9\%$ SEM, $p < 0.001$, Wilcoxon signed-rank test). **D**, MID-STRF Pearson correlation coefficient prediction values for all STG sites. **E**, MID-STRF predicted variance values for all sites. Sites with $>5\%$ of the variance predicted (red) were included in the analysis. MID, maximally informative dimension analysis; NRC, normalized reverse correlation analysis.

flanked by inhibition on one or both sides (Fig. 2, ■). These types of STRFs are characteristic of sites tuned to sound energy that fluctuates across frequency, but is relatively constant over time, such as the formant structure prevalent in vowel sounds. These data show that STG STRFs have characteristics similar to STRFs found in lower-order auditory areas and are predictive of responses to speech stimuli, demonstrating their ability to characterize spectrotemporal processing of speech in STG.

Organization of spectrotemporal processing of speech in human STG

Given that the STRF characterizes spectrotemporal processing at each ECoG site, the STRF map characterizes how processing varies across STG. However, as Figure 2 shows, the spectrotemporal structure within each STRF is relatively complex, making it difficult to empirically visualize or quantify organizational aspects from the raw STRF maps. One alternative representation of the STRF that contains identical spectrotemporal information (minus phase), but which exhibits visually less complex structure, is the MTF (Fig. 3). The MTF characterizes spectrotemporal modulation tuning of each STRF with a structure that is often more localized in modulation space than the corresponding STRF structure within the time–frequency domain (Fig. 3A, compare STRF, MTF). Additionally, STRF features that are shifted in frequency or time, but have the same overall spectrotemporal structure, have the same MTF representation because phase information is discarded. Using this alternative, but less complex and equivalent representation of the STRF, we examined the distribution of MTFs within the STG. A map of MTFs for one participant (EC6) is shown in Figure 4. The “temporal” and “spectral” STRFs shown in Figure 2 have corresponding temporal and spectral MTFs in Figure 4. Temporal transfer functions (●) are characterized by tuning shifted away from the vertical midline toward high temporal modulations (energy shifted away from the vertical axis) and sites in the anterior STG are tuned for slow temporal modulations, but high spectral modulations (energy falling along the vertical axis and shifted upward).

To more thoroughly characterize this apparent nonrandom distribution of spectrotemporal processing, representative MTFs of STG were first identified by k-means clustering of all MTFs from all participants. Silhouette criterion values were used to identify the number of data clusters that maximize intracluster similarity compared with similarity between neighboring clusters. The centroid MTFs from this analysis are shown in Figure 5A, along with their 50% energy contours. Compared with the average MTF for STG (Fig. 5E), centroid MTFs occupy discrete regions of modulation tuning space in an orderly fashion from high spectral/low temporal modulation tuning (Fig. 5A, left MTF with red contour) to high temporal/low spectral modulation tuning (Fig. 5A, right MTF with yellow contour). To examine the distribution of overall spectrotemporal modulation tuning across STG, a group MTF map was computed and each MTF within the map was classified according to its cluster identity (Fig. 5B).

Cluster identities from the group map are plotted in Figure 5C. This map shows significant local organization of the MTF type ($p < 1.0 \times 10^{-5}$, neighborhood similarity permutation test) and illustrates anterior-to-posterior organization of modulation tuning along the STG in which high spectral modulation/low temporal modulation tuned MTFs are located anteriorly (red). By contrast, high temporal modulation/low spectral modulation tuned MTFs are located posteriorly (yellow). This is further quantified by the average location of each MTF type along the anterior-to-posterior extent of the STG (Fig. 5D). Overall, this shows that the STG is tuned for temporally fast-changing speech sounds that have relatively constant energy across the frequency axis (low spectral modulation) in the posterior STG while the anterior STG is tuned for temporally slow-changing speech sound that have a high degree of spectral variation across the frequency axis (high spectral modulation).

To further illustrate the organized distribution of spectrotemporal processing along the anterior–posterior extent of the STG, we plot the distribution of MTF peak values across the STG. The peak of the MTF defines the bSTM, and in this context represents the dominant spectrotemporal modulation in the feature each site is tuned to (as captured by the STRF). It should be noted that examining the distribution of MTF peaks (bSTM values) derived for natural speech does not imply that the STG will respond to individual spectrotemporal modulations. Rather, the localized nature of modulation energy in MTFs makes the peak a good descriptor of the overall MTF tuning. The distribution of bSTMs can, therefore, be used to further characterize overall organization of spectrotemporal modulation tuning in the STG. The distribution of bSTMs for all sites and from all participants is shown in Figure 6A. The relationship between best temporal modulation and best spectral modulation shows a trade-off pattern in which spectral modulation decreases with increasing temporal modulation. This relationship between spectral and temporal modulation tuning has been observed previously in other auditory areas, including the human primary auditory cortex (Schönwiesner and Zatorre, 2009), the cat primary auditory cortex, the cat medial geniculate body, the cat inferior colliculus (Miller et



Figure 2. Participant EC6 cortical STRF map. **A**, STRF map for participant EC6 (STRFs calculated with MID analysis). STG STRFs showed clear contiguous excitatory and inhibitory regions and structure characteristic of STRFs found in other regions. Representative temporal STRFs (●, tuned to quick onsets or offsets), and spectral STRFs (■, tuned to constant sound energy that fluctuates across frequency), are shown. LS, Lateral sulcus; STS, superior temporal sulcus; MTG, medial temporal gyrus; CS, central sulcus.

al., 2002; Rodríguez et al., 2010; Atencio and Schreiner, 2012), and in models of auditory processing that predict receptive field structure based on efficient coding hypotheses (Carlson et al., 2012).

To examine the spatial distribution of bSTM tuning across the human STG, we plot bSTM values at their corresponding cortical position within each subject (Fig. 6A, color scale, *B*). Most participants show a significant degree of nonrandom local organization that is consistent across participants (EC6, $p < 1.0 \times 10^{-5}$; GP31, $p = 5.0 \times 10^{-4}$; EC36, $p = 0.029$; EC28, $p = 4.5 \times 10^{-5}$; EC53, $p = 0.015$; EC58, $p = 0.10$; EC56, $p = 0.06$; EC2, $p = 0.42$; two-parameter neighborhood similarity permutation test). Posterior regions of the STG show primarily high temporal modulation/low spectral modulation tuning (blue), while more anterior regions show primarily high spectral modulation/low temporal modulation tuning (green). With respect to interhemispheric differences, there was no significant difference in best temporal modulation tuning (left: mean, 0.89 ± 0.8 Hz; right: mean, 0.72 ± 0.8 Hz, $p = 0.17$, Wilcoxon rank-sum test) and a subtle difference in best spectral modulation tuning (left: mean, 0.14 ± 0.17 cycles/octave; right: mean, 0.08 ± 0.13 cycles/octave, $p = 0.015$, Wilcoxon rank-sum test). To examine data across subjects, we computed a group modulation tuning map, which is shown in Figure 6D. Similar to individual subject maps, the group map shows primarily high temporal modulation/low spectral modulation tuning posteriorly (blue) and high spectral modulation/low temporal modulation tuning anteriorly (green, $p < 1.0 \times 10^{-5}$, neighborhood similarity permutation test). The dominant gradient within the group modulation map runs in the posterior-to-anterior direction ($+176^\circ$ counterclockwise from the 3 o'clock position; Fig. 6D) nearly parallel with the anterior–posterior axis of the STG. Again, these data represent organized spectrotemporal processing along the STG, in which posterior regions are tuned to temporally fast sounds with relatively constant energy along the spectral axis (low spectral modulation) and more anterior regions are tuned for temporally slower sound with high variation in energy along the spectral axis (high spectral modulation). To further characterize modulation organization,

