Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries

Adeen Flinker[1,4], **Werner K. Doyle**[2], **Ashesh D. Mehta**[3], **Orrin Devinsky**[4], **David Poeppel**[1,5]

[1]Department of Psychology, New York University, New York, NY, USA

[2]Department of Neurosurgery, New York University School of Medicine, New York, NY, USA

[3]Department of Neurosurgery, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Manhasset, NY.

[4]Department of Neurology, New York University School of Medicine, New York, NY, USA

[5]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

Abstract

The principles underlying functional asymmetries in cortex remain debated. For example, it is accepted that speech is processed bilaterally in auditory cortex, but a left hemisphere dominance emerges when the input is interpreted linguistically. The mechanisms, however, are contested: what sound features or processing principles underlie laterality? Recent findings across species (humans, canines, bats) provide converging evidence that spectrotemporal sound features drive asymmetrical responses. Typically, accounts invoke models wherein the hemispheres differ in time-frequency resolution or integration window size. We develop a framework that builds on and unifies prevailing models, using spectrotemporal modulation space. Using signal processing techniques motivated by neural responses, we test this approach employing behavioral and neurophysiological measures. We show how psychophysical judgments align with spectrotemporal modulations and then characterize the neural sensitivities to temporal and spectral modulations. We demonstrate differential contributions from both hemispheres, with a left lateralization for temporal modulations and a weaker right lateralization for spectral modulations. We argue that representations in the modulation domain provide a more mechanistic basis to account for lateralization in auditory cortex.

Introduction

A hallmark of speech perception and language comprehension is that these perceptual and cognitive processes are subserved by an asymmetric distribution of cortical circuitry. The original observations of Broca [1] and Wernicke [2] provided striking evidence that damage to cortical regions in the left (dominant) but not the right hemisphere caused impairments in comprehension and production. A great deal of research has focused on elucidating the functional neuroanatomy of these (and other, subsequently identified) regions as well as their underlying computational principles [3–7]. In contrast to the historically established one-size-fits-all view on lateralization of speech and language - a perspective that remains the prevailing one in the clinical literature [8,9] - there is now emerging consensus that both left and right temporal cortices are heavily involved in speech perception proper (as well as some aspects of linguistic processing [10]), i.e. the mapping from acoustic input to the internal representations (informally speaking, words) that form the basis for language processing (for recent fMRI data supporting this hypothesis, see [11]). However, the functional and computational differences between the two hemispheres with respect to auditory processing remain incompletely understood and vigorously contested [12,13].

A series of influential behavioral studies provided suggestive evidence that the left hemisphere is sensitive to rapidly changing auditory cues [14–16]. This sensitivity manifested when temporal intervals between stimuli were reduced below 100–150 ms in aphasic patients [temporal order judgment] [14], children with developmental delays [discrimination] [16], as well as a sensitivity to the length of formant transition in healthy listeners [dichotic listening] [15]. In parallel, clinical reports argued that focal damage to the right hemisphere produces impairments in processing slower changing prosodic cues in speech [17,18] as well the ability to discriminate spectral information [19,20].

Based on these findings as well as on foundational neuroimaging studies [21,22], a framework emerged whereby left hemisphere structures - specifically language related regions, principally the temporal lobe but also inferior frontal gyrus - are more sensitive to temporal cues whereas right hemisphere structures are more sensitive to spectral cues [23]. A related, complementary framework explained the differences in auditory sensitivity as a function of temporal integration windows of neural ensembles in auditory cortex, proposing that the left auditory cortex integrates incoming auditory information over shorter timescales (~20–80ms) and right regions over longer timescales (~150–300 ms), asymmetric sampling in time [24]. A third view associates lateralization with function: if representations are linguistic, they are processed by the dominant, typically left hemisphere [13,25]. Notwithstanding the existing neural and behavioral evidence, the mechanisms underlying functional lateralization in hearing remain underspecified and are the source of significant disagreements. Focusing here solely on competing auditory theories, we advance these theories building upon their approaches and providing a unifying framework, motivated by a different analytic view on signals that is closely aligned with neural response properties.

Recent reports from different mammalian species including humans [26–29], dogs [30], and mustached bats [31,32] provide converging evidence that spectrotemporal sound features are processed in an asymmetrical manner, presumably reflecting shared neurocomputational

principles across species. Nevertheless, conceptual gaps have held back a cross-species model and interpretation. First, the stimuli as well as the type of stimulus manipulation employed are inconsistent across studies (both within human research and across species). Second, the models used to interpret the results [23,24] differ in their implementational specificity regarding the underlying cortical tissue.

To address these gaps, we reframe and test the asymmetry hypothesis [24] in the modulation domain, an acoustic 'space' that has been developed to mirror successful analysis approaches in the visual domain [33,34]. The modulation domain reflects energy fluctuations that vary across the temporal and spectral axes of a spectrographic representation (i.e. temporal and spectral modulations of power in the time frequency representation), similarly to horizontal and vertical spatial gratings that comprise an image (Figure 1a). The modulation domain quantifies temporal and spectral acoustic features (Figure 1b) and has been used to investigate communication vocalizations (and link them to neural mechanisms) in the ferret [35], zebra finch [36], and in humans [37–40]. Recently, spectral and temporal modulations have provided novel topographic maps of the human auditory cortex [41,42], and auditory models capitalizing on this domain have proved to be more accurate at reconstructing neural activity both in electrophysiology [43] and neuroimaging [42]. We take these recent successes using the modulation domain in neural representation as providing a new opportunity to investigate auditory cortical asymmetries, building upon the prevailing auditory theories, namely the temporal versus spectral view [23], and the asymmetric sampling in time hypothesis [24,44].

We hypothesize that left and right auditory cortical fields differ in how they integrate temporal and spectral modulations across the time-frequency representation of speech. We argue that the left hemisphere integrates over a wider range of temporal modulations (slow to fast temporal modulations) but over a limited range of spectral modulations (low spectral modulations); in contrast, the right hemisphere integrates over a wide range of spectral modulations (low to high) but a limited range of temporal modulations (slow) (Figure 2. noteworthy, both hemispheres have a range of temporal and spectral modulations that overlap, namely both low temporal and low spectral modulations are processed by both hemispheres). This hypothesis provides a computationally specific stimulus space (modulation domain) that is linked to an implementational computation in cortex (i.e. integration of neuronal inputs from subpopulations). In order to test this hypothesis, we provide a filtering technique in the modulation domain that is applied to speech stimuli, and we test this hypothesis in human listeners using psychophysical measures (diotic and dichotic listening), neuronal measures (magnetoencephalography; MEG) as well as direct recordings from cortex in neurosurgical patients (electrocorticography; ECoG).

Results

We developed an analytic technique that filters stimuli in the modulation domain based on a cochlear time-frequency representation and permits resynthesis of a new waveform that corresponds to the filtered representation (Figure 3a, Supplementary Figure 1, see Methods, Modulation Domain Filtering). We closely follow the framework introduced by Elliott & Theunissen [38], but we diverge from that approach in critical ways in how we decompose and

resynthesize the signal in the frequency domain. While Elliott & Theunissen employed a linear frequency scale (short time Fourier transform), we employ a logarithmic frequency scale following the frequency distribution of the cochlea [45] using a filter bank decomposition (see Methods, Modulation Domain Filtering). We used this technique on a wide selection of English sentences and different speakers in order to control parametrically the degree of temporal modulations (Figure 3b) or spectral modulations (Figure 3c) contained by each sentence. It is noteworthy that our approach can successfully filter modulation rates in continuous speech while still complying with the envelope projection test [46]. The envelope projection test is used to verify that a filtering technique claiming to limit temporal modulations indeed produces the desired modulation spectra in the final resynthesized waveform, unlike previous approaches that inadvertently reintroduced undesired modulations in the resynthesis process (e.g. Drullman [47]). Our filtering approach produces spoken sentences that sound natural but have a controlled amount of modulation rates (in contrast to a majority of studies employing non-speech stimuli or artificial noise carriers).

Using a resolution that is rather more fine-grained than the existing literature (1 Hz temporally and 0.1867 cycles/octave spectrally), we first examined behavioral responses in English speaking participants (N=20). Listeners were presented the materials diotically and reported intelligibility and speaker voice pitch (male or female). This approach has been employed infrequently in the literature [38,48], and here we report psychometric curves on voice pitch identification as well as intelligibility in the critical modulation range for speech (i.e. 2–8 Hz). Filtering out temporal modulations completely abolished intelligibility at the lowest cutoff (2 Hz) and showed a logistic relationship between the amount of temporal modulation present in the signal and the degree of intelligibility. The most prominent boost in intelligibility was observed when stimuli contained more than 5 Hz temporal modulations and, unsurprisingly, during the second block of the task, when all stimuli were repeated (within-subject non-parametric factorial permutation test main effect of block $P_{F\text{-value}}$=88.84, p<0.001, 95% CI of null hypothesis statistic=[0.029, 2.68], Figure 4a, Supplementary Figure 2). This effect is especially striking as each sentence is only presented once within a block, so the performance boost represents information from only one additional exemplar. As proportion data as well as categorical outcomes can violate ANOVA assumptions [49] we chose to use a non-parametric approach [50] and subsequently model psychometric curves within each participant using a logistic function[51]. The mean and standard error distribution of curve fits across participants are plotted as a continuous shaded line in Figure 4. To quantify the temporal modulation ranges where the psychometric curves showed a significant difference between the first and second block a paired Wilcoxon signed-rank test was performed for each value and a continuous block of 2.8 and 7.1 Hz was found significant (non-parametric Wilcoxon signed-rank test p<0.05). Spectral modulations abolished the ability to correctly identify the speaker's voice pitch with a sharp increase in the performance when stimuli contained above 0.74 cycles/octave spectral modulations (Figure. 4a right panel). There was no effect of block (within-subject non-parametric factorial permutation test main effect of block $P_{F\text{-value}}$= 0.1195, p=0.992, 95% CI of null hypothesis statistic=[0.0795, 2.96]) nor was there a significant difference between psychometric curves (non-parametric Wilcoxon signed-rank test p>0.17 for all paired tests). These results provide a compelling link between the two axes of the modulation space and

our ability to process the content of speech and speaker identity (speaker identity was maximally different in our stimulus at 1 cyc/oct, see Supplementary Figure 3).

We next asked whether differences in cortical processing across the two hemispheres could be detected using psychophysical measures. We designed a dichotic paradigm that leverages an asymmetry in the auditory pathway when different stimuli are presented to each ear [52]. While the classic dichotic design elicits a right ear advantage when short (usually consonant-vowel) competing stimuli are presented, we presented identical sentences that varied by the amount of temporal or spectral information in each ear (e.g. 3 Hz in the one ear and 4 Hz in the other, etc.). In a separate cohort of participants (N=60, see Methods, Psychophysical Experiments), we found a significant behavioral advantage when the right ear was presented with more temporal modulations than the left (main effect of ear within-subject non-parametric factorial permutation test $P_{F\text{-value}}$=18.963744 , p<0.001, 95% CI of null hypothesis statistic=[0.05455, 2.689], Figure 4b left panel). This right ear advantage (REA) reveals a specific hemispheric preference for processing high temporal modulations and no preference for low temporal modulations and the psychometric curves showed significant differences for values between 3.6 and 6 Hz (non-parametric Wilcoxon tailed signed-rank test p<0.05 for all paired tests). An analysis of the raw data in the voice pitch identification task did not exhibit a significant LEA during the task (main effect of ear within-subject non-parametric factorial permutation test $P_{F\text{-value}}$= 2.497, p=0.079, 95% CI of null hypothesis statistic=[0.03147, 2.916], Figure. 4b right panel). But, an analysis of the fitted within-subject psychometric curves showed significant differences for values between 0.373 – 0.541 cyc/oct (non-parametric Wilcoxon tailed signed-rank test p<0.05 for all paired tests).

