Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries

Adeen Flinker1,2*, Werner K. Doyle3, Ashesh D. Mehta4, Orrin Devinsky2 and David Poeppel1,5

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, such as what sound features or processing principles underlie laterality. Recent findings across species (humans, canines and 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 behavioural and neurophysiological measures. We show how psychophysical judgements 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.

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 Broca1 and Wernicke2 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 principles3–7. In contrast to the historically established one-size-fits-all view on lateralization of speech and language—a perspective that still prevails in the clinical literature8,9—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 processing9), that is the mapping from acoustic input to the internal representations (informally speaking, words) that form the basis for language processing (and there is recent fMRI data supporting this hypothesis10). However, the functional and computational differences between the two hemispheres with respect to auditory processing remain incompletely understood and vigorously contested11,12.

A series of influential behavioural studies provided suggestive evidence that the left hemisphere is sensitive to rapidly changing auditory cues13–18. This sensitivity manifested when temporal intervals between stimuli were reduced below 100–150 ms in aphasic patients (temporal order judgement)14, children with developmental delays (discrimination)15, as well as a sensitivity to the length of formant transition in healthy listeners (dichotic listening)16. In parallel, clinical reports argued that focal damage to the right hemisphere produces impairments in processing slower changing prosodic cues in speech17,18, as well the ability to discriminate spectral information19,20.

Based on these findings as well as on foundational neuroimaging studies21–23, a framework emerged whereby left hemisphere structures—specifically language-related regions, principally the temporal lobe but also the inferior frontal gyrus—are more sensitive to temporal cues, whereas right hemisphere structures are more sensitive to spectral cues24. 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 (~20–80 ms) and right regions over longer timescales (~150–300 ms), asymmetric sampling in time24. A third view associates lateralization with function: if representations are linguistic, they are processed by the dominant, typically left hemisphere15,25. Notwithstanding the existing neural and behavioural 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 by building on 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 humans26–29, dogs30 and mustached bats31,32 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 results23,24 differ in their implementational specificity regarding the underlying cortical tissue.
To address these gaps, we reframe and test the asymmetry hypothesis in the modulation domain, an acoustic ‘space’ that has been developed to mirror successful analysis approaches in the visual domain. The modulation domain reflects energy fluctuations that vary across the temporal and spectral axes of a spectrographic representation (that is, temporal and spectral modulations of power in the time–frequency representation), similarly to horizontal and vertical spatial gratings that comprise an image (Fig. 1a). The modulation domain quantifies temporal and spectral acoustic features (Fig. 1b) and has been used to investigate communication vocalizations (and link them to neural mechanisms) in the ferret, zebra finch and in humans. Recently, spectral and temporal modulations have provided novel topographic maps of the human auditory cortex, and auditory models capitalizing on this domain have proved to be more accurate at reconstructing neural activity both in electrophysiology and neuroimaging. We take these recent successes using the modulation domain in neural representation as providing a new opportunity to investigate auditory cortical asymmetries, building on the prevailing auditory theories, namely the temporal versus spectral view, and the asymmetric sampling in time hypothesis.

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) (Fig. 2). It is noteworthy that both hemispheres have a range of temporal and spectral modulations that overlap—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 (that is, integration of neuronal inputs from subpopulations). 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) and 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...
Fig. 2 | The modulation asymmetry hypothesis. 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.