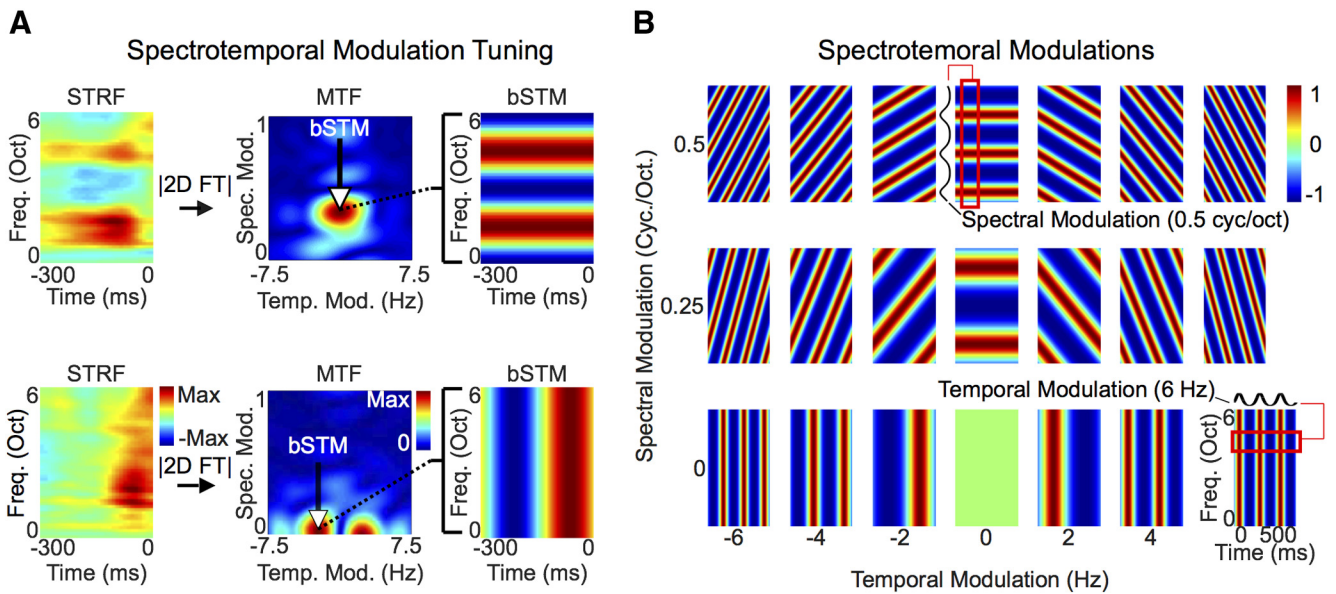


Figure 3. The modulation transfer function and best spectrotemporal modulation (bSTM). **A**, Computation of the modulation transfer function (MTF). The MTF is derived as the magnitude of the two-dimensional Fourier transform of the STRF. It characterizes spectrotemporal modulation tuning for each site. Like the BF of a frequency tuning curve, the peak of the MTF defines the bSTM and represents a good descriptor of the overall MTF given the localized nature of modulation tuning within each MTF. For the site with the STRF shown at the top, the MTF indicates that high spectral modulations and low temporal modulations drive activity at that site. In contrast, the site below has a bSTM at high temporal modulations and low spectral modulations, indicating that the site is driven by changes in temporal and not spectral energy. **B**, Ordered array of spectrotemporal modulations as a function of their temporal and spectral modulation parameters. Spectrotemporal modulations represent the envelope “frequency” components of the spectrogram. Any spectrogram can be reconstructed exactly by a weighted sum of spectrotemporal modulations since they form a complete orthonormal basis of functions.



Figure 4. Participant EC6 cortical modulation tuning map. **A**, Modulation tuning map for participant EC6. Each MTF is derived from the corresponding STRF. Representative temporal (●) and spectral (■) MTFs are shown. Although MTFs and STRFs contain equivalent information about spectrotemporal processing (except for phase information, which is discarded in MTFs), the overall structure of MTFs is less complex than that of STRFs. As shown, sites in the posterior STG are tuned for high temporal modulations (energy shifted away from the vertical midline) and sites in the anterior STG are tuned for slow temporal modulations and high spectral modulations (energy falling along the vertical midline and shifted upward). LS, Lateral sulcus; STS, superior temporal sulcus; MTG, medial temporal gyrus; CS, central sulcus.

average best temporal modulation and best spectral modulation tuning are plotted as a function of distance along the anterior–posterior extent of the STG (Fig. 6E,F, red). Additionally, modulation tuning was plotted as a function of map distance after individual maps have been aligned by their individual gradients (Fig. 6E,F, blue and green lines). This shows a transition from high temporal modulation to high spectral modulation along the

posterior–anterior aspect of the STG. A similar but less robust distribution of modulation tuning was also seen for STRFs based on normalized reverse correlation (Fig. 6E,F, insets). Further analysis showed no compelling additional organization within the maps based on inclusion of the temporal modulation sign (positive vs negative), no organized distribution of prediction performance that paralleled the distribution of spectrotemporal modulation tuning, and no difference in prediction performance between “spectral” and “temporal” STRFs (r value mean, 0.50 ± 0.13 vs 0.48 ± 0.11 , respectively; $p = 0.23$, Wilcoxon rank-sum test; “spectral” and “temporal” STRFs defined as having bSTM values above or below the diagonal in Fig. 5A, respectively). Collectively, these data reveal spatially organized spectrotemporal processing of speech in human STG in which tuning varies from high temporal/low spectral modulation tuning in the posterior STG to high spectral/low temporal modulation tuning in the middle STG. This organization is reminiscent of the posterior-to-anterior distribution of temporal modulation tuning found at the lateral aspect of the superior temporal plane in macaques as may have been predicted for directly adjacent areas (Baumann et al., 2015).