In order to further quantify and assess neurally the sensitivity of each hemisphere to temporal and spectral modulations, we next used the same stimuli and the diotic paradigm while MEG signals were being recorded from participants (N=19). Across recording channels, average neuronal power (quantified between 0.1–8 Hz) was highest when participants listened to sentences with the highest temporal modulation (6 Hz) or spectral modulation (0.93 cyc/octave) content (Figure 5, top panel, Supplementary Figures 4, 5). After the onset of a sentence, the average power elicited by different modulation rates diverges with a systematic order in magnitude from lower modulations (blue line) to higher modulations (red line) and converge by the time the sentence has ended (Figure 5, top panel). We quantified this effect by correlating neuronal power with the sentence modulation rate (filter cutoffs) and tested for significance using a permutation approach for each sensor. Neuronal power significantly correlated (p<0.05, permutation test) with stimulus temporal modulation cutoffs across time, showing an increase in neural power as the participants heard sentences with increased temporal modulation content (Figure 5 middle panel). These correlations were strongest (maximal value of R=0.0796 at 500 ms post stimulus onset) during the onset of the stimulus and dropped by the time of stimulus offset (R=0.0428) with significant correlations lasting up to 110ms post stimulus offset (R=0.0231) (Figure 5 middle panel, black horizontal lines denote significance). Similarly, correlations with spectral modulations were strongest (maximal value of R=0.0489 at 1060 ms post stimulus onset) during stimulus presentation and dropped after stimulus offset (R=0.0413) with significant correlations lasting up to 180 ms post stimulus offset (R=0.0238).

Next, in order to investigate how these correlations are distributed spatially, we averaged correlations across time for the first half of the sentence (Figure 5 middle panel, left topoplot) and second half of the sentence (Figure 5 middle panel, right topoplot) and statistically assessed correlations for each sensor. The distribution of significant correlations (p<0.05, permutation test with a cluster correction for multiple comparisons) showed a left-hemisphere biased topography (Figure 5 middle insets and lower panels), with a laterality index of 0.32 (−1 is maximally right hemisphere and 1 is maximally left hemisphere) for the temporal modulations. We statistically assessed laterality with a permutation test and found a significant left laterality where left sensor power was greater than right (permutation test, $P_{difference\_of\_mean\_R}$=0.01021, p=0.0020, 95% CI of null hypothesis statistic=[−0.00334 0.00309]). Analogously, neuronal power significantly correlated with spectral modulation stimulus cutoffs and showed a right-hemisphere biased topography, with a laterality index of −0.18. We statistically assessed laterality with a permutation test and found a significant right laterality where right sensor power was greater than left (permutation test $P_{differenc\_of\_mean\_R}$=−0.00498, p=0.016, 95% CI of null hypothesis statistic=[−0.00330 0.00293]). We systematically probed neural frequency power in band-limited steps in order to investigate whether these sensitivities to temporal and spectral modulations are spread across the neural frequency spectrum. We found significant correlations with temporal modulations in both low (<8 Hz) and high frequency (12–23 Hz) ranges, but spectral modulations only correlated with low frequency neural ranges (Supplementary Figure 6).

To further substantiate our analysis approach and elucidate the cortical sources underlying the effect, we projected our data on the cortical surface using minimum norm estimate (MNE) and followed the same approach we employed in sensor space, i.e. correlated neural power with stimulus modulation cutoff for each cortical source. The significant correlations (permutation test p<0.05) shown in Figure 5 (bottom) replicate the finding in sensor space, with an asymmetric leftward distribution for correlations with temporal modulations (Figure 5 bottom). Similarly to the sensor data, we statistically assessed this laterality with a permutation test comparing left and right neuronal power. Correlations with temporal modulations showed a significant left laterality where left source power was greater than right (permutation test, $P_{difference\_of\_mean\_R}$ = 0.0199, p=0.003, 95% CI of null hypothesis statistic =[−0.0059 0.0056]) but the spectral modulation asymmetry did not pass significance (permutation test, $P_{difference\_of\_mean\_R}$ = 0.0025, p=0.163, 95% CI of null hypothesis statistic =[−0.0057 0.0056]). Nevertheless, the same analysis limited to Heschl's gyrus was significant for both temporal modulations (permutation test, $P_{differenc\_of\_mean\_R}$ = 0.0395 p= 0.0180, 95% CI of null hypothesis statistic =[−0.0041 0.000]) as well as spectral modulations (permutation test, $P_{differenc\_of\_mean\_R}$ = −0.05846, p=0.0170, 95% CI of null hypothesis statistic =[−0.0039 0.0041]). These correlations were strongest at Heschl's gyrus (Temporal modulation mean correlation for left and right hemispheres = 0.1367, 0.0972 respectively; Spectral modulation mean correlation for left and right hemispheres = 0.040366, 0.0988 respectively) and decreased in higher order cortical regions (Temporal modulation correlation range across high order regions for left and right hemispheres = [0.07138 – 0.1294],[0.03734 – 0.05915] respectively; Spectral modulation correlation range across high order regions for left and right hemispheres = [−0.002693 – 0.032166], [0.027008 – 0.06067] respectively; Supplementary Figure 7).

In addition to a strong relationship between the stimulus modulation space and neural responses, we tested for significant correlations between neural responses and the participants' behavioral ratings (intelligibility on a scale from 1–4). While correlation values were lower overall, they still passed significance (p<0.05, permutation test) and exhibited a more left lateralized topography compared with modulation space correlations (Supplementary Figure 8). Correlations with the behavioral responses (male or female) in the voice pitch identification task did not pass significance (Supplementary Figure 9), providing evidence that cortical responses are sensitive to the spectral content during the task. In order to verify that the correlations in the intelligibility task were due to the temporal modulations rather than the intelligibility of speech content per se, we replicated the effect in a separate cohort of participants using *reversed speech* stimuli (unintelligible, following an identical filtering procedure in the modulation domain, see Methods, Psychophysical Experiments). The analysis revealed significant correlations with a left-hemisphere biased topography (Supplementary Figures 10, 11).

Taken together, these results demonstrate that the sources of the correlations have a specific spatial topography which evolves over time, culminating in a left-hemisphere biased distribution in the case of temporal modulations and a right-hemisphere biased distribution in the case of spectral modulations. This pattern of results is most consistent with the modulation-based asymmetry hypothesis.

Lastly, we sought to verify the asymmetry of neural sources with a technique that is unbiased by the computational assumptions of source localization. We recorded intracranial neural signals in a cohort of neurosurgical patients undergoing treatment for refractory epilepsy who were implanted with stereotactic depth and surface electrodes for clinical monitoring. In a patient with rare bilateral depth coverage of superior temporal cortices, we found a significant correlation with temporal modulations in a left STG electrode (permutation test, $P_{correlation}$=0.413, p<0.001, 95% CI of null hypothesis statistic =[−0.173279 0.196382], R^2=0.1708) and in a right STG electrode a significant correlation with spectral modulations (permutation test, $P_{correlation}$=0.241, p=0.015, 95% CI of null hypothesis statistic =[−0.171243 0.193584], R^2= 0.05831). These results replicate our non-invasive low frequency power correlations within participant and within electrode (Figure 6, Supplementary Figures 12-14); but these patterns could also be observed when examining correlations with neural high gamma activity (Supplementary Figure 15). An analysis of the correlations across all electrodes and participants showed that temporal modulation correlations were larger in the left hemisphere (MN=0.1350, SEM=0.0070) than the right hemisphere (MN=0.1011, SEM=0.0084) and the difference between left and right hemispheres was statistically significant (Wilcoxon rank sum test, Z=2.969, p=0.0030). Correlations with spectral modulations did not show a significant difference between hemispheres (Wilcoxon rank sum test, Z=−0.640, p=0.5223) with left hemisphere values (MN=0.1163, SEM=0.0063) similar to those in the right hemisphere (MN=0.1199, SEM=0.0076). Across patients, significant correlations with temporal modulations were more prominently visible in the left hemisphere with 8.042% of total left electrodes and 3.218% of total right electrodes showing significant correlations (permutation test, p<0.05; Supplementary Figure 16, 17). Correlations with spectral modulations showed a more

bilateral distribution with 6.818% of total left electrodes and 6.188% of total right electrodes showing significant correlations (permutation test, p<0.05; Supplementary Figures 16, 17).

Discussion

We provide a theoretical and computationally driven account of hemispheric asymmetries for processing speech and other continuous acoustic signals. We introduce a well-defined computational space to investigate hemispheric asymmetries (Figure 2), signal processing techniques to manipulate natural stimuli in this space (Figure 3), a psychophysical mapping of this space, as well as behavioral and multimodal (MEG, ECoG) neural evidence supporting a cortical asymmetry in processing temporal-spectral modulations in speech.

The modulation domain is arguably an ideal auditory stimulus space for representing speech intelligibility [38] as well as permitting unified computations across different cortical modalities [53]. Due to the complexity of manipulating and reconstructing acoustic signals in this space, only few studies have employed this approach to investigate speech intelligibility [37,38,54]. Our approach builds on the pipeline proposed by Elliott et al.[38] (i.e. linear time-frequency decomposition, modulation filtering, iterative convex projection) but diverges by employing a cochlear filter-bank which represents frequency in a logarithmic scale (similarly to Chi et al.[48]) which more closely reflects the biological representation of frequency in the cochlea. Unlike previous reports that mapped a wide range of the modulation space, we focused on a fine-grained mapping of the modulation space that is most relevant to speech intelligibility (below 10 Hz and below 1.5 cycles/octave). The more fine-grained analysis provided a 2–3 fold increase in resolution, revealing a sharp inflection point at the transition between 4 and 5 Hz, which is most likely due to the inclusion of critical temporal cues that dominate the temporal modulation spectra of speech [47,55,56]. The boost in intelligibility is also evident, showing a significant increase in percent words identified above 4 Hz when the stimulus is presented for the second time in a new block (the same sentence is never played with a higher modulation cutoff during the first block). We did not find this priming effect in the voice pitch identification task, but our fine-grained resolution does provide a psychometric curve that has been only partially reported in the past (see data in Elliott et al.[38]) and reflects the degradation of speaker identity (male or female) as spectral modulation content is removed. The critical (and statistically significant) differences between male and female voices in our materials are well represented in the spectral modulation domain in the range 0.8–1.3 cyc/oct (Supplementary Figure 3) which fits well with our behavioral results showing a clear increase in performance at 0.75 cyc/oct which approaches ceiling above 1 cyc/oct (Figure 4, Supplemental Figure 2). These ranges most likely represent modulation power due to pitch differences in the male and female speakers and are driven by the fundamental frequency (which in our case is smoothed by low spectral cutoffs and masks the voice pitch of the speaker). Our manipulation of spectral cues also provides evidence for an effect of spectral modulation on intelligibility as previously reported by Shannon et al. [57] and Chi et al.[48]. This effect is much smaller in comparison to degrading temporal modulations and is limited to low spectral modulation values which could be processed in both hemispheres according to our model.