The filtered representation (Fig. 3a and Supplementary Fig. 1; see also ‘Modulation domain filtering’ in Methods). We closely follow the framework introduced previously, but we diverge from that approach in critical ways in how we decompose and resynthesize the signal in the frequency domain. While the other study employed a linear frequency scale (short time Fourier transform), we employ a logarithmic frequency scale following the frequency distribution of the cochlea using a filter bank decomposition (see ‘Modulation domain filtering’ in Methods). We used this technique on a wide selection of English sentences and different speakers in order to control parametrically the degree of temporal modulations (Fig. 3b) or spectral modulations (Fig. 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. 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. 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 per octave spectrally), we first examined behavioural 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, and here we report psychometric curves on voice pitch identification as well as intelligibility in the critical modulation range for speech (that is, 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 \( P_{\text{value}} = 0.0795, 95\% \text{ CI of null hypothesis statistic} = (0.0011, 2.689) \); Fig. 4a, right 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 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 \( P_{\text{value}} = 2.497, P = 0.079, 95\% \text{ CI of null hypothesis statistic} = (0.0314, 2.916) \); Fig. 4b right panel). But, an analysis of the fitted within-subject psychometric curves showed significant differences for values between 0.373 and 0.541 cycle per octave (non-parametric Wilcoxon signed-rank test \( P < 0.05 \) for all paired tests).

To further quantify and assess neurally the sensitivity of each hemisphere to temporal and spectral modulations, we next used the same stimuli and the dichotic 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 cycle per octave) content (Fig. 5 top panel and Supplementary Figs. 4 and 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 converges by the time the sentence has ended (Fig. 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 \( (P < 0.05, \text{ permutation test}) \) with stimulus temporal modulation cutoffs across time, showing an increase in neural power as the participants heard sentences with increased each participant using a logistic function. The mean and standard error distribution of curve fits across participants are plotted as a continuous shaded line in Fig. 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 between 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 per octave spectral modulations (Fig. 4a, right panel). There was no effect of block (within-subject non-parametric factorial permutation test main effect of block \( P_{\text{value}} = 0.1195, P = 0.992, 95\% \text{ CI of null hypothesis statistic} = (0.0795, 2.96) \)), 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 cycle per octave; see Supplementary Fig. 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. 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 (for example, 3 Hz in the one ear and 4 Hz in the other, and so on). In a separate cohort of participants (n = 60), we ‘Psychophysical experiments’ in Methods, we found a significant behavioural 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 \( P_{\text{value}} = 18.963744, P < 0.001, 95\% \text{ CI of null hypothesis statistic} = (0.05455, 2.689) \); Fig. 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 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 \( P_{\text{value}} = 2.497, P = 0.079, 95\% \text{ CI of null hypothesis statistic} = (0.0314, 2.916) \); Fig. 4b right panel). But, an analysis of the fitted within-subject psychometric curves showed significant differences for values between 0.373 and 0.541 cycle per octave (non-parametric Wilcoxon signed-rank test \( P < 0.05 \) for all paired tests).

NATURE HUMAN BEHAVIOUR | VOL 3 | APRIL 2019 | 393–405 | www.nature.com/nathumbehav
temporal modulation content (Fig. 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 110 ms post stimulus offset ($R = 0.0231$) (Fig. 5 middle panel; black horizontal lines denote significance). Similarly, correlations with spectral modulations were strongest (maximal value of $R = 0.0489$ at 1,060 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 (Fig. 5, middle panel, left topoplot) and second half of the sentence (Fig. 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 (Fig. 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 right laterality where right sensor power was greater than left (permutation test $P_{\text{difference of mean, b}} = 0.00498$, $P = 0.016$, 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 Fig. 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, that is, correlated neural power with stimulus modulation cutoff and found a significant right laterality where right sensor power was greater than left (permutation test $P_{\text{difference of mean, b}} = 0.00199$, $P = 0.003$, 95% CI of null hypothesis statistic $= (−0.0059, 0.0056)$) but the spectral modulation asymmetry did not pass significance (permutation test, $P_{\text{difference of mean, b}} = 0.0025$, $P = 0.163$, 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, $P_{\text{difference of mean}_R} = 0.0395$, $P = 0.0180$, 95% CI of null hypothesis statistic = (−0.0041, 0.000)) as well as spectral modulations (permutation test, $P_{\text{difference of mean}_R} = −0.05846$, $P = 0.0170$, 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 and 0.0972 respectively; spectral modulation mean correlation for left and right hemispheres = 0.040366 and 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) and (0.03734–0.05915) respectively; spectral modulation correlation range across high-order regions for left and right hemispheres = (−0.002693–0.032166) and (0.027008–0.06067) respectively; Supplementary Fig. 7).

