eLife

Reviewed Preprint
v2 • June 5, 2024
Revised by authors

Reviewed Preprint
v1 • February 8, 2024

Neuroscience

Speech and music recruit frequency-specific distributed and overlapping cortical networks

Noémie te Rietmolen, Manuel Mercier, Agnès Trébuchon, Benjamin Morillon, Daniele Schön

Institute for Language, Communication, and the Brain, Aix-Marseille Université, Marseille, France • Aix Marseille Université, INSERM, INS, Institut de Neurosciences des Systèmes, Marseille, France • APHM, Hôpital de la Timone, Service de Neurophysiologie Clinique, Marseille, France

Abstract

To what extent does speech and music processing rely on domain-specific and domain-general neural networks? Using whole-brain intracranial EEG recordings in 18 epilepsy patients listening to natural, continuous speech or music, we investigated the presence of frequency-specific and network-level brain activity. We combined it with a statistical approach in which a clear operational distinction is made between *shared*, *preferred,* and domain-*selective* neural responses. We show that the majority of focal and network-level neural activity is shared between speech and music processing. Our data also reveal an absence of anatomical regional selectivity. Instead, domain-selective neural responses are restricted to distributed and frequency-specific coherent oscillations, typical of spectral fingerprints. Our work highlights the importance of considering natural stimuli and brain dynamics in their full complexity to map cognitive and brain functions.

eLife assessment

This study presents **valuable** intracranial findings on how two types of natural auditory stimuli - speech and music - are processed in the human brain, and demonstrates that speech and music largely share network-level brain activities, thus challenging the domain-specific processing view. The evidence supporting the claims of the authors is **solid**. The work will be of broad interest to speech and music researchers as well as cognitive scientists in general.

https://doi.org/10.7554/eLife.94509.2.sa3

Introduction

The advent of neuroscience continues the longstanding debate on the origins of music and language—that fascinated Rousseau and Darwin (Kivy, 1959; Rousseau, 2009)—on new biological ground: evidence for the existence of selective and/or shared neural populations involved in their processing. The question on functional selectivity versus domain-general mechanisms is closely related to the question of the nature of the neural code: Are representations

sparse (and localized) or *distributed*? While the former allows to explicitly represent any stimulus in a small number of neurons, it would require an intractable number of neurons to represent all possible stimuli. Experimental evidence instead suggests that stimulus identification is achieved through a population code, implemented by neural coupling in a distributed dynamical system (Bizley & Cohen, 2013; Rissman & Wagner, 2012). The question of the nature of the neural code has tremendous implications: it defines an epistemological posture on how to map cognitive and brain functions. This, in turn, affects both the definition of cognitive operations – what is actually computed – as well as the way we look at the data – looking for differences or similarities.

Neuroimaging studies report mixed evidence of selectivity and resource sharing. On one hand, one can find claims for a clear distinction between brain regions exclusively dedicated to language versus other cognitive processes (Chen et al., 2023; Fedorenko et al., 2011; Fedorenko & Blank, 2020; Friederici, 2020) and for the existence of specific and separate neural populations for speech, music, and song (Boebinger et al., 2021; Norman-Haignere et al., 2022). On the other hand, other neuroimaging studies suggest that the brain regions that support language and speech also support nonlinguistic functions (Albouy et al., 2020; Fadiga et al., 2009; Koelsch, 2011; Menon et al., 2002; Robert et al., 2023; Schön et al., 2010). This point is often put forward when interpreting the positive impact music training can have on different levels of speech and language processing (Flaugnacco et al., 2015; François et al., 2013; Kraus & Chandrasekaran, 2010; Schön et al., 2004).