The key conceptual advance of these studies is a link between the modulation domain as a representational space and cortical hemispheric asymmetries. The most prevalent models for hemispheric asymmetries in speech formulate computational differences between right and left (auditory) cortices as a difference in time-frequency resolution [23] or as a difference in temporal window sizes over which information is integrated [24]. The time-frequency resolution argument draws a parallel between cortical computation and the uncertainty principle in decomposing an auditory signal. The acoustic (i.e. Heisenberg-Gabor) uncertainty principle provides a theoretical limit on the degree a signal can be resolved in time and frequency simultaneously: a highly resolved signal in the time domain inherently limits the attainable frequency resolution and vice versa [58]. Proposals in both humans [23], as well as other model organisms [31], claim that hemispheric asymmetries address this uncertainty by optimizing acoustic processing with a higher temporal resolution in the left hemisphere and a higher frequency resolution in the right hemisphere.

Our approach closely follows this time-frequency dichotomy but is motivated by building closely on a biologically plausible pathway: we first represent an auditory signal using a filter bank mimicking the log-frequency spacing of the cochlea [45] and then move to the modulation space, which provides an explicit definition of temporal and spectral resolution as well as a cortically plausible operation of integration (i.e. integrating different cortical populations which are organized tonotopically). The modulation space in our model provides a plausible candidate for cortical computation in the realm of speech processing. First, the auditory pathway exhibits a low-pass characteristic, whereby ascending from subcortical to cortical structures elicits lower temporal modulation responses [59,60]. Second, this low-pass neural characteristic is also evident in the temporal modulation spectra of speech signals [55,56] and speech intelligibility [47].

Our finding of asymmetric power correlations fits well with previous time-frequency tradeoff accounts that have been based on changes in temporal and spectral cues in non-speech stimuli [12,21,23,61], but it also provides neural evidence for specific acoustic cues in more natural speech stimuli (i.e. sentences). Previous reports have provided evidence for hemispheric asymmetries to specific acoustic cues in non-speech stimuli [27,28,61–69], phonemic segments [21,22,70–72], as well as single words [73,74] but reports using sentence stimuli have been mostly linked to neural signatures rather than specific cues in the sentence stimuli [75–77]. This is partly due to the complexity of manipulating relevant acoustic cues in sentential stimuli, such as the temporal and spectral modulations, which we directly address with our technique.

The Asymmetric Sampling in Time (AST) model proposes that the initial neural representation of speech is bilaterally symmetric, say at the levels of inferior colliculus to primary/core auditory cortex; however the two hemispheres 'resample' information asymmetrically [24]. The right hemisphere predominantly extracts information from longer temporal windows while the left hemisphere extracts information over shorter temporal windows. This approach focuses on temporal sampling rather than time-frequency resolution per se (albeit frequency resolution is implicitly greater in the case of larger temporal window), but it is broadly consistent with our modulation domain model. In both models the speech signal is first represented bilaterally in a symmetric fashion (i.e. a high-resolution

time-frequency representation) and then sampled asymmetrically. In the modulation model, information is extracted by integrating ranges of modulation rates, while in the AST model information is extracted by sampling over windows. The modulation model can be viewed as a significant extension and modification of the AST hypothesis by extracting information over *many* windows both temporally and spectrally. In this view, the right hemisphere predominantly extracts information from long temporal windows due to integration over a limited range of temporal modulations (reflecting longer temporal cycles) while the left integrates a wide range of temporal modulations (reflecting both short and long temporal cycles). Accepting this new hypothesis demands a reframing of the AST model wherein the left hemisphere extracts both short and long temporal windows (but the right is still preferentially integrating long windows), which is consistent with reports of informative theta activity during speech perception [76,77].

One important aspect of the AST model is a specific oscillatory mechanism that performs temporal sampling and integration. There is evidence supporting left lateralized high frequency oscillations and right lateralized oscillations in rest as well as during speech processing [29,75,78–81]. We investigated a range of power correlations across the neural spectrum and found that sensitivity to temporal modulations existed in both low and high frequency ranges and were left lateralized (more so in the case of high frequencies, see Supplementary Figure 6). We interpret our current findings as a superposition of two signals or event types. The first is generated by impulse responses of neural populations to the incoming speech signal, it is indexed by low frequency power as well as broadband activity in the ECoG recordings, and it reflects integration of neural activity (and populations sensitive to specific temporal modulation ranges). The second is generated by ongoing oscillations in the theta and low gamma ranges that are indexed by corresponding frequency power and by our hypothesis reflect temporal sampling or 'chunking' of the auditory signal. It would make sense for such a sampling mechanism to be mostly associated with temporal cues, such as the temporal modulations of the signal. This view is consistent with our findings of low and high frequency power correlations with temporal modulations (Supplementary Figure 6). In contrast, the spectral correlations that are restricted to lower frequencies most likely reflect the first mechanism, that is neural populations that integrate ranges of spectral modulations produce impulse responses that are mostly reflected in low frequency activity.

While we provide strong evidence for left hemisphere asymmetry during processing of temporal modulations in intelligible speech in our dichotic behavioral task (Figure 4), neurophysiology results (Figures 5–6) as well as processing unintelligible reversed speech (Supplementary Figure 10), the data supporting asymmetrical processing of spectral modulations is more complex. In the behavioral dichotic task we only found a significant difference for the within-subject fitted curves and weaker correlations with spectral modulations over all experiments compared to the temporal modulations. While we found significant right hemisphere laterality in the sensor level MEG analysis, we did not find a significant asymmetry in the ECoG data. One possibility is that the choice of task (Voice Pitch Identification) is not ideal for investigating spectral modulations within the context of speech.

Another viewpoint could be raised under the lens of lateralization driven by domain-specific speech mechanisms [13,25]. Under this view, the right hemisphere is strongly selective for certain acoustic cues and the left hemisphere specializes in domain-specific (speech) mechanisms and is not selective to an acoustic regime [13]. The strong asymmetry of temporal modulation correlations to the left hemisphere partly challenges this view by providing an acoustic regime (temporal modulations) that well describes and represents speech both for intelligible and non-intelligible (reversed) speech. While temporal modulations could be a good characteristic of speech-like acoustics, thereby activating left hemisphere domain specific circuits, our data does not support a purely bottom up feed-forward acoustic effect. The degree of frontal-temporal recruitment both in our MEG and ECoG results suggest, unsurprisingly, that there is a larger network involved that may be sensitive both to the temporal modulations as well as the task and attentional demands. Indeed, the context and expectations during our tasks can influence perception of the modulations and the speech content.

Our findings and approach complement a growing number of studies in human neuroimaging reporting a spatial topography of auditory cortex sensitivity to temporal and spectral modulations [41,42,82–84]. Interestingly, these studies have reported bilateral activation patterns and modest asymmetries at best, and furthermore the modulation topography profiles differ across studies [41,42,82]. The limited asymmetry and discrepancy across results are most likely due to the varying stimuli used to probe cortical responses (modulated noise, ripples, environmental sounds, and speech). In the case of speech, both electrophysiology [43] and neuroimaging [42] have shown that modulation domain cues are critical for more accurate modeling of auditory responses. Our stimuli constrain the amount of modulation information contained in the speech signal that drives asymmetric responses not readily seen in previous reports employing the full modulation domain.

In summary, our approach offers a unifying framework to standardize stimulus manipulation across the unit of interest (non-speech, phonemic segment, word, etc.) as well as across species of interest. We provide behavioral and neurophysiological evidence that the two hemispheres are differentially sensitive to ranges of temporal and spectral modulations. We view these sensitivities as an asymmetry in cortical architecture reflecting a neuronal integration of acoustic modulations and a unifying framework for hemispheric models critical to understanding the nature of human speech processing.

Methods

Stimulus Construction

All sentences were extracted from the Texas Instruments/Massachusetts Institute of Technology (TIMIT) database (2–4 s, 16 kHz) [85]. A set of 28 unique speakers from the database was selected, 14 were female and 14 male. In order to select male and female speakers with similar fundamental frequencies, speakers were selected such that the first peak of their FFT spectra was matched for male and females speakers. We then verified fundamental frequency using a sawtooth waveform inspired pitch estimator (Camacho & Harris, JASA 2008) showing a mean F0 with relatively low female pitch values (MN = 161.34 Hz, SD = 27) and relatively high male pitch (MN = 118.90 Hz, SD = 32.03). For

each speaker three unique sentences were processed and filtered in the modulation domain once with a low-pass cutoff of 2,3,4,5,6,7,8 Hz and once with a low-pass cutoff of 0.1867, 0.3733, 0.56, 0.7467, 0.933, 1.12, 1.3067 cycles/octave (28 speakers X 3 sentences X 14 filters). Stimuli length varied between 1.5–4.5 s (Mean: 2.396 s, SD: 0.576 s) containing 3–13 words (Mean: 7.40, SD: 2.21)

Modulation Domain Filtering

All modulation filtering was performed by a toolbox written in Matlab for the purposes of this manuscript and is freely available. Sound waveforms were first transformed into a time-frequency representation (spectrogram) using a filter-bank approach. Waveforms were filtered using 128 different frequency domain Gaussians that were designed to estimate cochlear critical bands [45]. Gaussian center frequencies logarithmically spanned the frequency space and the full width at half maximum (FWHM) corresponded to the equivalent rectangular bandwidth [86] (i.e. Bandwidth = 24.7*(F *4.37+1), where F is the center frequency in KHz). The output of the filter operation was then Hilbert transformed in order to extract the analytic amplitude and log-transformed. The output of this filter-bank processing is a spectrogram, which provides a time-frequency representation estimating the output of the cochlea. Next, spectrograms were filtered in the modulation domain, which essentially is a multiplication in the two dimensional frequency domain of the spectrogram matrix (i.e. 2D FFT). A given spectrogram was transformed to the modulation domain using a 2D FFT and then multiplied with a low pass filter (cosine ramp) removing all components above the cutoff (similarly to modulation filtering in Elliott et al. [38]). The filtered modulation domain was then inverse transformed back to a spectrogram (inverse 2D FFT) producing a new smoothed spectrogram with modulation frequencies only below the cutoff. All the filtering steps producing a smoothed (low-passed modulation) spectrogram are linear, invertible and relatively straightforward. The last operation in our filtering pipeline is transforming the new spectrogram into a corresponding time domain waveform. While such a direct transformation based only on spectrogram power estimates is not feasible there is a convex projection technique that iteratively produces a waveform which is maximally close to the desired spectrogram [87]. While the original procedure by Griffin and Lim requires iteratively inverting the short time fourier transform, the same logic can be used to invert a filter bank (e.g. [88]) and we follow this procedure with our inverted Gaussian filters with 10 iterations to produce a new waveform. In brief, each new smoothed spectrogram is inverted to a sound waveform by initially using random phase estimates, inverting each filter in the filter-bank given the power and phase estimates of that critical band and summing the output to produce a temporary waveform. This temporary waveform is then decomposed (forward filter-bank) in to a spectrogram that does not fully match the smoothed spectrogram but contains more accurate phase information than the previous iteration (i.e. the spectrogram more closely matches the desired spectrogram). The new phase information, together with the smoothed spectrogram power estimates are used to produce a new temporary waveform, and this procedure is repeated with updated phase estimates until there is a sufficiently small difference between the desired spectrogram and the spectrogram matching the sound waveform.