In addition to a strong relationship between the stimulus modulation space and neural responses, we tested for significant correlations

**Fig. 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 s.e.m. across participants depicted in the error bars. Additionally, the within-subject curve fits are shown, with a solid curve depicting the mean and s.e.m. 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 represents two different modulation values presented to each ear, which are denoted in parentheses below the tick. The colour 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.
between neural responses and the participants’ behavioural 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 Fig. 8). Correlations with the behavioural responses (male or female) in the voice pitch identification task did not pass significance (Supplementary Fig. 9), providing evidence that cortical responses are sensitive to the spectral content during the task. To verify that the correlations in the intelligibility task were due to the temporal modulations rather than the intelligibility of speech content, 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 ‘Psychophysical experiments’ in Methods). The analysis revealed significant correlations with a left-hemisphere biased topography (Supplementary Figs. 10 and 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, $P_{\text{correlation}} = 0.413$, $P < 0.001$, 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, $P_{\text{correlation}} = 0.241$, $P = 0.015$, 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 (Fig. 6 and Supplementary Figs. 12–14); but these patterns could also be observed when examining correlations with neural high gamma activity (Supplementary Fig. 15). An analysis of the correlations across all electrodes and participants showed that temporal modulation correlations were larger in the left hemisphere (mean = 0.1350, s.e.m. = 0.0070) than the right hemisphere (mean = 0.1011, s.e.m. = 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 (mean = 0.1163, s.e.m. = 0.0063) similar to those in the right hemisphere (mean = 0.1199, s.e.m. = 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 Figs. 16 and 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 Figs. 16 and 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 (Fig. 2), signal processing techniques to manipulate natural stimuli in this space (Fig. 3), a psychophysical mapping of this space, as well as behavioural and multimodal (MEG and 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 as well
as permitting unified computations across different cortical modalities. Due to the complexity of manipulating and reconstructing acoustic signals in this space, only a few studies have employed this approach to investigate speech intelligibility. Our approach builds on the pipeline proposed previously (that is, linear time–frequency decomposition, modulation filtering and iterative convex projection) but diverges by employing a cochlear filter bank that represents frequency in a logarithmic scale (similarly to another study), 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 per 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. 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 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 cycle per octave (Supplementary Fig. 3), which fits well with our behavioural results showing a clear increase in performance at 0.75 cycle per octave, which approaches ceiling above 1 cycle per octave (Fig. 4 and Supplementary Fig. 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. 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 or as a difference in temporal window sizes over which information is integrated. The time–frequency resolution argument draws a parallel between cortical computation and the uncertainty principle in decomposing an auditory signal. The acoustic (that is, Heisenberg–Gabor) uncertainty principle provides a theoretical limit on the degree to which 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. Proposals in both humans and other
model organisms\textsuperscript{41} 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 on a biologically plausible pathway: we first represent an auditory signal using a filter bank mimicking the log-frequency spacing of the cochlea\textsuperscript{4} 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 (that is, integrating different cortical populations that 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\textsuperscript{79,80}. Second, this low-pass neural characteristic is also evident in the temporal modulation spectra of speech signals\textsuperscript{55,56} and speech intelligibility\textsuperscript{48}.