Several elements may account for these different findings. The very first may rely on the definition of a brain region. This can be considered as a set of functionally homogeneous but spatially distributed voxels, or, alternatively, as an anatomical landmark as those used in brain atlases (e.g. inferior frontal gyrus). However, observing functional regional selectivity in a distributed pattern is not incompatible with the observation of an absence of anatomical regional selectivity: a selective set of voxels may exist within an anatomically non-selective region. A second element concerns the choice of the stimuli. Some of the studies claiming functional selectivity used rather short auditory stimuli (Boebinger et al., 2021; Norman-Haignere et al., 2015; Norman-Haignere et al., 2022). Besides the low ecological validity of such stimuli that may reduce the generalizability of the findings (Theunissen et al., 2000), their comparison further relies on the assumption that speech and music share similar cognitive time constants. However, speech unfolds faster than music (Ding et al., 2017), and while a linguistic phrase is typically shorter than a second (Inbar et al., 2020), a melodic phrase is an order of magnitude longer. Moreover, balancing the complexity/simplicity of linguistic and musical stimuli can be challenging, and musical stimuli are often reduced to very simple melodies played on a synthesizer. These simple melodies mainly induce pitch processing in associative auditory regions (Griffiths et al., 2010) but do not recruit the entire dual-stream auditory pathways (Zatorre et al., 2007). Overall, while short and simple stimuli may be sufficient to induce linguistic processing, they might not be cognitively relevant musical stimuli. Finally, another element concerns the data at stake. Most studies that compared language and music processing, examined functional MRI data (Chen et al., 2023; Fedorenko et al., 2011; Nieto-Castañón & Fedorenko, 2012). Here, we would like to consider cognition as resulting from interactions among functionally specialized but widely distributed brain networks and adopt an approach in which large-scale and frequency-specific neural dynamics are characterized. This approach rests on the idea that the canonical computations that underlie cognition and behavior are anchored in population dynamics of interacting functional modules (Buzsáki & Vöröslakos, 2023; Safaie et al., 2023) and bound to spectral fingerprints consisting of network– and frequency-specific coherent oscillations (Siegel et al., 2012). This framework requires relying on time-resolved neurophysiological recordings (M/EEG) and—rather than focusing only on the amplitude of the high-frequency activity, a common approach in the literature involving human intracranial EEG recordings (Martin et al., 2019; Norman-Haignere et al., 2022; Oganian & Chang, 2019)—to investigate the entire

frequency spectrum of neural activity. Indeed, while HFa amplitude is a good proxy of focal neural spiking (Le Van Quyen et al., 2010; Ray & Maunsell, 2011), large-scale neuronal interactions mainly rely on slower dynamics (Kayser et al., 2012; Kopell et al., 2000; Siegel et al., 2012).

Following the reasoning developed above, we suggest that the study of selectivity of music and language processing should carefully consider the following points: First, the use of ecologically valid stimuli, both in terms of content and duration. Second, a within-subject approach comparing both conditions. Third, aiming for high spatial sensitivity. Fourth, considering not only one type of neural activity (broadband, HFa amplitude) but the entire frequency spectrum of the neurophysiological signal. Fifth, use a broad range of complementary analyses, including connectivity, and take into account individual variability. Finally, we suggest that terms should be operationally defined based on statistical tests, which results in a clear distinction between shared, selective, and preferred activity. That is, be *A* and *B* two investigated cognitive functions, "shared" would be a neural population that (compared to a baseline) significantly and equally contributes to the processing of both *A* and *B*; "selective" would be a neural population that exclusively contributes to the processing of *A* or B (e.g. significant for A but not *B)*; and "preferred" would be a neural population that significantly contributes to the processing of both *A* and *B*, but more prominently for *A* or *B* (**Figure 1A**).

In an effort to take into account all the above challenges and to precisely quantify the degree of shared, preferred, and selective responses both at the levels of the channels and anatomical regions (**Figure 1C-D**), we conducted an experiment on 18 pharmacoresistant epileptic patients explored with stereotactic EEG (sEEG) electrodes. Patients listened to long and ecological audio-recordings of speech and music (10-minutes each). We investigated stimulus encoding, spectral content of the neural activity, and brain connectivity over the entire frequency spectrum (from 1-120 Hz; i.e. delta band to HFa). Finally, we carefully distinguished between the three different categories of neural responses described above: shared, selective, and preferred across the two investigated cognitive domains. Our results reveal that the majority of neural responses are shared between natural speech and music, and they highlight an absence of anatomical regional selectivity. Instead, we found neural selectivity to be restricted to distributed and frequency-specific coherent oscillations, typical of spectral fingerprints.