Psychophysical Experiments

In Experiment 1 (diotic, N=20) participants listened to two blocks of 84 unique pseudo randomly ordered sentences filtered either at 2,3,4,5,6,7,8 Hz low pass temporal modulation cutoff (and 11 cycles/octave spectral cutoff) or 0.1867, 0.3733, 0.5600, 0.7467, 0.9333, 1.1200, 1.3067 cycles/octave spectral cutoff (and 32 Hz temporal cutoff). After each sentence the participant was asked to rate the intelligibility from 1–4 then type out the sentence on the keyboard and finally select whether the speaker was male or female (keyboard response, 2 alternative forced choice). Each trial was self paced and continued after the participant responded to all the prompts, half the participants were prompted for intelligibility rating first and then voice pitch identification and half were prompted in the reverse order. Participants heard 84 unique sentences spoken by 28 unique speakers (3 sentences per speaker). For each speaker three cutoffs were randomly picked (either temporal or spectral) such that each filter cutoff appeared 6 different times (6 X 14 filters). In order to avoid priming effects of both speaker identity as well as sentence content, a sentence was never repeated within a block and a high cutoff filter never appeared before a lower cutoff filter for that speaker. In the second block the same sentences selected for the first block were repeated but in a different pseudorandom order, this manipulation ensured that the only priming effects of repeating a sentence or hearing more modulation content for that speaker were due to the second block. For each participant a different set of random filter and sentence permutations were selected.

In Experiment 2 (dichotic, N=60) participants listened to four blocks (2 blocks rating intelligibility and 2 marking male or female) of 120 pseudorandomly ordered sentences per block (40 unique sentences) constructed from the same filters as in Experiment 1 but were aligned in the audio stereo such that right ear and left ear received the same sentence but one ear received a filter at a higher cutoff (temporal dichotic pairs: [2,3], [3,4], [4,5], [5,6], [6,7] Hz and spectral dichotic pairs [0.1867,0.3733], [0.3733,0.5600], [0.5600.0.7564], [0.7564,0.9333], [0.9333,1.1200] cyc/oct). Each sentence was repeated once with a higher cutoff on the left ear (and lower on the right), once with a higher cutoff on the right ear (and lower in the left) and once as a diotic pair of the lower cutoff (e.g. [2,2] Hz). Half the sentences were presented first with a higher cutoff on the left (and then right and diotic) and half were presented first with a higher cutoff on the right (and then left and diotic). Each filter cutoff dichotic pair was heard in 8 unique sentences, four of which appeared first with higher modulation cutoffs on the left ear and four of which appeared first with higher modulation rates on the right ear. As in experiment 2 in order to avoid priming effects of both speaker identity as well as sentence content, a dichotic sentence did not repeat within a block and a higher cutoff dichotic pair filter did not appear before a lower cutoff dichotic pair filter for that speaker. For each participant a different set of random filter and sentence permutations were selected. Participants were asked to rate intelligibility and transcribe the sentence in two consecutive intelligibility blocks (the second block contained the same sentences as the first but in a different pseudo-random order) and in two consecutive voice pitch blocks, the participants were asked to mark the voice pitch of the speaker. Half the participants performed the intelligibility blocks first and then the voice pitch blocks and half the participants performed the voice pitch blocks first. Forty of the sixty participants first

performed an audiometer task to assess hearing thresholds in the left and right ears and to ensure differences were not greater than 5 dB between both ears.

Experiment 3 (MEG, N=19) was identical to Experiment 2 but only diotic stimuli were used, i.e. participants listened to four blocks (2 blocks rating intelligibility and 2 marking voice-pitch) of 120 pseudorandomly ordered sentences per block (40 unique sentences) using the same temporal and spectral cutoffs: 2,3,4,5,6 Hz or 0.1867, 0.3733, 0.5600, 0.7467, 0.9333 Cyc/Oct. Participants were only required to respond with an intelligibility rating (1–4) in the intelligibility blocks (no sentence transcription) and mark male or female in the voice pitch blocks. One participant was excluded due to excessive movement in the MEG leaving 19 out of the 20 recruited participants.

Experiment 4 (MEG, N=10) consisted of speech sentence stimuli as well as reversed sentences, identical to the pre-processed sentences in Experiment 1–3. Sentences were first reversed (audio vectors flipped left to right) and then filtered with identical procedures and cutoff values as in Experiment 3 (2,3,4,5,6 Hz or 0.1867, 0.3733, 0.5600, 0.7467, 0.9333 cyc/oct). Participants were asked to detect a tone that was embedded in 8% of stimuli while they listened to pseudo-randomly presented sentences. Stimuli with an embedded tone and stimuli with responses were removed from analysis. To ensure participants were paying attention to the non-reversed intelligible sentence stimuli, questions about the content were asked during the end of each block. An identical amount of reversed speech stimuli were presented per cutoff as in Experiment 3. One participant was excluded due to falling asleep repeatedly during the experiment leaving a total of 10 out of 11 recruited participants.

Across experiments 1–4, participants were counter balanced in the hand they were instructed to respond with, where half responded with the left hand and half with the right hand (transcriptions in Experiments 1 and 2 were performed freely with both hands). Experiments 1 and 2 were performed in a sound proof psychophysical booth and Experiment 2–4 was performed in the MEG scanner. Across all experiments participants first performed a practice block consisting of four unique exemplars that did not appear in the main experiment.

Experiment 5 (ECoG, N=8) was identical to Experiment 3 (MEG) but consisted of 80 pseudorandomly ordered sentences per block (40 unique sentences) and only one block of Intelligibility and one block of Voice Pitch Identification were administered per patient. Patients heard stimuli through a speaker presented in front of them in the hospital bedside environment.

Psychophysical Analysis

Intelligibility transcriptions were processes by an algorithm to assess how many words were correct in each sentence allowing up to two spelling mistakes (Levenshtein distance of 2) and was verified by a human rater providing minor corrections where appropriate. The proportion of words correct as well as correct voice pitch (male or female) responses are plotted in raw form across participants as mean and SEM in Figure 4. In order to assess statistically significant effects of Filter and Block (or Ear in the dichotic case) we first applied a within-subject (a.k.a "repeated measures") analysis of variance (ANOVA) and used

the raw scores as proportion (percent correct) data are notoriously susceptible to violation of ANOVA assumptions. While the Block and Ear conditions did not violate sphericity assumptions we did find a significant violation of sphericity for Filter as reported by Mauchly's Test for Spehricity (diotic Intelligibility W=0.0401, p=7.343e-5; diotic Voice Pitch Identification W=0.0212, p=1.88e-6; dichotic Intelligibility W=0.528, p=3.09e-5; dichotic Voice Pitch Identification W=0.724, p=0.0296). Instead of reporting the corrected ANOVA statistics with sphericity corrections we opted to use a non-parametric approach of analysis of variance using a factorial permutation test. We used 1000 permutations as recommended by the default value of the ez-package [50] but also verified that all effects held with 10000 permutations. Both ANOVA analysis, sphericity tests and final non-parametric factorial test were implemented in R using the ez-package [50] which is designed for within-subject analysis of factorial experiments (we used ezANOVA for the ANOVA and sphericity tests and ezPerm for the permutation test as well as ezBoot for validating data). When permutation test p-values equaled 0 we reported the more conservative estimate of $1/(m+1)$ where m is the number of permutations. It is noteworthy that all effects reported in the manuscript using the non-parametric approach mirrored the results of the ANOVA after correction for sphericity (i.e significant effects and non-significant effects). In addition to testing the effects *across participants* using a repeated measure factorial design for statistical assessment we also performed a within-subject analysis by fitting a logistic function for each participant's data using a maximum likelihood criterion with fixed gamma and lambda parameters implemented by the Palamedes Matlab toolbox [51]. The mean fitted curve across participants and SEM are depicted in Figure 4. These curves were then used to assess the exact range of modulations that showed a significant effect by performing a paired non-parametric Wilcoxon signed-rank test (implemented in Matlab). This test was performed on each data point of the curve and values with an alpha criterion of 0.05 (p<0.05) and to reduce type I errors (multiple comparisons) only a range of values showing continuous statistical significance was used, i.e. if any intermediate values did not pass significance the range of values reported was decreased to the largest block of continuous significant modulations.

No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to, and in experiments 2 and 3 larger than, those reported in previous publications[38,63,75–77,96].

In all experiments each participant listened to a different random order of speech stimuli (combination of filter, speaker and sentence). This randomization process ensured a random order of stimuli for each participant while reducing priming effects within block (e.g. for both speaker identity as well as sentence content, a high cutoff filter never appeared before a lower cutoff filter for that stimulus). In experiments 1–4, the order of tasks and response hand were counter balanced across participants (within experiment). Data collection and analysis were not performed blind to the conditions of the experiments.

Participants

Twenty participants participated in Experiment 1 (9 males; mean age 19.8, range 18–24), all participants were native speakers of American English and were right handed. Sixty

participants participated in Experiment 2 (27 males; mean age 24.67, range 18–46) all participants were native speakers of American English and all but three were right handed. Twenty participants participated in Experiment 3 (10 males; mean age 25.1, range 19–46) all participants were native speakers of American English and all were right handed. Ten participants participated in Experiment 4 (5 males; mean age 25 range 22–46) all participants were native speakers of American English and all but one participant were right handed. One participant in Experiment 3 was rejected due to consistent artifact and noise during the MEG recording. Participants in all experiments self-reported normal hearing and no neurological deficits. Six additional participants performed tasks but were removed from analysis due to a hearing threshold greater than 20 dB (in Experiment 1 and 2) and one participant that fell asleep during the MEG task. Participants were either paid for taking part in the study or received course credit and provided written informed consent. The protocol was approved by the local Institutional Review Board (New York University's Committee on Activities Involving Human Subjects) for all of these studies.

MEG Data Acquisition and Analysis

Neuromagnetic responses were recorded using a 157-channel whole-head axial gradiometer system (KIT, Kanazawa Institute of Technology, Japan) in a magnetically shielded room using a 1000 Hz sampling rate. Five electromagnetic coils were attached to the participant's head to monitor head the position during the MEG recording. The coils were localized to the MEG sensors, at three different time points: at the beginning of the experiment between the intelligibility and voice pitch tasks and at the end of the experiment. The position of the coils with respect to three anatomical landmarks: the naison, and left and right tragus were determined using 3D digitizer software (Source Signal Imaging, Inc.) and digitizing hardware (Polhemus, Inc.). This measurement allowed a coregistration of the MEG data with a MRI template brain (MNI). An online hardware bandpass filter between 1 and 200 Hz and a notch filter at 60 Hz was applied during data acquisition. Participants were asked to perform a practice block with their eyes open but performed the main experiment (Experiment 3) with their eyes closed in order to minimize eye blinks. The heard each filtered sentence and responded with one of four button presses in order to continue to the next trial.