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\textsuperscript{21,22,70–72} but it also provides neural evidence for specific acoustic cues in more natural speech stimuli (that is, sentences). Previous reports have provided evidence for hemispheric asymmetries to specific acoustic cues in non-speech stimuli\textsuperscript{20,61–65} and phonemic segments\textsuperscript{21,22,70–72} and single words\textsuperscript{1,73} but reports using sentence stimuli have been mostly linked to neural signatures rather than specific cues in the sentence stimuli\textsuperscript{74–77}. 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\textsuperscript{8}. The right hemisphere predominantly extracts information from longer temporal windows whereas the left hemisphere extracts information over shorter temporal windows. This approach focuses on temporal sampling rather than time–frequency resolution (although frequency resolution is implicitly greater in the case of long windows), which is consistent with reports of informative theta activity during speech perception\textsuperscript{51,52}. This right hemisphere laterality is evident in both long and short temporal windows whereas the left hemisphere integrates over 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 is primed to laterally transmit longer temporal cycles whereas the right hemisphere is tuned to shorter temporal cycles. While we provide strong evidence for left hemisphere asymmetry during processing of temporal modulations in intelligible speech in our dichotomic behavioural task (Fig. 4), neurophysiology results (Figs. 5 and 6) and processing unintelligible reversed speech (Supplementary Fig. 10), the data supporting asymmetrical processing of spectral modulations is more complex. In the behavioural dichotomic task we found a significant difference for the left hemisphere (Fig. 5), but weaker correlations with spectral modulations in the right hemisphere (Fig. 6). While we provide strong evidence for left hemisphere asymmetry during processing of temporal modulations in intelligible speech, this provides strong support for the hypothesis that the left hemisphere specializes in domain-specific speech mechanisms\textsuperscript{13,25}. Under this view, the right hemisphere is strongly selective for certain acoustic cues, whereas the left hemisphere specializes in domain-specific (speech) mechanisms and is not selective to an acoustic regime\textsuperscript{11}. The strong asymmetry of temporal modulation correlations to the left hemisphere partly challenges this view by providing an acoustic regime (temporal modulations) that 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 suggests, 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.

Another viewpoint could be raised under the lens of lateralization driven by domain-specific speech mechanisms\textsuperscript{8,20}. Under this view, the right hemisphere is strongly selective for certain acoustic cues, whereas the left hemisphere specializes in domain-specific speech mechanisms and is not selective to an acoustic regime\textsuperscript{11}. The strong asymmetry of temporal modulation correlations to the left hemisphere partly challenges this view by providing an acoustic regime (temporal modulations) that 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 suggests, 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\textsuperscript{41,42,81–83}. These studies have reported bilateral activation patterns and modest asymmetries at best, and the modulation topography profiles differ across studies\textsuperscript{41,42,81}. 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\textsuperscript{31} and neuroimaging\textsuperscript{42} have shown that modulation domain cues are critical for more accurate modelling 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, and so on) as well as across species of interest. We provide behavioural 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)84. A set of 28 unique speakers from the database was selected; 14 female and 14 male. 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 estimator85 showing a mean F0 with relatively low female pitch values (mean = 161.34 Hz, SD = 27) and relatively high male pitch (mean = 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 or 8 Hz and once with a low-pass cutoff of 0.1867, 0.3733, 0.56, 0.7467, 0.9333, 1.12 or 1.3067 cycles per octave (28 speakers × 3 sentences × 14 filters). Stimmuli 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 bands86. Gaussian centre frequencies logarithmically spanned the frequency space and the full width at half maximum (FWHM) corresponded to the equivalent rectangular bandwidth87 (bandwidth = 24.7 × (F × 4.37 + 1), where F is the centre frequency in kHz). The output of the filter operation was then Hilbert transformed 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 is essentially a multiplication in the two-dimensional frequency domain of the spectrogram matrix (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 a previous study88). 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. Although such a direct transformation based only on spectrogram power estimates is not feasible, there is a convex projection technique that iteratively produces a waveform that is maximally close to the desired spectrogram89. And although 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 bank90, 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) into a spectrogram that does not fully match the smoothed spectrogram but contains more accurate phase information than the previous iteration (that is, the spectrogram more closely matches the desired spectrogram). The new phase information and 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 (dichotic, n = 20) participants listened to blocks of 84 unique pseudo-randomly ordered sentences filtered either at 2, 3, 4, 5, 6, 7 or 8 Hz low-pass temporal modulation cutoff (and 11 cycles per octave spectral cutoff) or 0.1867, 0.3733, 0.56, 0.7467, 0.9333, 1.1200 or 1.3067 cycles per 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 sentences in the key order of the first filter block and finally select whether the speaker was female or male (keyboard response, two alternative forced choices). 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 the other half were prompted in the reverse order. Participants heard 84 unique sentences spoken by 28 unique speakers (three sentences per speaker). For each participant, three cutoffs were randomly picked (either temporal or spectral) such that each filter cutoff appeared six times different filters (6 × 14 filters).