Results

Anatomical regional neural activity is mostly non-domain selective to speech or music

To investigate the presence of domain-selectivity during ecological perception of speech and music, we first analyzed the neural responses to these two cognitive domains in both a spatially and spectrally resolved manner, with respect to two baseline conditions: one in which patients passively listened to pure tones (each 30 ms in duration), the other in which they passively listened to isolated syllables (/ba/ or /pa/, see Methods). Here we will report the results using pure tones data as baseline, but note that the results using syllables data as baseline are highly similar (see Figures S1-5). We classified, for each canonical frequency band, each channel into one of the categories mentioned above, i.e. shared, selective, or preferred (**Figure 1A**), by examining whether speech and/or music differ from baseline and whether they differ from each other. We also considered both activations and deactivations, compared to baseline, as both index a modulation of neural population activity, and both have been linked with cognitive processes (Pfurtscheller & Lopes da Silva, 1999; Proix et al., 2022). However, because our aim was not to interpret specific increase or decrease with respect to the baseline, we here simply consider significant deviations from the baseline. In other words, when estimating selectivity, it is the strength of the response that matters, not its direction (activation, deactivation). Overall, neural responses are predominantly shared between the two domains, accounting for ~70% of the

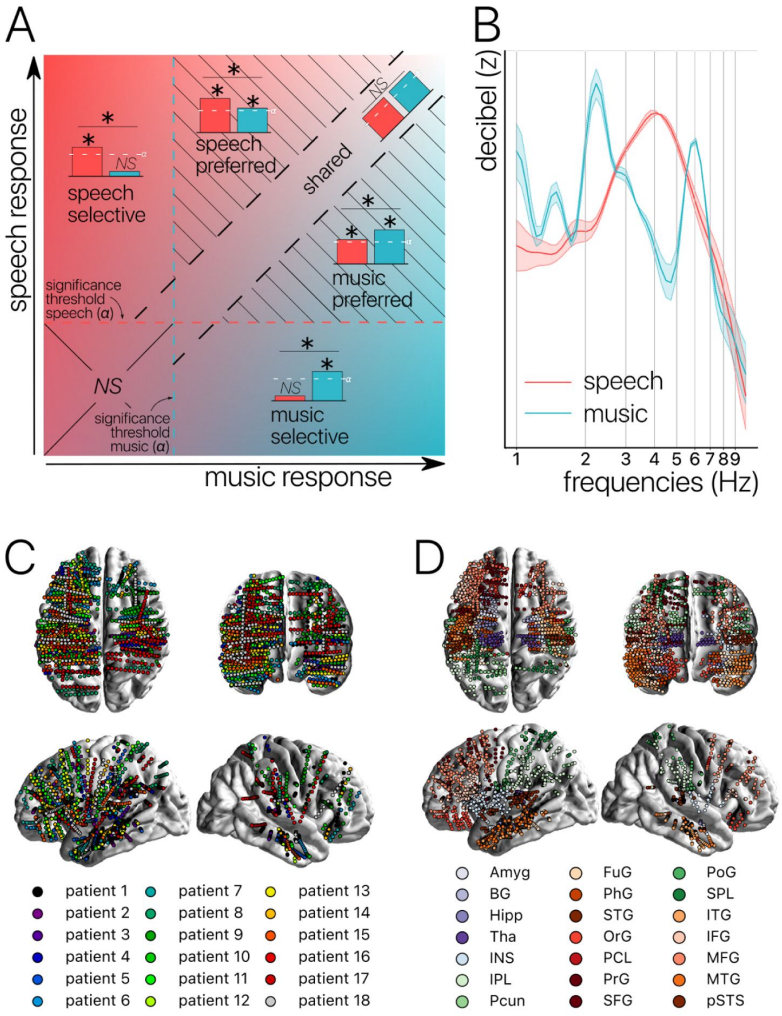


Figure 1.

Concepts, Stimuli and Recordings.

(A) Conceptual definition of selective, preferred and shared neural processes. Illustration of the continua between speech and music selectivity, speech and music preference and shared resources. “Selective” responses are neural responses significant for one domain but not the other, and with a significant difference between domains (for speech top-left; for music bottom-right). “Preferred” responses correspond to neural responses that occur during both speech and music processing, but with a significantly stronger response for one domain over the other (striped triangles). Finally, “shared” responses occur when there are no significant differences between domains, and there is a significant neural response to at least one of the two stimuli (visible along the diagonal). If neither domain produces a significant neural response, the difference is not assessed (lower left square). **(B) Stimuli.** Modulation spectrum of the acoustic temporal envelope of the continuous, 10-minutes long speech and music stimuli. **(C) Anatomical localization of the sEEG electrodes for each patient** (N=18). **(D) Anatomical localization of the sEEG electrodes for each anatomical region**. Abbreviations according to the Brainnetome Atlas (Fan et al., 2016).

channels which showed a significant response compared to baseline (**Figures 2**-**3**). The preferred category is also systematically present, accounting for 3 to 15% of significant neural responses, across frequency bands. Selective responses are more present in the lower frequency bands (~30% up to the alpha band), and quite marginal in the HFa band (6-12%).