Data processing and analysis was conducted using custom MATLAB scripts and functions from the FieldTrip toolbox[89]. For each participant's dataset, noisy channels were visually identified and rejected. Two denoising algorithms were run in sequence to regress out environmental noise measured with three reference sensors located away from the participant's head. First, a least squares estimate between data sensors and reference sensors was estimated with 5 minutes of data recorded with no participants (empty room) and these weights were then used to remove environmental noise in the recording session (Experiment 3) similarly to Adachi et al. [90]. Any remaining environmental noise components were removed using a time-shifted PCA procedure [91]. MEG signals were then zero-meaned and artifact components (eye blinks, heartbeat, stationary noise) were identified and removed using an independent component analysis (ICA) of the PCA space. Signals were then downsampled to 100 Hz for time efficiency and marked with onsets and offsets of each sentence.

Power estimates were extracted for the full length of the data using a frequency domain FWHM Gaussian filter, a Hilbert transform and log transforming the power. This procedure was conducted for every frequency range of interest (e.g. 0.1–8 Hz in Figure 5) and for all frequencies spanning 0.1–50 Hz (Figure S6) when examining the entire spectra (center frequencies were logarithmically spaced and fractional bandwidth was used).

Correlation analysis was performed between the filter type (ordinal 1–5) and the neural power estimate across the 240 sentences. Correlations were then averaged across participants, and significance was estimated by comparing the correlation with a null distribution of permutations. Each permutation was constructed by randomly reordering the filter type labels and repeating the same correlation and averaging previously described. This procedure was repeated 1000 times, providing a null distribution for assessing significance. When evaluating correlations on the level of sensor (topographic plots in Figure 5, Supplemental Figures 4–6, 9-11) power estimates were first averaged across time and then correlated with the filter type. When evaluating correlations over time (Figure 5 middle panel, Supplemental Figures 8-11) a sliding window approach was employed, averaging power across a window of 250 ms for each correlation. In order to correct for multiple comparisons, only significant clusters were included. For the temporal correlations, a minimum of two consecutive correlations in time were used as a threshold. For the spatial topographies (sensor and source) a cluster correction was used based on random permutations (1000) in order to only include clusters that had an average significant correlation.

In this approach no sensors were pre-selected but rather all correlations were computed within sensor. For spatial topography plots, only significant data that survived the permutation test and correction for multiple comparisons are shown. To clarify the process for each sensor: power estimates were averaged across time, producing one estimate per sentence, these estimates were then correlated with their corresponding filter cutoff producing one correlation per sensor. This process was repeated with randomly shuffled labels (cutoff values) for each sensor in order to estimate the permuted null distribution. Final topographies contain the average correlations across participants that were significant compared to the permuted null distribution (which was verified to hold a normal distribution) and survived a cluster size correction for multiple comparisons (based on permuted data). A laterality index (varying form −1 to 1) for topographies was computed as (L-R)/(L+R) where L corresponds to the significant correlations for left sensors and R corresponds to right sensors. In order to test for statistical significance we followed the same permutation approach of randomly shuffling labels and producing a null distribution of laterality indices. This distribution was found to violate normality so we opted to test significance on the un-normalized indices i.e. (L-R) instead of (L-R)/(L+R). The distribution was verified to be normal and the actual (L-R) laterality index was compared to the null distribution to report p-values. The same correlation procedure was followed in a separate pipeline for each vertex in source space after data was projected to estimate cortical sources.

Source reconstruction was employed using a cortically-constrained minimum norm estimation (MNE) [92]. This method approximates the cortical surface as a large number of current dipoles, and estimates the dipole's amplitude configuration with minimum overall

energy that generates the measured magnetic field. Power estimates in source space were then computed similarly to sensor space (0.1–8 Hz filter, Hilbert transform, log transform) and correlations were performed in sensor space with the same permutation and cluster procedures. In order to assess correlations in anatomical regions of interest (Supplemental Figure 7), uncorrected correlations were averaged per anatomical ROI (base on a fieldtrip provided atlas) in order to avoid statistical double dipping.

Intracranial Data Acquisition

Three (two patients with bilateral sterotactic implants and one patient with a left grid implant) patients were recruited from North Shore University Hospital and five patients (one left grid implant, one right grid implant and three implants with bilateral surface strips) from New York University Langone Medical Center. Patients were undergoing neurosurgical treatment for refractory epilepsy and consented to participate in cognitive tasks during lulls in clinical treatment. Electrode placement and treatment were dictated solely by the clinical needs of each patient. All participants gave written consent to participate in the study as well as an additional oral consent immediately before recording the task. The study protocol was approved by the NYU, NYU Langone Medical Center, North Shore University Hospital Committees on Human Research. Synchronized electrophysiology and trigger onsets for sentence stimuli were acquired to ensure time-locked analysis. At NYU Langone Medical Center all signals were acquired clinically with Nicolet ONE clinical amplifier (Natus) and sampled with a rate of 512 Hz. At North Shore University Hospital, signals were acquired with a Xltek EMU 128 clinical recording system (two stereotactic patients) with a sampling rate of 500 Hz and Xltek 512 channel clinical recording system (grid patient) with a sampling rate of 512 Hz.

Intracranial Electrode Localization

Electrode localization at NYULMC in participant space as well as MNI space was based on co-registering a pre-operative (no electrodes) and post-operative (with electrodes) structural MRI using a rigid-body transformation and then projecting electrodes to the surface of cortex (pre-operative segmented surface) to correct for edema induced shifts following previous procedures[93], registration to MNI space was based on a non-linear DARTEL algorithm[94]. Electrode localization at North Shore University Hospital was based on a rigid-body transformation between a preoperative MR and a postoperative CT with electrodes projected to the pial surface to correct for edema induced shift and registered to MNI coordinates based on a spherical warping transformation as implemented by freesurfer [95]. In both cases depth electrodes were not corrected for the edema induced shift.

Intracranial Data Analysis

Preprocessing of the data followed previous reported procedures[96] and included visual inspection and rejection of noisy electrodes (60 Hz line noise, poor contact with abnormal voltage readings). A common average referencing (CAR) approach was used by zero-meaning each channel and then removing the mean of all channels (the CAR) from each electrode (the CAR was also removed from noisy electrodes to ensure no residual signal was left). Electrodes over epileptiform regions were excluded as well as electrodes that were outside of cortical tissue. Ictal and epileptiform activity that spread across to other electrodes

was manually identified and any trials overlapping with the activity were excluded from analysis. Low frequency (0.1–8 Hz) or high frequency (70–150 Hz) activity was extracted with the same band-pass filtering approach as in MEG (i.e. frequency domain FWHM Gaussian filter, a Hilbert transform and log transformation for low power and percent change from baseline for high gamma). For low power correlations we followed a similar approach to MEG data analysis by correlating low frequency power with filter type (ordinal 1–5) and assessing significance by a permutation where all labels were randomly assigned and then correlated (1000 times) to create a null-distribution (Figure 6, Supplemental Figures 12-17). All electrodes were also tested for significant power change from baseline (paired t-test, $p < 0.05$) as a measure of signal response in the electrode. For analysis of correlations across patients and their asymmetry we focused on all electrodes that showed a significant increase of power from baseline and calculated their correlations focusing on positive correlations and aggregating correlations (reported mean, SEM and asymmetry significance in the Results) regardless of whether the correlations themselves were significant (in order to provide an unbiased estimate of signal and effect size). When reporting the percent of significant electrodes (Results), their location value and the asymmetry index (Supplementary Figures 16,17) we only report electrodes with a significant correlation via a permutation test ($p < 0.05$).

Statistical Reporting of Permutation Tests

While there is no consistent format for reporting statistical permutation tests, we chose to follow the approach of Collingridge[97] by employing the nomenclature "$P_{\text{test-statistic}} = X$" to denote the permutation test (P) of a specific test-statistic (subscript) with the real observed test-statistic reported as X. Given that Confidence Intervals (CI) are not well-defined nor meaningful for a permutation test (and can vary in meaning depending on the reported parametric statistic) we chose to report the 5% and 95% percentiles of the null hypothesis distribution (permutation results) which reflect the confidence intervals of the test statistic under random labels and if X falls outside the CI range we can reject the null hypothesis of the permutation test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Broca P Remarques sur le siege de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletins et mémoires de la Société Anatomique de Paris 36, 330–356 (1861).

2. Wernicke C Symptomenkomplex. Eine psychologische Studie auf anatomischer Basis. Cohn und Weigert, Breslau (1874).

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

4. Hagoort P & Indefrey P The neurobiology of language beyond single words. Annu. Rev. Neurosci 37, 347–362 (2014). [PubMed: 24905595]

5. Friederici AD The brain basis of language processing: from structure to function. Physiol. Rev 91, 1357–1392 (2011). [PubMed: 22013214]

6. Rauschecker JP & Scott SK Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience 12, 718–724 (2009). [PubMed: 19471271]

7. Price CJ A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. NeuroImage 62, 816–847 (2012). [PubMed: 22584224]

8. Goetz CG Textbook of clinical neurology 355, (2007).

9. Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA & Hudspeth AJ Principles of neural science 4, 457–469 (2000).

10. Bozic M, Tyler LK, Ives DT, Randall B & Marslen-Wilson WD Bihemispheric foundations for human speech comprehension. Proc. Natl. Acad. Sci. U.S.A 107, 17439–17444 (2010). [PubMed: 20855587]

11. Overath T, McDermott JH, Zarate JM & Poeppel D The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. Nature Neuroscience 18, 903–911 (2015). [PubMed: 25984889]

12. Zatorre RJ & Gandour JT Neural specializations for speech and pitch: moving beyond the dichotomies. Philos. Trans. R. Soc. Lond., B, Biol. Sci 363, 1087–1104 (2008). [PubMed: 17890188]

13. McGettigan C & Scott SK Cortical asymmetries in speech perception: what's wrong, what"s right and what"s left? Trends in Cognitive Sciences 16, 269–276 (2012). [PubMed: 22521208]

14. Efron R Temporal Perception, Aphasia and Déjà Vu. Brain 86, 403–424 (1963). [PubMed: 14063892]

15. Schwartz J & Tallal P Rate of acoustic change may underlie hemispheric specialization for speech perception. Science 207, 1380–1381 (1980). [PubMed: 7355297]

16. Tallal P & Piercy M Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241, 468–469 (1973). [PubMed: 4705758]

17. Ross ED & Mesulam M-M Dominant language functions of the right hemisphere?: Prosody and emotional gesturing. Arch Neurol 36, 144–148 (1979). [PubMed: 435134]

18. Tucker DM, Watson RT & Heilman KM Discrimination and evocation of affectively intoned speech in patients with right parietal disease. Neurology 27, 947–947 (1977). [PubMed: 561908]

19. Robin DA, Tranel D & Damasio H Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. Brain Lang 39, 539–555 (1990). [PubMed: 2076495]

20. Zatorre RJ Pitch perception of complex tones and human temporal-lobe function. The Journal of the Acoustical Society of America 84, 566–572 (1988). [PubMed: 3170948]

21. Zatorre RJ, Evans AC, Meyer E & Gjedde A Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849 (1992). [PubMed: 1589767]

22. Belin P et al. Lateralization of speech and auditory temporal processing. J Cogn Neurosci 10, 536–540 (1998). [PubMed: 9712682]

23. Zatorre RJ, Belin P & Penhune VB Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences 6, 37–46 (2002). [PubMed: 11849614]

24. Poeppel D The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41, 245–255 (2003).

25. Scott SK & McGettigan C Do temporal processes underlie left hemisphere dominance in speech perception? Brain Lang 127, 36–45 (2013). [PubMed: 24125574]

26. Gervain J Near-infrared spectroscopy: recent advances in infant speech perception and language acquisition research. Frontiers in Psychology 5, 481 (2014). [PubMed: 24904494]

27. Luo H & Poeppel D Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. Frontiers in Psychology 3, 170 (2012). [PubMed: 22666214]

28. Okamoto H & Kakigi R Hemispheric Asymmetry of Auditory Mismatch Negativity Elicited by Spectral and Temporal Deviants: A Magnetoencephalographic Study. Brain Topogr 28, 471–478 (2015). [PubMed: 24366694]

29. Thompson EC et al. Hemispheric Asymmetry of Endogenous Neural Oscillations in Young Children: Implications for Hearing Speech In Noise. Sci. Rep 6, 19737 (2016). [PubMed: 26804355]

30. Ratcliffe VF & Reby D Orienting Asymmetries in Dogs' Responses to Different Communicatory Components of Human Speech. Current Biology 24, 2908–2912 (2014). [PubMed: 25454584]

31. Washington SD & Tillinghast JS Conjugating time and frequency: hemispheric specialization, acoustic uncertainty, and the mustached bat. Front. Neurosci 9, (2015).