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 were selected for the first block were repeated but in a different pseudo-random 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 (two blocks filtering intelligibility and two marking male or female) of 120 pseudo-randomly ordered sentences per block (40 unique sentences) constructed from the same filters as in experiment 1, but 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 dichotonic pairs of (2, 3), (3, 4), (4, 5), (5, 6) and (6, 7) Hz, and spectral dichotonic pairs (0.1867, 0.3733), (0.3733, 0.56), (0.56, 0.7467), (0.7564, 0.9333) and (0.9333, 1.1200) cycle per octave. 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 (for example, (2, 2) Hz). Half the sentences were presented first with a higher cutoff on the left (and then right) and half with the opposite. Each participant received a filter with a higher cutoff on the right and then left and diotic. Each filter cutoff dichotic pair was heard in eight 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 1, in order to avoid priming effects of both speaker identity as well as sentence content, a dichotic sentence did not appear within a block and a higher cutoff dichotic filter was never used before a lower cutoff dichotic filter pair 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 the other half of the participants performed the voice pitch blocks first. 40 of the 60 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, that is participants listened to four blocks (two blocks rating intelligibility and two marking voice pitch) of 120 pseudo-randomly ordered sentences per block (40 unique sentences) using the same temporal and spectral cutoffs 3, 4, 5, 6 or 7 Hz 0.1867, 0.3733, 0.56, 0.7467 or 0.9333 cycle per octave. 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 stimuli presented as well as reversed sentences, identical to the pre-processed sentences in experiments 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 or 6 Hz or 0.1867, 0.3733, 0.56, 0.7467 or 0.9333 cycles per octave). 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 experiments 2–4 were 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 pseudo-randomly 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.