To determine whether organization of spectrotemporal modulation tuning within the STG is present during processing of more traditional stimuli used to characterize modulation tuning,

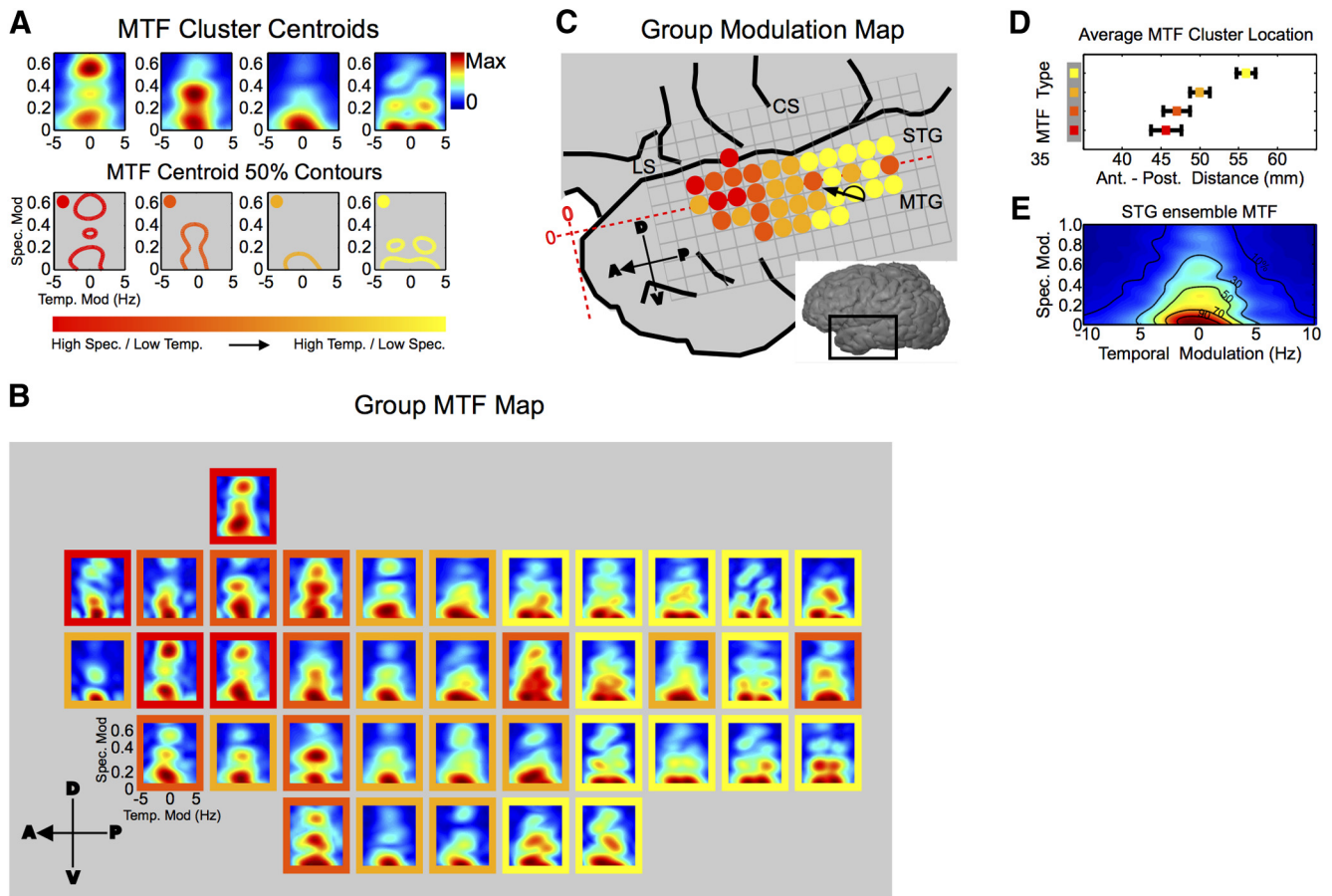


Figure 5. Organization of MTFs in STG. **A**, K-means cluster centroids generated from all MTFs across all participants. Their respective 50% contours are shown below. The overall tuning within an individual MTF centroid is fairly well localized. The collection of MTF centroid types span modulation space from high spectral/low temporal regions (red, left) to high temporal/low spectral regions (yellow, right) as shown by the 50% contours. **B**, Group MTF map. The map represents the average MTF across participants at each STG position. Only locations with ≥ 2 MTFs contributing to the average are included. Each MTF within the map is color-coded by its cluster membership. **C**, MTF cluster identity map. MTF cluster identities from **B** are plotted. The cluster identity map shows a transition from high spectral/low temporal MTFs anteriorly (red) to high temporal/low spectral MTFs posteriorly (yellow) and a significant degree of local organization ($p < 1.0 \times 10^{-5}$, neighborhood similarity permutation test). **D**, Average MTF cluster distance along the anterior–posterior extent of STG. Distances are measured from the anterior temporal pole illustrated in **C** (red horizontal axis). High spectral/low temporal MTFs are located anteriorly (red). High temporal/low spectral MTFs are located posteriorly (yellow, error bars represent SEM). **E**, Ensemble MTF (population average) for STG. Contour lines represent percentage maximum. LS, lateral sulcus; STG, superior temporal gyrus; MTG, medial temporal gyrus; CS, central sulcus.

DMR stimuli (Depireux et al., 2001; Escabi and Schreiner, 2002) were presented to four subjects while recording from the STG. Consistent with previous work (Schönwiesner and Zatorre, 2009), these synthetic stimuli did not sufficiently activate the human STG (Fig. 7). Additionally, TORC stimuli, a variant of moving ripple stimuli (Klein et al., 2006), were presented in two participants without significant sustained activation of the STG (data not shown). These data are consistent with only minimal temporal lobe activation outside of the superior temporal plane by nonspeech stimuli as seen by Overath et al. (2015) and Schönwiesner and Zatorre (2009) and implies a degree of specificity of the organization demonstrated here for speech in that more traditional nonspeech synthetic stimuli do not drive robust activity in the STG.