The spatial distribution of the spectrally-resolved responses corresponds to the network typically involved in speech and music perception. This network encompasses both ventral and dorsal auditory pathways, extending well beyond the auditory cortex and hence beyond auditory processing that may result from differences in the acoustic properties of our baseline and experimental stimuli. This is the case for overall responses but also when only looking at shared responses. For instance, HFa shared responses represent 74-86% of the overall significant HFa responses, and are visible in the left superior and middle temporal gyri, inferior parietal lobule, and the precentral, middle and inferior frontal gyri (**Figures 2F** & 3F). The left hemisphere appears to be more strongly involved, but this result is biased by the inclusion of a majority of patients with a left hemisphere exploration (**Figure 1C-D** and Table S1). Also, when inspecting left and right hemispheres separately, the patterns of shared, selective, and preferred responses remain similar across hemispheres across frequency bands (see Figures S6-7 for activation and deactivation, respectively). Both domains displayed a comparable percentage of selective responses across frequency bands (**Figure 4**, first values of each plot). When considering separately activation (**Figure 2**) and deactivation (**Figure 3**) responses, speech and music showed complementary patterns: for low frequencies (<15 Hz) speech selective (and preferred) responses were mostly deactivations and music responses activations compared to baseline, and this pattern reversed for high frequencies (>15 Hz).

Next, we investigated whether the channels selectivity (to speech or music) observed in a given frequency band was robust across frequency bands (**Figure 4**). We estimated the cross-frequency channel selectivity, that is the percentage of channels that selectively respond to speech or music across different frequency bands. We first computed the percentage of total channels selective for speech and music (either activated or deactivated compared to baseline) in a given frequency band. We then verified whether these channels were unresponsive to the other domain in the other frequency bands. This was done by examining each frequency band in turn and deducting any channels that showed a significant neural response to the other domain. When considering the entire frequency spectrum, the percentage of total channels being selective to speech or music is ~4 times less than when considering a single frequency band. For instance, while up to 8% of the total channels are selective for speech (or music) in the theta band, this percentage always drops to ~2% when considering the cross-frequency channel selectivity.

Critically, we found no evidence of anatomical regional selectivity, i.e. of a simple anatomo-functional spatial code (see **Figure 1D** for the definition of anatomical regions). We estimated, for each frequency band, activation/deactivation responses, and anatomical region, the proportion of patients showing selectivity for speech or music, by means of a population prevalence analysis (**Figures 5**-**6**; see Methods). This analysis revealed that, for the majority of patients, first of all, in most regions there were channels that responded to both speech and music (indicative of shared responses at the anatomical regional level), and, second of all, for the minority of anatomical regions for which a selectivity for the same domain (speech or music) was observed across multiple patients, this selectivity does not hold when also considering other frequency bands and activation/deactivation responses. For instance, while the left anterior middle temporal gyrus shows delta activity selective to music (**Figure 2A** and **5A**), it shows low-gamma activity selective to speech (**Figure 2E** and **5E**). The left STG and pSTS, which show selective activations in the theta and alpha bands for music (**Figure 5B-C**), show selective deactivations in the same bands for speech (**Figure 6B-C**) and a majority of shared activations in the HFa (**Figure 5F**). This absence of anatomical regional selectivity is also evident when looking at the uncategorized, continuous results (Figure S9).