No statistical analyses were performed to determine sample sizes, but our sample sizes are similar to previous publications84,18,37–72.
Psychophysical analysis. Intelligibility transcriptions were processed 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 judging 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 s.e.m. in Fig. 4. In order to assess statistically significant effects of filter and block (or ear in the dichotic case), we first applied a within-subject (‘repeated measures’) analysis of variance (ANOVA) and used the raw scores due to the fact that proportion (percent correct) data are susceptible to violation of ANOVA assumptions. Although 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 sphericity (diotic intelligibility $W = 0.9401$, $P = 7.343 \times 10^{-3}$; diotic voice pitch identification $W = 0.0212$, $P = 1.88 \times 10^{-3}$; dichotic intelligibility $W = 0.528$, $P = 3.09 \times 10^{-3}$; dichotic voice pitch identification $W = 0.29$, $P = 0.0078$). In an attempt to resolve the violation of sphericity corrections, we opted to use a non-parametric approach of analysis of variance using a factorial permutation test. We used 1,000 permutations as recommended by the default value of the ez-package$^7$ but also verified that all effects held with 10,000 permutations. ANOVA analysis, sphericity tests and final non-parametric factorial tests were implemented in R using the ez-package$^7$, 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. In cases where we believe that all effects were captured in the non-parametric approach mirrored the results of the ANOVA after correction for sphericity (that is, 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 parameters (and fixed parameters). Non-parametric 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 as previously described. This procedure was repeated 1,000 times, providing a null distribution for assessing significance. When evaluating correlations on the level of sensor (topographic plots in Fig. 5 and Supplementary Figs. 4–6 and 9–11), power estimates were first averaged across time and then correlated with the filter type. When evaluating correlations over time (Fig. 5 middle panel and Supplementary Figs. 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 (1,000 times in order to include only 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 reduce the process for each sensor, power estimates were normalized by the number of current dipoles, and estimates the dipole’s amplitude configuration as well as the neural power estimate across the 240 sentences. Correlations were then correlated with their corresponding filter cutoff, producing one correlation per sentence. These estimates were then correlated with their corresponding filter cutoff, producing one correlation per sentence. 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 performed the same sphericity violations of the MEG data and the non-parametric approach of randomly shuffling labels and producing a null distribution of laterality indices. This distribution was found to be normality so we opted to test significance on the un-normalized indices, that is $(L−R)$ 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 estimate (MNE)$^3$. 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 the computed source space were filtered through a low-pass 4 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 (Supplementary Fig. 7), uncorrected correlations were averaged per anatomical ROI (base on a fieldtrip-provided atlas) to avoid statistical double dipping.
Intracranial data acquisition. Three (two patients with bilateral stereotactic 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 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 stimulus 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 filtered with a range of 5–70 Hz and 70–150 Hz (70–150 Hz line noise, poor contact with abnormal voltage readings). A common electrode at NYU Langone Medical Center 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 oedema-induced shifts (that is, frequency domain FWHM Gaussian filter, a Hilbert transform and log activity was extracted with the same band-pass filtering approach as in MEG procedures) and included visual inspection and rejection of noisy electrodes (Results), their location value and the asymmetry index (Supplementary Figs. 16 and 17) we only report electrodes with a significant correlation via a permutation test, their location value and the asymmetry index (Supplementary Figs. 16 and 17) we only report electrodes with a significant correlation via a permutation test. For analysis of correlations across patients and their asymmetry, we focused on all electrodes (Results), their location value and the asymmetry index (Supplementary Figs. 16 and 17) we only report electrodes with a significant correlation via a permutation test.

Statistical reporting of permutation tests. While there is no consistent format for reporting statistical permutation tests, we chose to follow a previously published approach by employing the nomenclature $P_{\text{corr,mem}} = \text{X}$ 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 or 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.

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

Code availability. All stimulus construction code is available on a public repository https://github.com/flinkelas/SpectroTemporalModulationFilter. Experiment and analysis code is available from the corresponding author upon reasonable request.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 3 May 2018; Accepted: 28 January 2019; Published online: 4 March 2019.

References.

1. Broca, P. Remarques sur le siege de la faculé du langage articulé, suivies d’une observation d’aphémie (perte de la parole). Bull. et mémoires de la Société Anatomique de Paris 36, 330–356 (1861).
2. Wernicke, C. Der aphasisch Symptomkomplex; eine psychologische Studie auf anatominischer Basis (Cohn & Weigert, Breslau, 1874).
3. Zatorre, R. J., Gandour, J. T. Neural specializations for speech and pitch: moving beyond the dichotomies. Trends Cognit. Sci. 24, 403–410 (2020).
4. Zatorre, R. J. Pitch perception of complex tones and human temporal-lobe function. J. Acoust. Soc. Am. 84, 566–572 (1988).
5. Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849 (1992).
6. Belin, P. et al. Lateralization of speech and auditory temporal processing. J. Cogn. Neurosci. 10, 536–540 (1998).
7. Zatorre, R. J., Belin, P. & Zesch, D. The perception of speech: an auditory spatial cue. J. Acoust. Soc. Am. 115, 968–972 (2004).
8. Zatorre, R. J., Gervain, J. Near-infrared spectroscopy: recent advances in infant speech perception and language acquisition research. Front. Psychol. 5, 481 (2014).
9. Zatorre, R. J. & Poeppe, D. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. Front. Psychol. 3, 170 (2012).
10. Okamoto, H. & Kakigi, R. Hemispheric asymmetry of auditory mismatch negativity elicited by spatial and temporal deviants: a magnetoencephalographic study. Brain Topogr. 28, 471–478 (2015).
11. Thompson, E. C. et al. Hemispheric asymmetry of endogenous neural oscillations in young children: implications for hearing speech in noise. Sci. Rep. 6, 19737 (2016).
30. Ratclife, V. F. & Reby D. Orienting asymmetries in dogs’ responses to different communicatory components of human speech. *Carr. Biol.* 24, 2908–2912 (2014).