Topography of spectral tuning

Last, we analyzed the spatial distribution of frequency tuning in human STG to examine the relationship between spectral frequency and modulation tuning in this region. To examine the distribution of spectral tuning, two metrics derived from the STRF were used to characterize BF. The first metric is defined as the peak excitatory value of the STRF (STRF-BF) and represents

the frequency with the largest gain at any time point within the STRF. The second metric is defined as the peak of the spectral receptive field (SRF). The SRF is obtained by summing across the temporal dimension of the STRF. The peak of the SRF (SRF-BF) takes into account all temporal lags of the STRF and identifies the frequency associated with the largest net gain. Both metrics of BF showed similar tuning across ECoG sites (Pearson correlation coefficient, 0.693) and the average difference in BF between an ECoG site and its directly adjacent neighbors (anterior, posterior, dorsal, ventral) was not significantly different between the two metrics, reflecting equal degrees of neighborhood similarity (mean difference in BF between adjacent sites: STRF-BF, 0.7 ± 0.65 octaves SE; SRF-BF, 0.84 ± 0.72 octaves SE, $p = 0.065$, Wilcoxon rank-sum test). Individual maps for the topography of STRF-BF values are shown in Figure 8B. For both metrics, only two of eight participants showed a significant degree of local organization based on neighborhood similarity (STRF-BF p /SRF-BF p : EC6, $p = 0.12/0.23$; GP31, $p = 0.004/0.007$; EC36, $p = 0.10/0.04$; EC28, $p = 0.02/0.75$; EC53, $p = 0.29/0.13$; EC58, $p = 0.27/0.07$; EC56, $p = 0.39/0.20$; EC2, $p = 0.11/0.25$; neighborhood similarity permutation test). Examination of the group map also showed a low degree of local organization based on

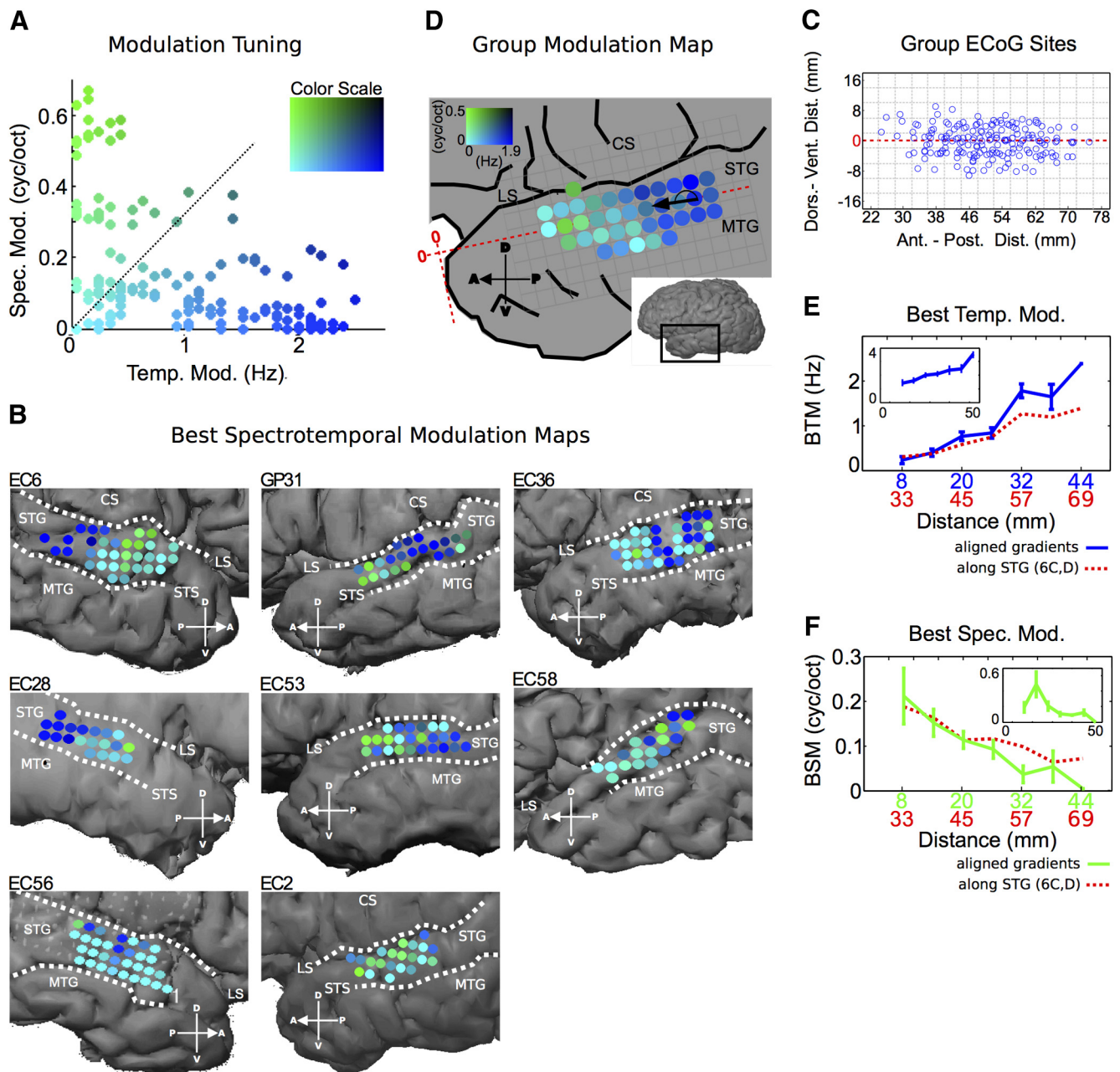


Figure 6. Organization of bSTM tuning in STG. **A**, bSTM tuning values from all participants. The distribution of bSTM values shows a particular relationship in which spectral modulation tuning decreases as temporal modulation tuning increases. **B**, Individual subject bSTM maps (**A**, color scale). Most participants show high temporal/low spectral modulation tuning posteriorly (blue), and high spectral/low temporal modulation tuning anteriorly (green) with significant nonrandom organization (EC6, $p < 1.0 \times 10^{-5}$; GP31, $p = 5.0 \times 10^{-4}$; EC36, $p = 0.029$; EC28, $p = 4.5 \times 10^{-5}$; EC53, $p = 0.015$; EC58, $p = 0.10$; EC56, $p = 0.06$; EC2, $p = 0.42$; two-parameter neighborhood similarity permutation test). **C**, Topographic ECoG site distribution along the anterior–posterior/dorsal–ventral extent of the STG. An example of the coordinate system used to measure distances is shown in **D** (red axis). Distances along the long axis of STG are measured from the anterior temporal pole. Distances along the short axis of STG are measured from the dorsal–ventral midpoint of the STG. **D**, Group spectrotemporal modulation tuning map. Only sites with data from ≥ 2 participants are included. High temporal/low spectral modulation tuned sites are located posteriorly (blue) and high spectral/low temporal modulation tuned sites are located anteriorly (green) with significant nonrandom organization and a mean gradient of $+176^\circ$ counterclockwise from the long axis of the STG ($p < 1.0 \times 10^{-5}$, neighborhood similarity permutation test). Data were binned at 4×4 mm resolution (same interelectrode distance as the ECoG array). **E**, Average best temporal modulation tuning as a function of distance along the dominant spectrotemporal modulation gradient. Individual subject maps were aligned by their gradients before averaging (blue, error bars represent SEM). Absolute distance is from the edge of data after maps have been aligned. The red function represents the raw average of temporal modulation tuning as a function of distance along the STG from the anterior temporal pole (**D**, horizontal red line; no map alignment by gradient before averaging). The inset represents data from normalized reverse correlation-based STRFs. **F**, Average best spectral modulation tuning as a function of distance along the dominant spectrotemporal modulation gradient. Maps were aligned by their gradients before averaging (green, error bars represent SEM). Absolute distance is from the edge of data after maps have been aligned. The red function represents the raw average of spectral modulation as a function of distance along the STG from the anterior temporal pole (**D**, horizontal red line; no map alignment by gradient before averaging). The inset represents data from normalized reverse correlation-based STRFs. LS, Lateral sulcus; STS, superior temporal sulcus; MTG, medial temporal gyrus; CS, central sulcus.