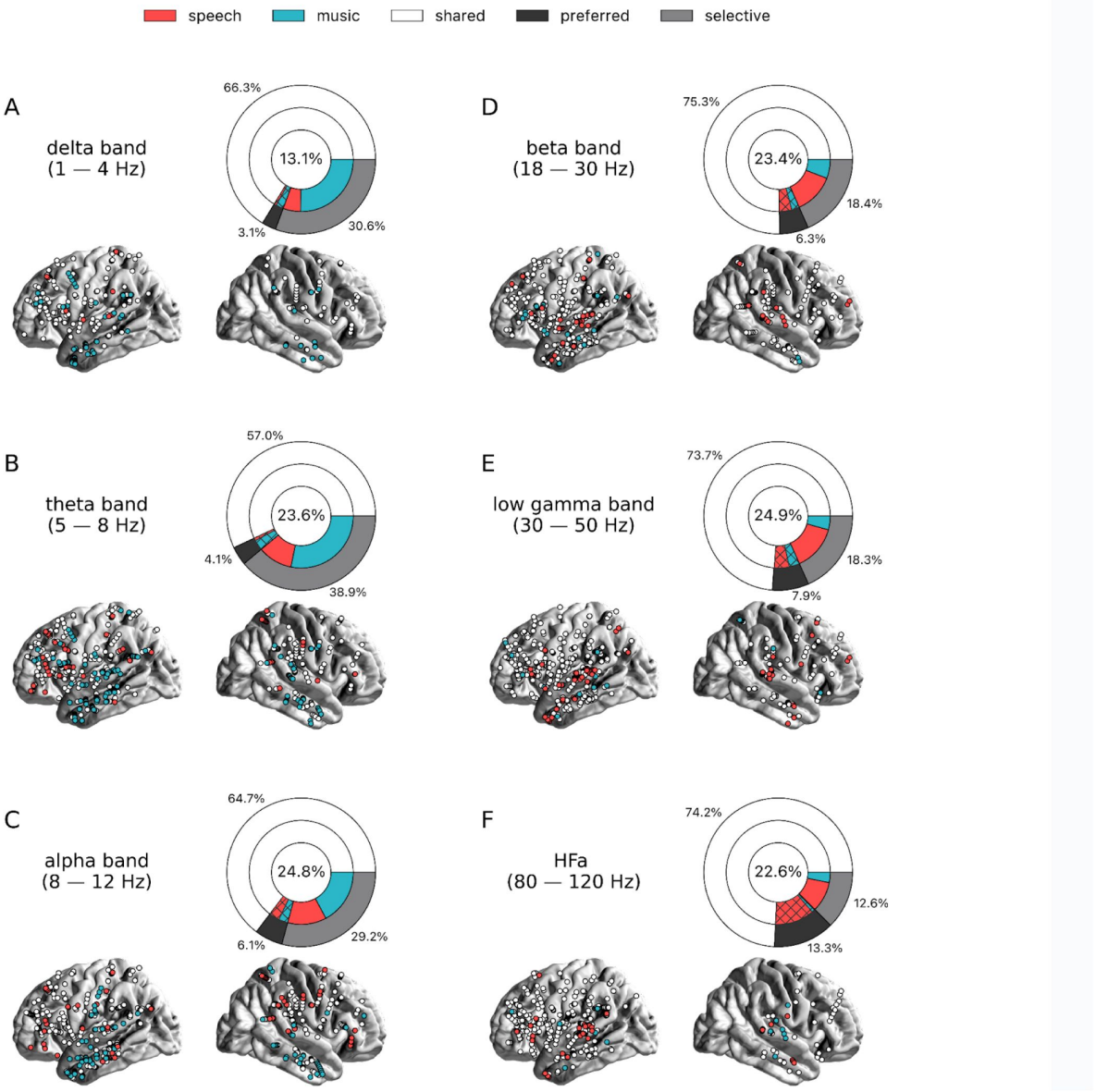


Figure 2.

Power spectrum analyses of activations (speech or music > tones).

(A)-(F) Neural responses to speech and/or music for the 6 canonical frequency bands. Only significant activations compared to the baseline condition (pure tones listening) are reported (see Figure S9 for uncategorized, continuous results). **Nested pie charts indicate**: (1) in the center, the percentage of channels that showed a significant response to speech and/or music. (2) The outer pie indicates the percentage of channels, relative to the center, classified as shared (white), selective (light gray), and preferred (dark gray). (3) The inner pie indicates, for the selective (plain) and preferred (pattern) categories, the proportion of channels that were (more) responsive to speech (red) or music (blue). **Brain plots indicate**: Distribution of shared (white) and selective (red/blue) sEEG channels projected on the brain surface. Results are significant at $q < 0.01$ (N=18).