31. Washington, S. D. & Tillinghast, J. S. Conjugating time and frequency: hemispheric specialization, acoustic uncertainty, and the mustached bat. *Front. Neurosci.* 9, 143 (2015).

32. Washington, S. D. & Kanwal, J. S. Sex-dependent hemispheric asymmetries for processing frequency-modulated sounds in the primary auditory cortex of the mustached bat. *J. Neurophysiol.* 108, 1548–1566 (2012).

33. De Valois, R. L. & De Valois K. K. *Spatial Vision* (Oxford Univ. Press, 1990).

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

35. Kowalski, N., Depireux, D. A. & Shamma, S. A. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J. Neurophysiol.* 76, 3503–3533 (1996).

36. Woolley, S. M. N., Fremouw, T. E., Hsu, A. & Theunissen, E. F. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.* 8, 1371–1379 (2005).

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

38. Elliott, T. M. & Theunissen, F. E. The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5, e1000302 (2009).

39. Arnal, L. H., Finkler, A., Kleinschmidt, A., Giraud, A.-L. & Poeppel, D. Human sound occupation: a privileged niche in the communication soundscape. *Carr. Biol.* 25, 2051–2056 (2015).

40. Gross, J. et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11, e1001752 (2013).

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

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).

43. Pasley, B. N. et al. Reconstructing speech from human auditory cortex. *PLoS Biol.* 10, e1001251 (2012).

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

45. Moore, B. C. I. Auditory filter shapes derived in simultaneous and forward masking. *J. Acoust. Soc. Am.* 70, 1003 (1986).

46. Ghalager, R. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *J. Acoust. Soc. Am.* 110, 1628–1640 (2001).

47. Druhl, R., Festen, J. M. & Plomp, R. Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* 95, 2670–2680 (1994).

48. Chi, T., Gao, Y., Guyton, M. C., Ru, P. & Shamma, S. Spectro-temporal modulation transfer functions and speech intelligibility. *J. Acoust. Soc. Am.* 106, 2719 (1999).

49. Jaeger, T. F. Categorial data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446 (2008).

50. Lawrence, M. A. The Package ez. *http://github.com/mike-lawrence/ez (R package, 2016).*

51. Prins, N. & Kingdom, F. A. A. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox. *Front. Psychol.* 9, 1250 (2018).

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 Cognit. Sci.* 5, 340–348 (2001).

54. Venezia, J. H., Hickok, G. & Richards, V. M. Auditory ‘bubbles’: efficient classification of the spectrotemporal modulations essential for speech intelligibility. *Acoust. Soc. Am.* 107, 1072–1088 (2005).

55. Houtgast, T. & Steeneken, H. J. M. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* 77, 1069–1077 (1985).

56. Ding, N. et al. Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 31, 187–201 (2007).

57. Shannon, R. V. Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. Speech recognition with primarily temporal cues. *IEEE Trans. Audio. Speech. Lang. Process.* 14, 1127–1134 (2006).

58. Gabor, D. Theory of communication. Part I: the analysis of information. *J. Inst. Electr. Eng.* 93, 429–441 (1946).

59. Joris, P. X., Schreiner, C. E. & Ress, A. Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577 (2004).

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

61. Zatorre, R. J. & Belin, P. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953 (2001).
Acknowledgements
This work was supported by NIH F32 DC011985 and Charles H. Revson Senior Fellowships in Biomedical Science 15–28 to A.F., by NIH 2R01DC05660 to D.P. and by NIMH R21 MH114166-01 to A.D.M. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We would like to thank I. T. Kim and N. Mei, who assisted in the setup and acquisition of psychophysical dichotic data, B. Mahmood and M. Hofstrader, who assisted in NYU ECoG data acquisition and setup, D. Groppe, who assisted in North Shore ECoG data acquisition and electrode reconstruction, and H. Wang, who provided electrode reconstruction at NYU.