neighborhood similarity (Fig. 8A, group: STRF-BF, $p = 0.1$; SRF-BF, $p = 0.06$; neighborhood similarity permutation test). Despite the general lack of significant tonotopic organization within individual subjects based on neighborhood similarity, mean BF in the group data decreased as a function of distance along the STG (Fig. 8C). One aspect of spectral tuning that increases the complexity of characterizing spectral tuning organization is the widespread presence of multi-peaked SRFs in the STG. Figure 8D shows an example of a multi-peaked SRF with four peaks at the 50% maximum level (red line). We quantified the degree of multi-peaked tuning in the STG by measuring the average number of spectral tuning peaks as a function of percentage maximum level (Fig. 8E). At 80% maximum, the average number of spectral peaks was 1.4 ± 0.7 with the number of peaks increasing to 2.6 ± 1.1 at 50% maximum. We also quantified the distribution of SRF types (classified by peak number) for each level of percentage maximum. Figure 8F shows the distribution of SRF types at each percentage maximum level. At 90% maximum, 72% of the SRFs were single peaked, 25% were double peaked, and 3% had three peaks. By 70% maximum, there are more multi-peaked SRFs than single-peaked SRFs (52% multi-peaked versus 48% single peaked). Last, we examined the overall distribution of BFs and found the majority of BFs are concentrated below 1000 Hz (Fig. 8G). This low-frequency tuning is consistent with previous work characterizing tonotopic organization across the human auditory cortex in which the STG occupies a low-frequency reversal region between putative cochleotopic maps that span the planum temporale, the STG, and the superior temporal sulcus (Striem-Amit et al., 2011; Moerel et al., 2012). Collectively, these data show that the STG has a substantial proportion of multi-peaked spectral tuning, with low-to-moderate tonotopic organization within the STG, and a high concentration of low BF frequencies consistent with its placement as a low-frequency region within larger-scale cochleotopic maps (Striem-Amit et al., 2011; Moerel et al., 2012).

Discussion

In this study we used natural speech stimuli in conjunction with MID analysis and ECoG from human participants to investigate the organization of spectrotemporal processing of speech in the human STG. Based on receptive field maps, we found tuning for temporally fast sound with relatively constant energy across the frequency axis (low spectral modulation) in the posterior STG and a transition to tuning for temporally slower sound with higher variation in spectral energy across the frequency axis (high spectral modulation) in the anterior STG. Additionally, we found that the STG shows BF tuning < 1000 Hz, which is consistent with its placement as a low-frequency reversal region in larger-scale cochleotopic maps (Striem-Amit et al., 2011; Moerel et al., 2012). These data expand our view of how the spectrotemporal processing of speech in the STG is organized and demonstrate organized tuning for different acoustic aspects of speech sounds along the human STG.

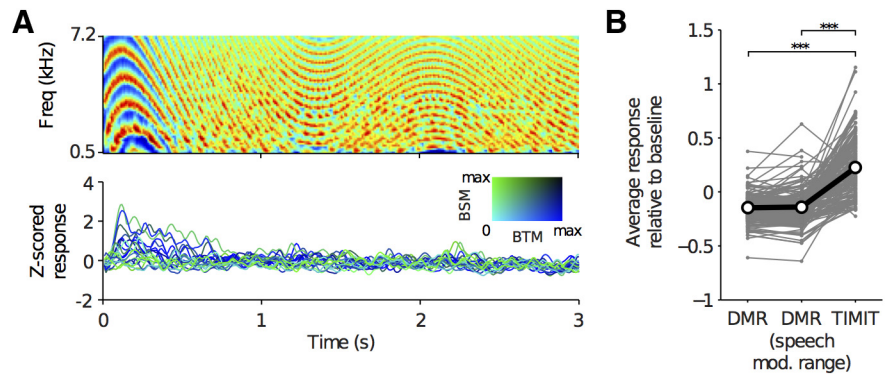


Figure 7. DMR stimuli do not activate STG as robustly as speech stimuli from TIMIT. **A**, Spectrogram of a 3 s segment of the 5 min DMR stimulus (top) and examples of corresponding Z-scored high-gamma responses from STG electrodes in one subject (GP31). Z-score was calculated using a silent baseline period. Electrode responses are colored according to their best spectral modulation (BSM) and best temporal modulation (BTM) as derived from TIMIT stimuli. As demonstrated in the bottom panel, STG electrodes showed responses to the onset of DMR stimuli, but did not elicit strong responses during the rest of the 5 min stimulus. **B**, Comparison between the average response to the DMR stimulus and the average response to TIMIT sentences for STG electrodes ($N = 4$ subjects: EC63, GP30, GP31, GP33). The response to speech was significantly higher than the response to DMR stimuli (speech vs DMR, $p = 1 \times 10^{-26}$; speech vs DMR (speech modulations), $p = 1.6 \times 10^{-25}$, Wilcoxon signed rank test, 198 STG electrodes). The average response was calculated across the entire DMR stimulus (labeled DMR) and across time points during which the DMR included modulations within the tuning range for STG (labeled speech modulations: spectral modulations, ≤ 1 cycle/octave; the absolute value of the temporal modulations, ≤ 3 Hz). Gray lines connect the mean response for the same electrode across stimuli; black line indicates the average. For both stimuli, Z-scored responses were recalculated using a silent baseline to allow for comparisons across stimuli.