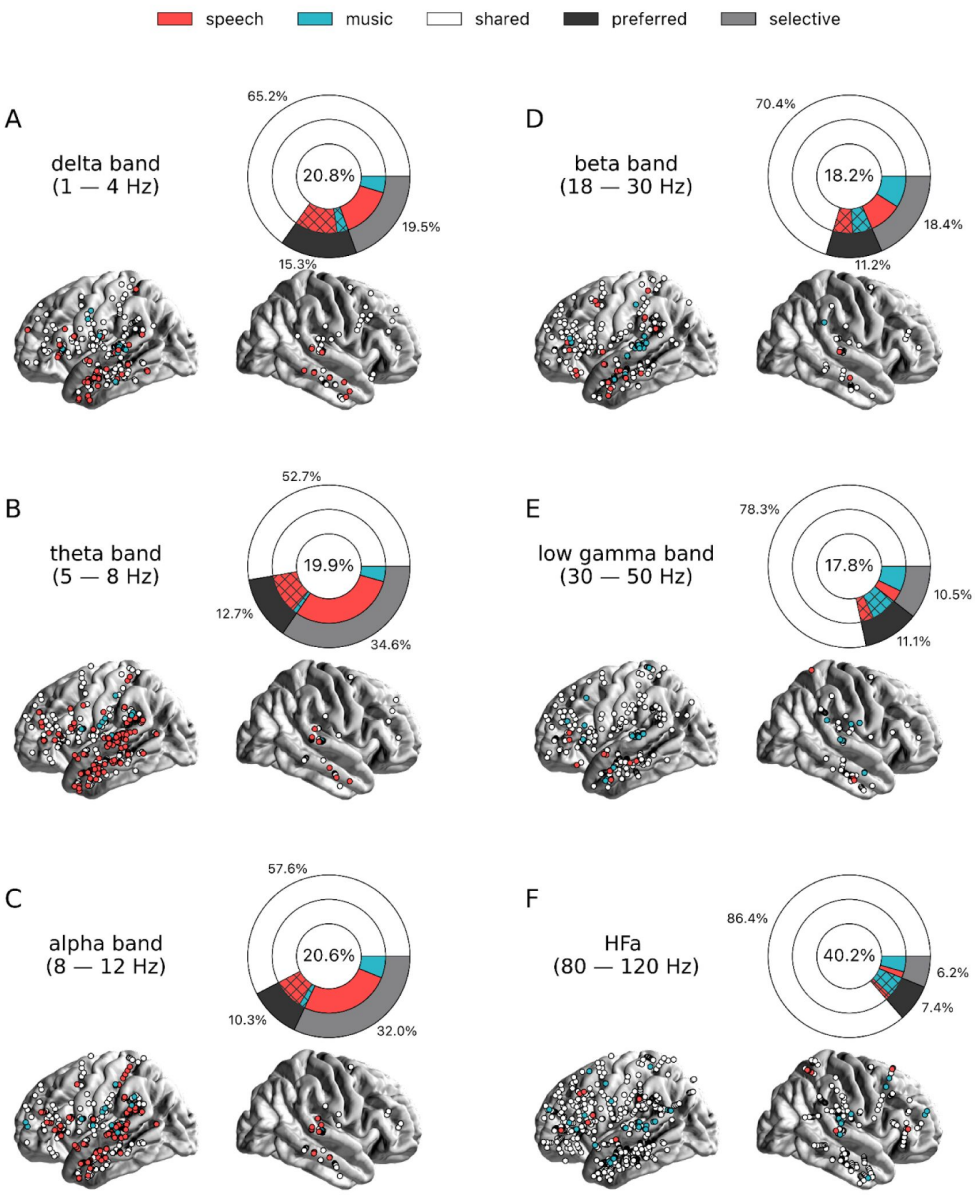


Figure 3.

Power spectrum analyses of deactivations (speech or music < tones).

(A)-(F) Neural responses to speech and/or music for the 6 canonical frequency bands. Only significant deactivations compared to the baseline condition (pure tones listening) are reported. Same conventions as in **Figure 2**. Results are significant at q < 0.01 (N=18).