32. Washington SD & Kanwal JS Sex-dependent hemispheric asymmetries for processing frequency-modulated sounds in the primary auditory cortex of the mustached bat. J. Neurophysiol 108, 1548–1566 (2012). [PubMed: 22649207]

33. L DeValois R & K DeValois K Spatial Vision (Oxford University Press USA, 1990).

34. Shapley R & Lennie P Spatial frequency analysis in the visual system. Annu. Rev. Neurosci 8, 547–581 (1985). [PubMed: 3920946]

35. Kowalski N, Depireux DA & Shamma SA Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J. Neurophysiol 76, 3503–3523 (1996). [PubMed: 8930289]

36. Woolley SMN, Fremouw TE, Hsu A & Theunissen FE Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nature Neuroscience 8, 1371–1379 (2005). [PubMed: 16136039]

37. Elhilali M, Chi T & Shamma SA A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. Speech Communication 41, 331–348 (2003).

38. Elliott TM & Theunissen FE The Modulation Transfer Function for Speech Intelligibility. PLoS Comput Biol 5, e1000302 (2009). [PubMed: 19266016]

39. Arnal LH, Flinker A, Kleinschmidt A, Giraud A-L & Poeppel D Human Screams Occupy a Privileged Niche in the Communication Soundscape. Current Biology 25, 2051–2056 (2015). [PubMed: 26190070]

40. Gross J et al. Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. Plos Biol 11, e1001752 (2013). [PubMed: 24391472]

41. Schönwiesner M & Zatorre RJ Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc. Natl. Acad. Sci. U.S.A 106, 14611–14616 (2009). [PubMed: 19667199]

42. Santoro R et al. Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. PLoS Comput Biol 10, e1003412 (2014). [PubMed: 24391486]

43. Pasley BN et al. Reconstructing Speech from Human Auditory Cortex. Plos Biol 10, e1001251 (2012). [PubMed: 22303281]

44. Poeppel D Pure word deafness and the bilateral processing of the speech code. Cognitive Science 25, 679–693 (2001).

45. Moore BCJ Auditory filter shapes derived in simultaneous and forward masking. The Journal of the Acoustical Society of America 70, 1003 (1981). [PubMed: 7288037]

46. Ghitza O On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. The Journal of the Acoustical Society of America 110, 1628–1640 (2001). [PubMed: 11572372]

47. Drullman R, Festen JM & Plomp R Effect of reducing slow temporal modulations on speech reception. The Journal of the Acoustical Society of America 95, 2670–2680 (1994). [PubMed: 8207140]

48. Chi T, Gao Y, Guyton MC, Ru P & Shamma S Spectro-temporal modulation transfer functions and speech intelligibility. The Journal of the Acoustical Society of America 106, 2719 (1999). [PubMed: 10573888]

49. Jaeger TF Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language 59, 434–446 (2008). [PubMed: 19884961]

50. Lawrence MA & Lawrence MMA Package 'ez' (2016).

51. Prins N & Kingdom FAA Matlab routines for analyzing psychophysical data http://www.palamedestoolbox.org.

52. Kimura D Functional Asymmetry of the Brain in Dichotic Listening. Cortex 3, 163–178 (1967).

53. Shamma S On the role of space and time in auditory processing. Trends in Cognitive Sciences 5, 340–348 (2001). [PubMed: 11477003]

54. Venezia JH, Hickok G & Richards VM Auditory 'bubbles': Efficient classification of the spectrotemporal modulations essential for speech intelligibility. The Journal of the Acoustical Society of America 140, 1072–1088 (2016). [PubMed: 27586738]

55. Houtgast T & Steeneken HJM A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. The Journal of the Acoustical Society of America 77, 1069–1077 (1985).

56. Ding N et al. Temporal modulations in speech and music. Neuroscience & Biobehavioral Reviews (2017). doi:10.1016/j.neubiorev.2017.02.011

57. Shannon RV, Zeng FG, Kamath V, Wygonski J & Ekelid M Speech recognition with primarily temporal cues. Science 270, 303–304 (1995). [PubMed: 7569981]

58. Gabor D Theory of communication. Part 1: The analysis of information. Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering 93, 429–441 (1946).

59. Joris PX, Schreiner CE & Ress A Neural Processing of Amplitude-Modulated Sounds. Physiol. Rev 84, 541–577 (2004). [PubMed: 15044682]

60. Overath T, Zhang Y, Sanes DH & Poeppel D Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. J. Neurophysiol 107, 2042–2056 (2012). [PubMed: 22298830]

61. Zatorre RJ & Belin P Spectral and temporal processing in human auditory cortex 11, 946–953 (2001).

62. Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J & Chauvel P Temporal envelope processing in the human left and right auditory cortices 14, 731–740 (2004).

63. Okamoto H, Stracke H, Draganova R & Pantev C Hemispheric Asymmetry of Auditory Evoked Fields Elicited by Spectral versus Temporal Stimulus Change 19, bhn245–2297 (2009).

64. Wang Y et al. Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J. Neurophysiol 107, 2033–2041 (2012). [PubMed: 21975451]

65. Schönwiesner M, Rübsamen R & Cramon, Von, D. Y. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. European Journal of Neuroscience 22, 1521–1528 (2005). [PubMed: 16190905]

66. Jamison HL, Watkins KE, Bishop DVM & Matthews PM Hemispheric specialization for processing auditory nonspeech stimuli 16, 1266–1275 (2006).

67. Boemio A, Fromm S, Braun A & Poeppel D Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nature Neuroscience 8, 389–395 (2005). [PubMed: 15723061]

68. Overath T, Kumar S, Kriegstein von, K. & Griffiths TD Encoding of spectral correlation over time in auditory cortex. J. Neurosci 28, 13268–13273 (2008). [PubMed: 19052218]

69. Hyde KL, Peretz I & Zatorre RJ Evidence for the role of the right auditory cortex in fine pitch resolution. Neuropsychologia 46, 632–639 (2008). [PubMed: 17959204]

70. Liégeois-Chauvel C, de Graaf JB, Laguitton V & Chauvel P Specialization of left auditory cortex for speech perception in man depends on temporal coding 9, 484–496 (1999).

71. Kriegstein von, K., Smith DRR, Patterson RD, Kiebel SJ & Griffiths TD How the human brain recognizes speech in the context of changing speakers. J. Neurosci 30, 629–638 (2010). [PubMed: 20071527]

72. Arsenault JS & Buchsbaum BR Distributed Neural Representations of Phonological Features during Speech Perception. J. Neurosci 35, 634–642 (2015). [PubMed: 25589757]

73. Millman RE, Woods WP & Quinlan PT Functional asymmetries in the representation of noise-vocoded speech. NeuroImage 54, 2364–2373 (2011). [PubMed: 20946961]

74. Obleser J, Eisner F & Kotz SA Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J. Neurosci 28, 8116–8123 (2008). [PubMed: 18685036]

75. Morillon B et al. Neurophysiological origin of human brain asymmetry for speech and language. Proc. Natl. Acad. Sci. U.S.A 107, 18688–18693 (2010). [PubMed: 20956297]

76. Peelle JE, Gross J & Davis MH Phase-locked responses to speech in human auditory cortex are enhanced during comprehension 23, 1378–1387 (2013).

77. Luo H & Poeppel D Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. Neuron 54, 1001–1010 (2007). [PubMed: 17582338]

78. Thompson EC et al. Hemispheric Asymmetry of Endogenous Neural Oscillations in Young Children: Implications for Hearing Speech In Noise. Sci. Rep 6, 19737 (2016). [PubMed: 26804355]

79. Morillon B, Liégeois-Chauvel C, Arnal LH, Bénar C-G & Giraud A-L Asymmetric Function of Theta and Gamma Activity in Syllable Processing: An Intra-Cortical Study. Frontiers in Psychology 3, 248 (2012). [PubMed: 22833730]

80. Giraud A-L et al. Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. Neuron 56, 1127–1134 (2007). [PubMed: 18093532]

81. Giraud A-L & Poeppel D Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience 15, 511–517 (2012). [PubMed: 22426255]

82. Herdener M et al. Spatial representations of temporal and spectral sound cues in human auditory cortex. Cortex 49, 2822–2833 (2013). [PubMed: 23706955]

83. Barton B, Venezia JH, Saberi K, Hickok G & Brewer AA Orthogonal acoustic dimensions define auditory field maps in human cortex. Proc. Natl. Acad. Sci. U.S.A 109, 20738–20743 (2012). [PubMed: 23188798]

84. Joanisse MF Sensitivity of human auditory cortex to rapid frequency modulation revealed by multivariate representational similarity analysis. Front. Neurosci 8, (2014).

85. Garofolo JS et al. TIMIT acoustic-phonetic continuous speech corpus. Linguistic data consortium, Philadelphia 33, (1993).

86. Moore BC J. Hearing (1995).

87. Griffin D & Jae Lim. Signal estimation from modified short-time Fourier transform. IEEE Trans. Acoust., Speech, Signal Process 32, 236–243 (1984).

88. Chi T & Shamma S NSL Matlab Toolbox Maryland: Neural Systems Lab (2005).

89. Oostenveld R, Fries P, Maris E & Schoffelen J-M FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience 2011, 1–9 (2010).

90. Adachi Y, Shimogawara M, Higuchi M, Haruta Y & Ochiai M Reduction of nonperiodical environmental magnetic noise in MEG measurement by continuously adjusted least squares method. IEEE Transactions on Applied Superconductivity 11, 669–672 (2001).