Our data reveal topographic organization of spectrotemporal modulation tuning in the STG derived from speech. A previous study examined modulation tuning in the human auditory cortex using fMRI and spectrotemporal ripple stimuli (Schönwiesner and Zatorre, 2009). They observed cortical areas responsive to ripple stimuli in primary and secondary auditory cortex within superior temporal plane, but did not find evidence of organized modulation tuning. Consistent with our study, individual spectrotemporal ripple stimuli did not drive activity in the STG. Taken alone, this may indicate organized spectrotemporal modulation processing emerges in the STG. However, recent work has shown topographic organization of temporal modulation tuning alone in the human primary auditory cortex using fMRI and amplitude modulated white noise stimuli (Herdener et al., 2013). Consistent with this, Schönwiesner et al. (2009) endorse a tendency for temporal modulation tuning to increase from the medial Heschl's gyrus (HG) to the lateral HG in a manner similar to that supported by data from Herdener et al. (2013). Additionally, organized temporal modulation tuning has also been found in subcortical areas, such as the cat and the primate midbrain (Schreiner and Langner, 1988; Baumann et al., 2011). Organized joint spectrotemporal modulation tuning has been characterized in the cat inferior colliculus (Rodríguez et al., 2010) and the cat auditory cortex (Atencio and Schreiner, 2012). Collectively, these studies provide evidence for organized spectrotemporal modulation tuning in various subcortical and cortical auditory areas and indicate that such processing may be a prevalent form of functional organization within the auditory system.

In contrast to robust activation of STG by speech stimuli, our data with individual ripples (DMR stimulus) and TORCs indicate that these stimuli do not robustly activate STG, which is consistent with previous work (Schönwiesner and Zatorre, 2009). The two primary differences of speech from ripple stimuli are acoustic complexity and behavioral relevance. Although spectrotemporal modulations represent fundamental elements of complex sounds, such that a linear combination of

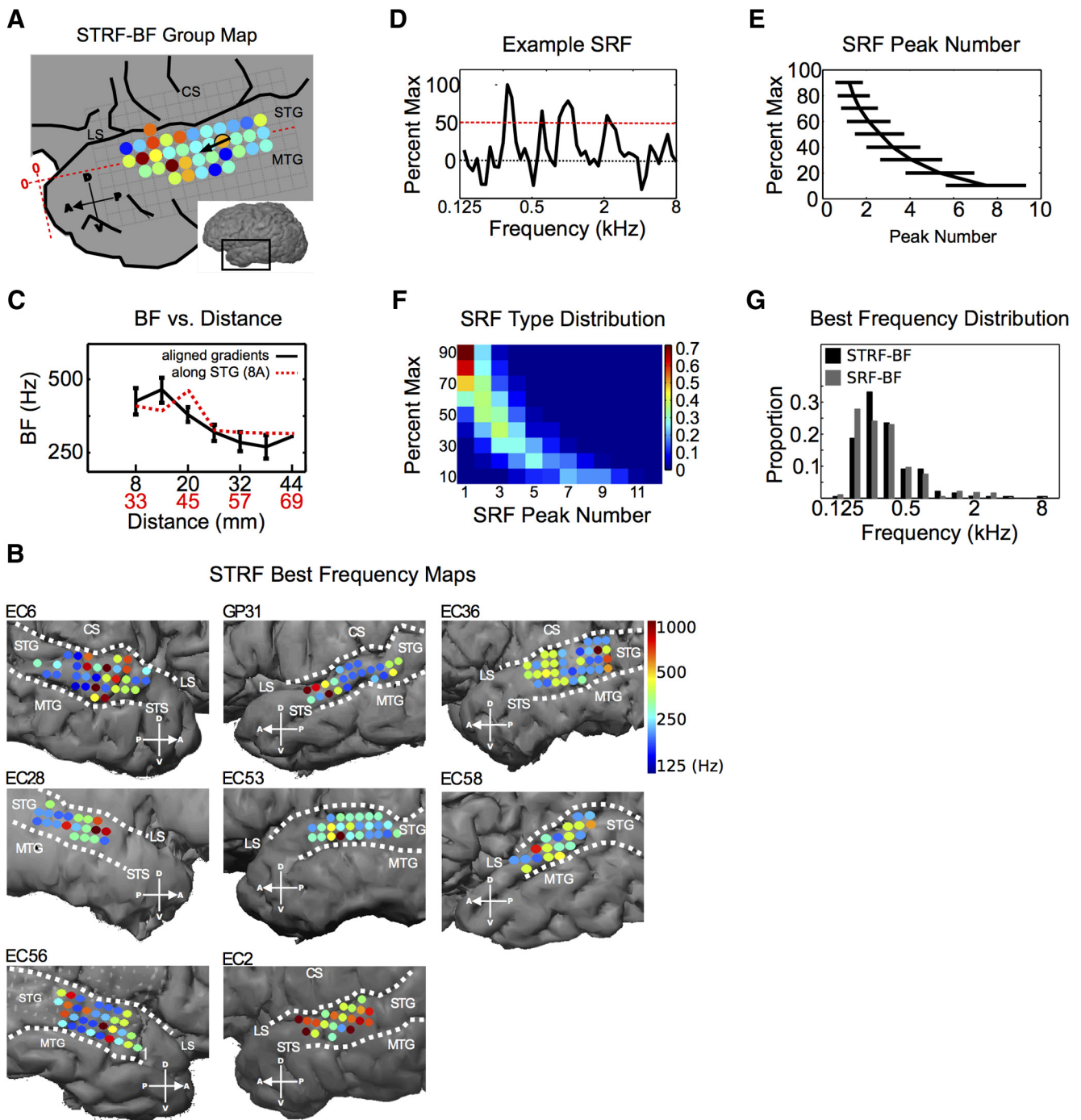


Figure 8. Topography of spectral tuning. **A**, STRF-BF group map (color scale in **B**; only sites with data from ≥ 2 subjects are shown). The BF gradient runs in the anteroventral direction at $+193^\circ$ counterclockwise from the 3 o'clock position. Neither group map shows significant local organization. (STRF-BF, $p = 0.1$; SRF-BF, $p = 0.06$; neighborhood similarity permutation test). **B**, Individual participant STRF-BF maps. Two of eight maps show significant local organization for both metrics (STRF-BF/SRF-BF: EC6, $p = 0.12/0.23$; GP31, $p = 0.004/0.007$; EC36, $p = 0.10/0.04$; EC28, $p = 0.02/0.75$; EC53, $p = 0.29/0.13$; EC58, $p = 0.27/0.07$; EC56, $p = 0.39/0.20$; EC2, $p = 0.11/0.25$; neighborhood similarity permutation test). **C**, STRF-BF as a function of distance. Absolute distance is from the edge of data after maps have been aligned by their dominant gradient (black). The red function represents the raw average of BFs as a function of distance along the STG from the anterior temporal pole (A, horizontal red line; no map alignment by gradient before averaging). **D**, Example SRF with the 50% maximum line (red). At 50% maximum, this SRF has four peaks. **E**, Average peak number as a function of percentage maximum. **F**, Distribution of neuron types, in terms of peak number, as a function of percentage maximum level. Color scale represents proportion of neurons. At 90% maximum, 72% of the SRFs are single peaked, 25% are double peaked, and 3% have three peaks. **G**, BF distribution. The concentration of BFs in the STG is < 1000 Hz, which is consistent with STG's placement as a low-frequency region in larger-scale cochleotopic maps. LS, Lateral sulcus; STS, superior temporal sulcus; MTG, medial temporal gyrus; CS, central sulcus.