Author contributions
A.F. and D.P. designed the study and hypotheses, and wrote the manuscript. A.F. constructed the stimuli and filtering techniques and collected and analysed the data. A.D.M., O.D. and W.K.D. recruited clinical patients and performed clinical care.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0548-z.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.F.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature Limited 2019
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Behavioral data was collected using custom Matlab scripts based on Psychtoolbox. Concurrent neural data was collected using a commercial Yokogawa MEG system or commercial clinical EEG systems used by the Epilepsy centers (Natus Medical Incorporated).

Data analysis
All stimuli was constructed and data analyzed using custom Matlab and R scripts written by the first author. Source reconstruction of MEG data was based on scripts from the Fieldtrip toolbox.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences
**Life sciences study design**

All studies must disclose on these points even when the disclosure is negative.

| Sample size     | Sample sizes were taken in accordance to the standard published and acceptable size in the field of research. |
|-----------------|---------------------------------------------------------------------------------------------------------|
| Data exclusions | Data exclusions occurred only if the participant did not perform the task or data quality was so low it could not be analyzed (due to artifacts), such cases are detailed in the Methods. |
| Replication     | Our findings replicated across three modalities (psychophysics, MEG, ECoG). In the case of MEG we provide two datasets replicating the main effect of temporal correlations but not spectral correlations due to task differences. We have also replicated the main MEG experiment in a sub-population (6 participants) with a separate MEG system providing the same results (unpublished). |
| Randomization   | There was no randomization according to group or control groups. All statistics are within subjects or across subjects but not comparing one group to another. 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). |
| Blinding        | Data collection and analysis were not performed blind to the conditions of the experiments. Blinding was not relevant to the study as there was no comparison of treatment vs. control group. |

**Reporting for specific materials, systems and methods**

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

**Materials & experimental systems**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Antibodies            |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology         |
| ☒   | Animals and other organisms |
| ✗   | Human research participants |
| ✗   | Clinical data         |

**Methods**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChIP-seq              |
| ☒   | Flow cytometry        |
| ✗   | MRI-based neuroimaging |

**Human research participants**

Policy information about studies involving human research participants

**Population characteristics**

For MEG and behavioral paradigms subjects were recruited from the NYU community and characteristics represent the overall population in the New York metropolitan area as well as New York University. Subjects with a history of neurological conditions or any metal implants were not included in the study. For the ECoG experiment, patient characteristics represent the overall patient cohort at NYU Langone Medical Center and North Shore University Hospital with refractory Epilepsy.

**Recruitment**

For MEG and behavioral paradigms subjects were recruited using fliers in the University, students in the Psychology department and previous subjects in MEG experiments at NYU. For the ECoG experiment, patients were recruited if they were undergoing neurosurgery for refractory epilepsy for clinical purposes and volunteered to participate in research during their planned hospital stay. In both cases there can be a self-selection bias to the healthy population at NYU and the epileptic population that would undergo neurosurgical treatment.

**Ethics oversight**

The study protocol was approved by the NYU Committee on Human Research for all MEG and behavioral data collection and analysis. In a separate protocol, NYU approved the analysis of anonymized ECoG data acquired at the hospital. ECoG data collection was approved by the NYU Langone Medical Center Committee on Human Research for NYU Langone ECoG patients. Stereotactic and ECoG data collection was approved by the North Shore University Hospital Committee on Human Research for North Shore University Hospital patients.

Note that full information on the approval of the study protocol must also be provided in the manuscript.