91. de Cheveigné A & Simon JZ Denoising based on time-shift PCA. Journal of Neuroscience Methods 165, 297–305 (2007). [PubMed: 17624443]

92. Dale AM et al. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67 (2000). [PubMed: 10798392]

93. Yang AI et al. Localization of dense intracranial electrode arrays using magnetic resonance imaging. NeuroImage 63, 157–165 (2012). [PubMed: 22759995]

94. Ashburner J A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113 (2007). [PubMed: 17761438]

95. Groppe DM et al. iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. Journal of Neuroscience Methods 281, 40–48 (2017). [PubMed: 28192130]

96. Flinker A et al. Redefining the role of Broca's area in speech. Proc. Natl. Acad. Sci. U.S.A 201414491 (2015). doi:10.1073/pnas.1414491112
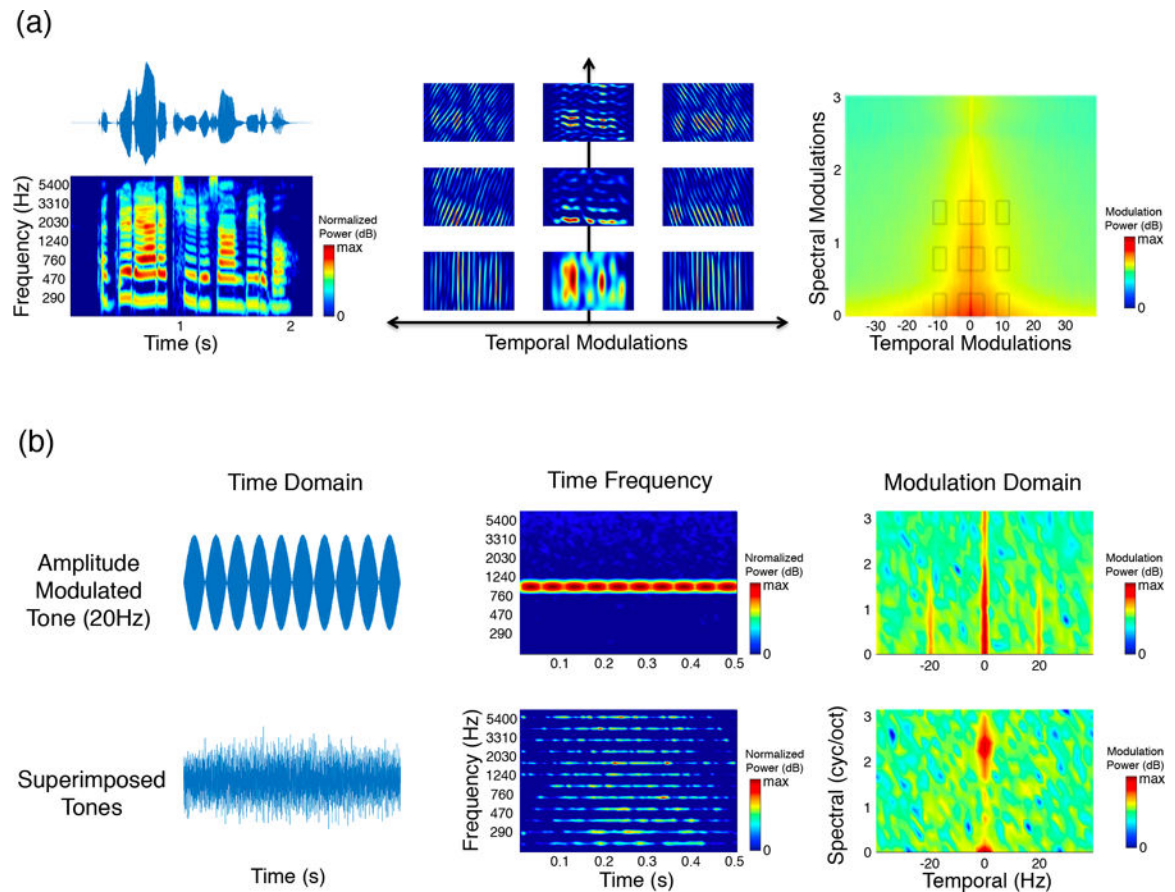
Figure 1.

Time, Time-Frequency, and Modulation domain representations for sound waveforms. (a) Sound waveform of a spoken sentence (left panel, top) is shown along with its corresponding spectrogram representation (left panel, bottom). The spectrogram can be represented as a decomposition in the modulation domain (middle panel) of horizontal (temporal, typically as cycles per second) and vertical (spectral, typically as cycles per octave) modulations. The degree (power intensity) of temporal and spectral modulations in the spectrogram is depicted in the right panel (showing the average modulation spectra of all our speech material, N=84). Superimposed gray squares correspond to the approximate temporal and spectral ranges shown in the middle panel. (b) To provide better intuitive insight into these representations, two artificially created audio signals are shown with a temporal modulation peak (left panel, top waveform) and a spectral modulation peak (left panel, bottom waveform). For both audio signals representations are shown in the time domain (left), the time-frequency domain (middle), and the modulation domain (right).
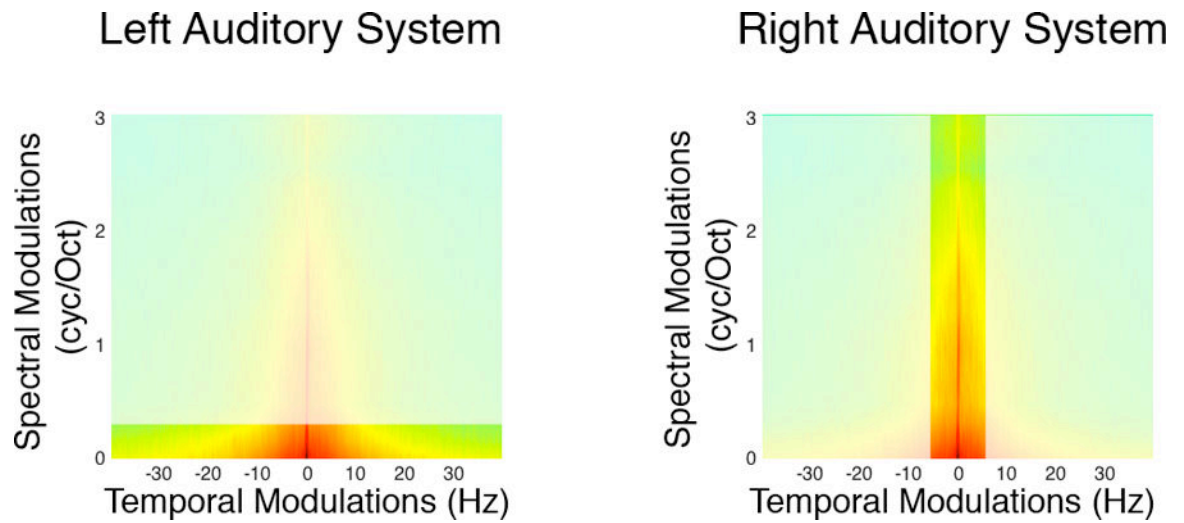
Figure 2.
A schematic depiction of the modulation asymmetry hypothesis whereby the left auditory system integrates a wide range of temporal modulations but a limited range of spectral modulations and the right auditory system integrates a wide range of spectral modulations but a limited range of temporal modulations.
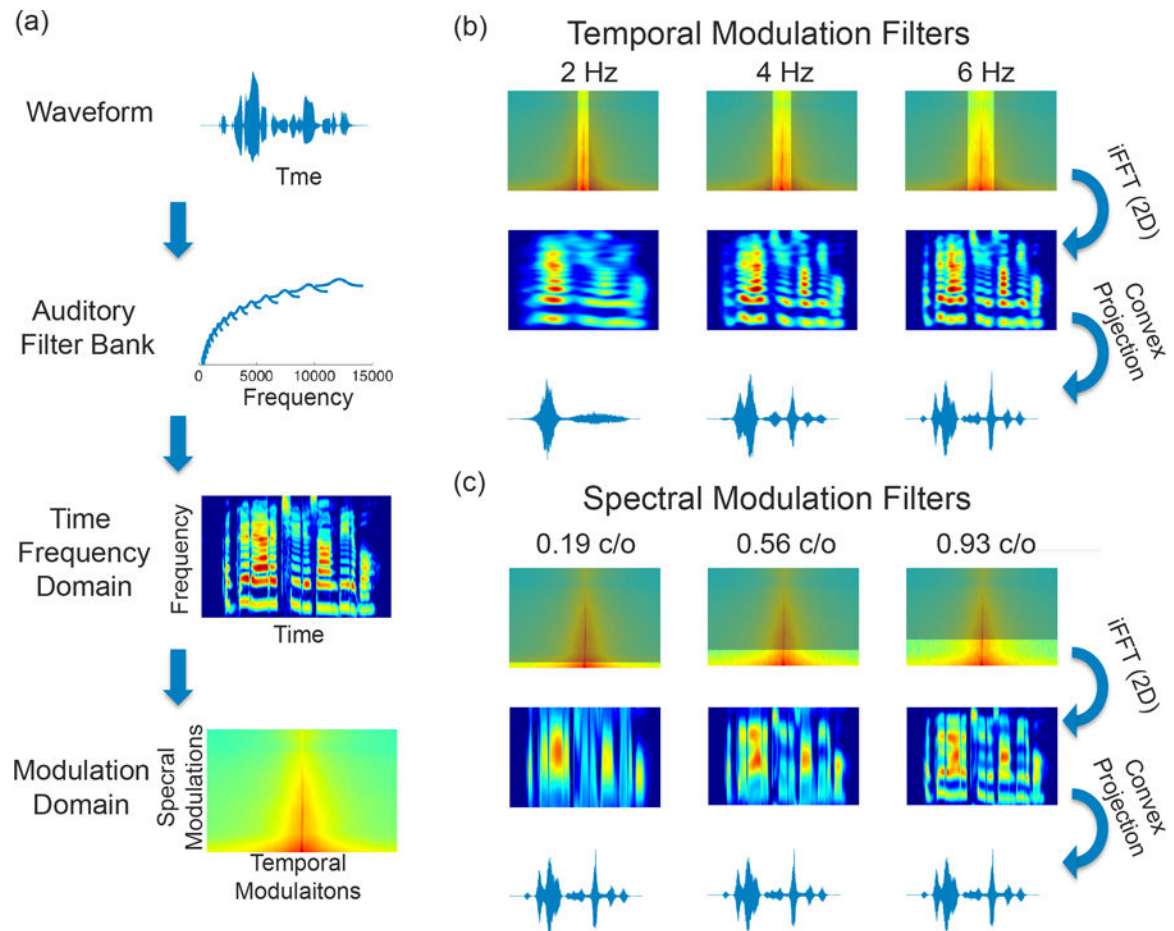
Figure 3.
Overview of the filtering technique used to produce modulation domain filtered stimuli. (a) An audio waveform is filtered in the frequency domain using a cochlear filter bank, and the subsequent spectrogram is Fourier transformed, two dimensionally, to produce a modulation domain representation. In the modulation domain, signals are low-pass filtered using a temporal (b) or spectral (c) cutoff and inversed Fourier transformed to produce the desired modulation-limited spectrogram. Finally, an iterative convex projection technique is employed to produce an audio signal that maximally matches the desired modulation-limited spectrogram.
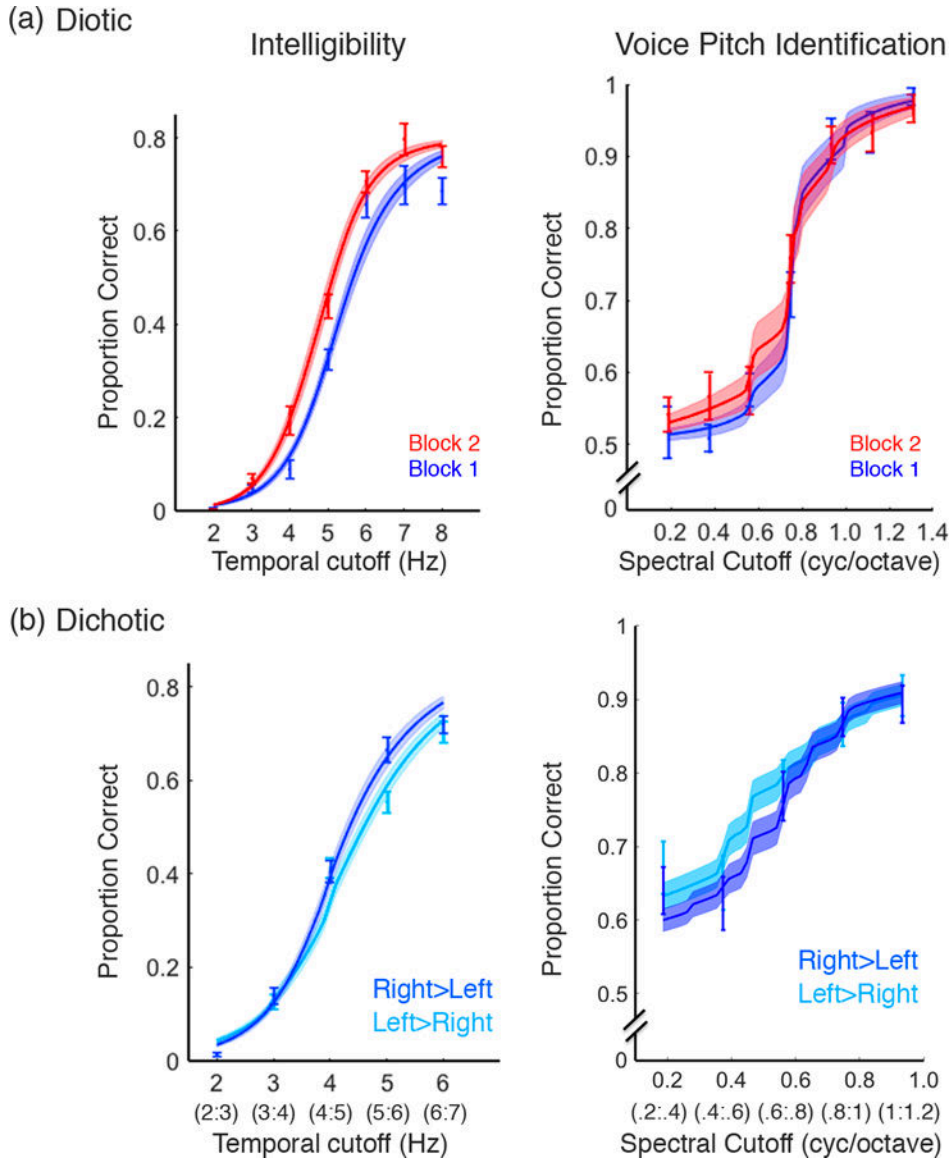
Figure 4.
Psychophysical performance as a function of temporal and spectral modulations in two separate experiments (diotic and dichotic). (a) Intelligibility (proportion of words transcribed correctly; N=20) at different temporal modulation cutoffs (left) and voice pitch identification (percent of correct responses; same participants N=20) at different spectral modulation cutoffs (right). Raw data is shown for each modulation value with mean and standard error of the mean across participants depicted in the error bars. Additionally the within-subject curve fits are shown with a solid curve depicting the mean and standard error of the mean across in the shaded area. Stimuli are repeated again in a subsequent block (block 2, in red). (b) The same tasks are used in a separate experiment employing dichotic stimulus presentation (N=60), with a different level of modulation information in each ear. Each value on the x-axis tick represents two different modulation values presented to each ear which are denoted in parentheses below the tick. The color code denotes in which ear the higher of the