these features can generate any segment of speech, speech is more acoustically complex due to such features as increased temporal envelope variability from frequent onsets in words or syllables, harmonic and formant structure, and formant transitions. These additional acoustic characteristics of speech

may be necessary to drive speech activation and may underlie selectivity seen in areas of the human auditory cortex (Overath et al., 2015). Future work, akin to the analysis-by-synthesis approach by McDermott and Simoncelli (2011), may be able to determine which acoustic properties of sound are most

important for activation of the STG. Our work shows that in the context of speech processing, the STG shows organized tuning for temporal and spectral modulation properties of speech.

Acoustic properties may not be the only etiology of speech selectivity in the STG. Along the auditory pathway, acoustic information is eventually transformed to semantic information, which confers behavioral relevance to the stimulus. Synthetic modulation stimuli lack semantic information and thus lack an important component of speech stimuli that may be responsible for activation of the STG. It is difficult to quantify the contribution of acoustic versus semantic information for driving activity in the STG in our dataset. The robust ability of STRFs to predict responses to novel speech stimuli relies on the acoustic content of sound alone. This suggests a significant component of the encoded information represents acoustic rather than semantic information and is consistent with recent work in the superior temporal sulcus showing significant selectivity for speech based on acoustic properties alone (Overath et al., 2015).

The range of temporal tuning within the human auditory cortex gives insight into the timescale of analysis being performed. Fast temporal tuning reflects processing on the phonemic timescale while slower temporal tuning reflects processing on syllabic or prosodic timescales. Currently there is considerable variation in reported temporal tuning and shape of the temporal MTF for both HG and the STG (HG, range of average best temporal modulation: 1.5–10 Hz; STG, range of average best temporal modulation: 2–8 Hz; Binder et al., 1994; Tanaka et al., 2000; Harms and Melcher, 2002; Langers et al., 2003; Liégeois-Chauvel et al., 2004; Rinne et al., 2005; Schönwiesner and Zatorre, 2009; Overath et al., 2012; Pasley et al., 2012; Wang et al., 2012; Gross et al., 2013). There are a number of possible contributors to this variability. Within these studies, a wide range of stimuli have been used to characterize temporal tuning, including sinusoidally amplitude-modulated (SAM) tones, SAM noise, moving spectrotemporal ripples, tone pips, harmonic complexes, noise bursts, consonant–vowel pairs, and continuous speech. This variation is relevant because temporal tuning is not invariant to stimulus type (Eggermont, 2002; Malone et al., 2007, 2013; Zheng and Escabi, 2008) and differences in temporal tuning in the various studies may be partially attributable to the vast array of stimulus types used to characterize the system. In this study, we use speech stimuli to determine temporal tuning in the STG and thus characterize temporal processing relevant for speech processing in the STG.

Variability in temporal tuning within the human auditory cortex may also be attributable to differences in the range and spacing of temporal modulation content used to characterize the system (Edwards and Chang, 2013). The ideal stimulus set would extend from 0 Hz to temporal modulations well beyond the limits of the system and have resolution fine enough to resolve peaks and troughs within the temporal MTF. However, variable resolution and differences in the upper and lower limits of temporal modulations used to test temporal tuning have led to variable interpretations of the shape of the temporal MTF (low pass vs bandpass) and average temporal tuning for HG and the STG (Edwards and Chang, 2013). In this study, we used speech stimuli to define the temporal tuning relevant for speech processing. Speech contains a continuous distribution of temporal modulation content rather than discretely spaced temporally modulated stimuli (ripples, amplitude-modulated white noise, click trains) that may miss peaks and

troughs in temporal tuning functions. Speech also spans the range of temporal modulations of interest for speech processing. Thus, an analysis using speech will not miss upper or lower ranges of tuning that may be missed using a predefined and narrower range of temporal modulations.

Finally, differences in metrics used to characterize temporal tuning have led to variable characterizations of temporal tuning within the STG and other areas. Nonsynchronized metrics of temporal tuning (mean activity as a function of temporal modulation) and synchronized metrics of temporal tuning (vector strength and STRF-based MTFs) are in prevalent use within the literature, but can show different tuning properties (Eggermont, 2002; Zheng and Escabi, 2008). Consistent with this, recent work in the STG has shown that a nonsynchronized metric and a synchronized metric show low-pass and bandpass tuning, respectively (Pasley et al., 2012). In this study we characterized the synchronized component of temporal tuning, the dominant form of temporal tuning for the range of temporal modulations most important for speech perception (Elliott and Theunissen, 2009; Pasley et al., 2012), and find a low-pass distribution of temporal tuning in the STG most consistent with processing on prosodic and syllabic timescales.

References

- Atencio CA, Schreiner CE (2012) Spectrotemporal processing in spectral tuning modules of cat primary auditory cortex. *PLoS One* 7:e31537. [CrossRef Medline](#)
- Atencio CA, Sharpee TO, Schreiner CE (2008) Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58:956–966. [CrossRef Medline](#)
- Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A, Rees A (2011) Orthogonal representation of sound dimensions in the primate midbrain. *Nat Neurosci* 14:423–425. [CrossRef Medline](#)
- Baumann S, Joly O, Rees A, Petkov CI, Sun L, Thiele A, Griffiths TD (2015) The topography of frequency and time representation in primate auditory cortices. *Elife* 4. [CrossRef Medline](#)
- Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Hyde JS (1994) Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Res Cogn Brain Res* 2:31–38. [CrossRef Medline](#)
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and non-speech sounds. *Cereb Cortex* 10:512–528. [CrossRef Medline](#)
- Boatman D (2004) Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92:47–65. [CrossRef Medline](#)
- Boatman D, Hall C, Goldstein MH, Lesser R, Gordon B (1997) Neuroperceptual differences in consonant and vowel discrimination: as revealed by direct cortical electrical interference. *Cortex* 33:83–98. [CrossRef Medline](#)
- Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. *Neuron* 26:695–702. [CrossRef Medline](#)
- Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM (2011) A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS One* 6:e16104. [CrossRef Medline](#)
- Carlson NL, Ming VL, Deweese MR (2012) Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. *PLoS Comput Biol* 8:e1002594. [CrossRef Medline](#)
- Chang EF, Edwards E, Nagarajan SS, Fogelson N, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2011) Cortical spatio-temporal dynamics underlying phonological target detection in humans. *J Cogn Neurosci* 23:1437–1446. [CrossRef Medline](#)
- Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. *Brazier Award-winning article*, 2001. *Clin Neurophys* 112:565–582. [CrossRef Medline](#)
- Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J Neurosci* 16:3351–3362. [Medline](#)
- David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the

- receptive field structure of v1 neurons. *J Neurosci* 24:6991–7006. [CrossRef Medline](#)
- David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* 18:191–212. [CrossRef Medline](#)
- Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85:1220–1234. [Medline](#)
- Dong D, Atick J (1995) Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Netw Comput Neural Syst* 6:159–178. [CrossRef](#)
- Edwards E, Chang EF (2013) Syllabic (~2–5 Hz) and fluctuation (~1–10 Hz) ranges in speech and auditory processing. *Hear Res* 305:113–134. [CrossRef Medline](#)
- Eggermont JJ (2002) Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. *J Neurophysiol* 87:305–321. [Medline](#)
- Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. *PLoS Comput Biol* 5:e1000302. [CrossRef Medline](#)
- Escabi MA, Schreiner CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci* 22:4114–4131. [Medline](#)
- Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* 412:787–792. [CrossRef Medline](#)
- Felsen G, Touryan J, Han F, Dan Y (2005) Cortical sensitivity to visual features in natural scenes. *PLoS Biol* 3:e342. [CrossRef Medline](#)
- Garofolo J, Lamel LF, Fisher WM, Fiscus J, Pallett D, Dahlgren N, Zue V (1993) TIMIT acoustic-phonetic continuous speech corpus. Philadelphia: Linguistic Data Consortium.
- Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol* 11:e1001752. [CrossRef Medline](#)
- Harms MP, Melcher JR (2002) Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol* 88:1433–1450. [Medline](#)
- Herdener M, Esposito F, Scheffler K, Schneider P, Logothetis NK, Uludag K, Kayser C (2013) Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex* 49:2822–2833. [CrossRef Medline](#)
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402. [CrossRef Medline](#)
- Hsu A, Woolley SM, Fremouw TE, Theunissen FE (2004) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *J Neurosci* 24:9201–9211. [CrossRef Medline](#)
- Kim G, Doupe A (2011) Organized representation of spectrotemporal features in songbird auditory forebrain. *J Neurosci* 31:16977–16990. [CrossRef Medline](#)
- Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9:85–111. [CrossRef Medline](#)
- Klein DJ, Simon JZ, Depireux DA, Shamma SA (2006) Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J Comput Neurosci* 20:111–136. [CrossRef Medline](#)
- Langers DR, Backes WH, van Dijk P (2003) Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. *Neuroimage* 20:265–275. [CrossRef Medline](#)
- Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14:731–740. [CrossRef Medline](#)
- Malone BJ, Scott BH, Semple MN (2007) Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol* 98:1451–1474. [CrossRef Medline](#)
- Malone BJ, Beitel RE, Vollmer M, Heiser MA, Schreiner CE (2013) Spectral context affects temporal processing in awake auditory cortex. *J Neurosci* 33:9431–9450. [CrossRef Medline](#)
- McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71:926–940. [CrossRef Medline](#)
- Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87:516–527. [Medline](#)
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32:14205–14216. [CrossRef Medline](#)
- Nagel KI, Doupe AJ (2008) Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron* 58:938–955. [CrossRef Medline](#)
- Nourski KV, Steinschneider M, Oya H, Kawasaki H, Jones RD, Howard MA (2014) Spectral organization of the human lateral superior temporal gyrus revealed by intracranial recordings. *Cereb Cortex* 24:340–352. [CrossRef Medline](#)
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609. [CrossRef Medline](#)
- Overath T, Zhang Y, Sanes DH, Poeppel D (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. *J Neurophysiol* 107:2042–2056. [CrossRef Medline](#)
- Overath T, McDermott JH, Zarate JM, Poeppel D (2015) The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat Neurosci* 18:903–911. [CrossRef Medline](#)
- Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15:243–262. [CrossRef Medline](#)
- Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing speech from human auditory cortex. *PLoS Biol* 10:e1001251. [CrossRef Medline](#)
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724. [CrossRef Medline](#)
- Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9:e1000610. [CrossRef Medline](#)
- Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc Biol Sci* 262:259–265. [CrossRef Medline](#)
- Rinne T, Pekkola J, Degerman A, Autti T, Jääskeläinen IP, Sams M, Alho K (2005) Modulation of auditory cortex activation by sound presentation rate and attention. *Hum Brain Mapp* 26:94–99. [CrossRef Medline](#)
- Rodríguez FA, Read HL, Escabi MA (2010) Spectral and temporal modulation tradeoff in the inferior colliculus. *J Neurophysiol* 103:887–903. [CrossRef Medline](#)
- Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, and Formisano E (2014) Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comp Biol* 10:e1003412. [CrossRef Medline](#)
- Schönwiesner M, Zatorre RJ (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. [CrossRef Medline](#)
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60:1823–1840. [Medline](#)
- Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16:223–250. [CrossRef Medline](#)
- Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD (2006) Adaptive filtering enhances information transmission in visual cortex. *Nature* 439:936–942. [CrossRef Medline](#)
- Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386:69–73. [CrossRef Medline](#)
- Steinschneider M, Fishman YI, Arezzo JC (2008) Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb Cortex* 18:610–625. [CrossRef Medline](#)
- Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One* 6:e17832. [CrossRef Medline](#)
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91:1282–1296. [Medline](#)
- Talebi V, Baker CL Jr (2012) Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. *J Neurosci* 32:1560–1576. [CrossRef Medline](#)

- Tanaka H, Fujita N, Watanabe Y, Hirabuki N, Takanashi M, Oshiro Y, Nakamura H (2000) Effects of stimulus rate on the auditory cortex using fMRI with “sparse” temporal sampling. *Neuroreport* 11:2045–2049. [CrossRef Medline](#)
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20:2315–2331. [Medline](#)
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12:289–316. [CrossRef Medline](#)
- Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. *J Neurophysiol* 107:2033–2041. [CrossRef Medline](#)
- Wilkinson F, James TW, Wilson HR, Gati JS, Menon RS, Goodale MA (2000) An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Curr Biol* 10:1455–1458. [CrossRef Medline](#)
- Yang X, Wang K, Shamma SA (1992) Auditory representations of acoustic signals. *IEEE Trans Information Theory* 38:824–839. [CrossRef](#)
- Zheng Y, Escabí MA (2008) Distinct roles for onset and sustained activity in the neuronal code for temporal periodicity and acoustic envelope shape. *J Neurosci* 28:14230–14244. [CrossRef Medline](#)