two values was presented (Blue, right ear was presented with a higher value; Cyan, left ear was presented with a higher value). Higher intelligibility is seen for sentences with more temporal modulations in the right ear compared to the left (dark blue curve). This right ear advantage is evident for temporal modulations but not spectral modulations.
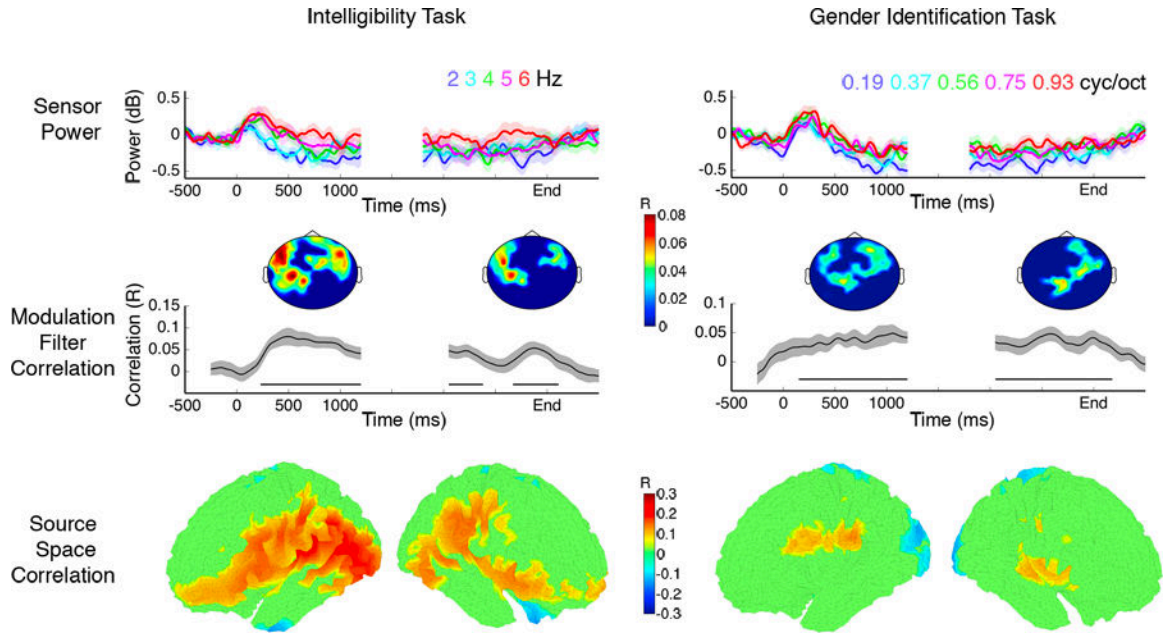
Figure 5.
MEG results showing significant correlations between neural power and degree of stimulus modulation. For each task, the average neural power for each modulation cutoff is shown (top) locked to the onset of a sentence (left most x-axis) as well as the offset (right most x-axis) with mean across sensors and participants and shaded error bars depicting standard error of the mean across participants (N=19). The correlations are shown in the middle panel as a function of time (mean across sensors and participants and shaded SEM across participants, N=19) and spatial sensor locations (mean across time and participants, N=19) in the topography plots above. Power estimates are projected to source space using a minimum norm estimate, and significant correlations in source space are shown in the bottom panels (mean correlations across participant, N=19).
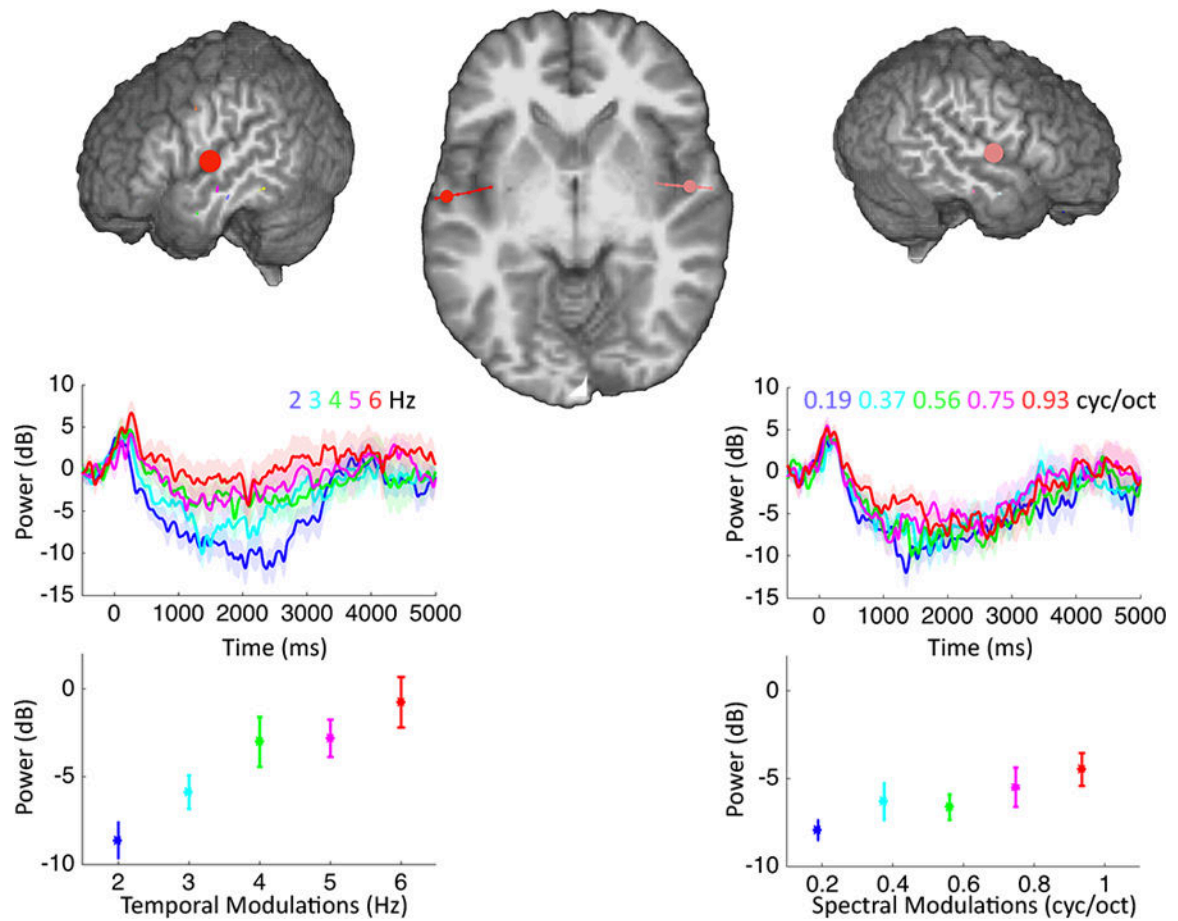
Figure 6.
Intracranial ECoG recordings in a patient with bilateral stereotactic depth electrodes, sampling superior temporal cortices. Reconstruction of electrode locations (MNI coordinates = [−60.17, −7.76, 2.82] , [52.81, −7.95, 2.64]) are shown (top) for an axial slice Z=3. Two significant electrodes with low frequency (0.1–8 Hz) mean (shaded SEM across trials, N=16) power traces shown for each modulation filter (middle) as well as average power (bottom) collapsed across time with mean and SEM across trials (N=16).