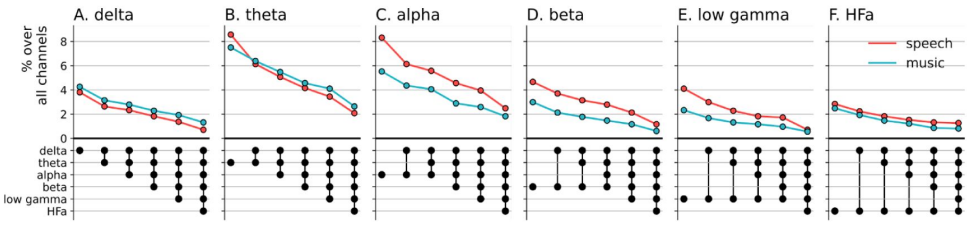


Figure 4.

Cross-frequency channel selectivity for the power spectrum analyses.

Percentage of channels that exclusively respond to speech (red) or music (blue) across different frequency bands. For each plot, the first (leftmost) value corresponds to the percentage (%) of channels displaying a selective response in a specific frequency band (either activation or deactivation, compared to the baseline condition of pure tones listening). In the next value, we remove the channels that are significantly responsive in the other domain (i.e. no longer exclusive) for the following frequency band (e.g. in panel A: speech selective in delta; speech selective in delta XOR music responsive in theta; speech selective in delta XOR music responsive in theta XOR music responsive in alpha; and so forth). The black dots at the bottom of the graph indicate which frequency bands were successively included in the analysis. Note that channels remaining selective across frequency bands did not necessarily respond selectively in every band. They simply never showed a significant response to the other domain in the other bands.

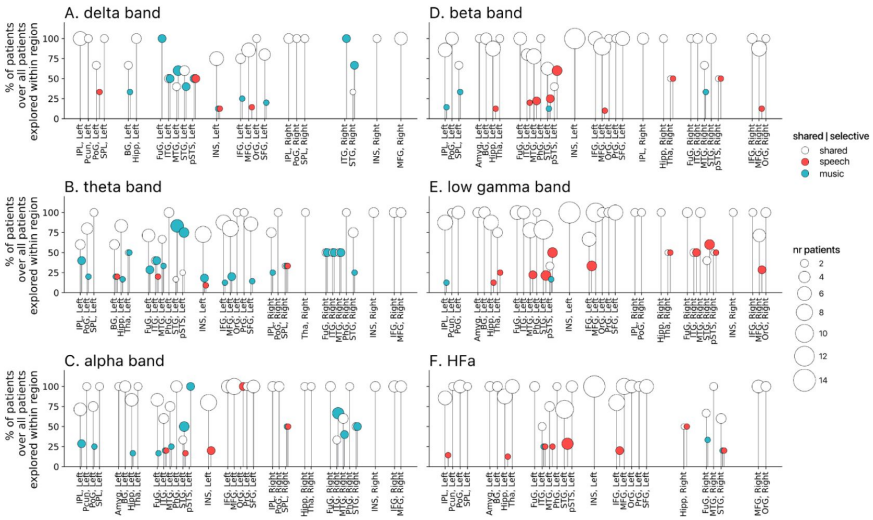


Figure 5.

Population prevalence for the power spectral analyses of activations (speech or music > tones; N = 18).

(A)-(F) Population prevalence of shared or selective responses for the 6 canonical frequency bands, per anatomical region (note that preferred responses are excluded). Only significant activations compared to the baseline condition (pure tones listening) are reported. Regions (on the x-axis) were included in the analyses if they had been explored by minimally 2 patients with minimally 2 significant channels. Patients were considered to show regional selective processing when all their channels in a given region responded selectively to either speech (red) or music (blue). When regions contained a combination of channels with speech selective, music selective or shared responses, the patient was considered to show shared (white) processing in this region. The height of the lollipop (y-axis) indicates the percentage of patients over the total number of explored patients in that given region. The size of the lollipop indicates the number of patients. As an example, in panel F (HFa band), most lollipops are white with a height of 100%, indicating that, in these regions, all patients presented a shared response profile. However, in the left inferior parietal lobule (IPL, Left) one patient (out of the 7 explored) shows speech selective processing (filled red circle). A fully selective region would thus show a fixed-color full height across all frequency bands. Abbreviations according to the Brainnetome Atlas (Fan et al., 2016).

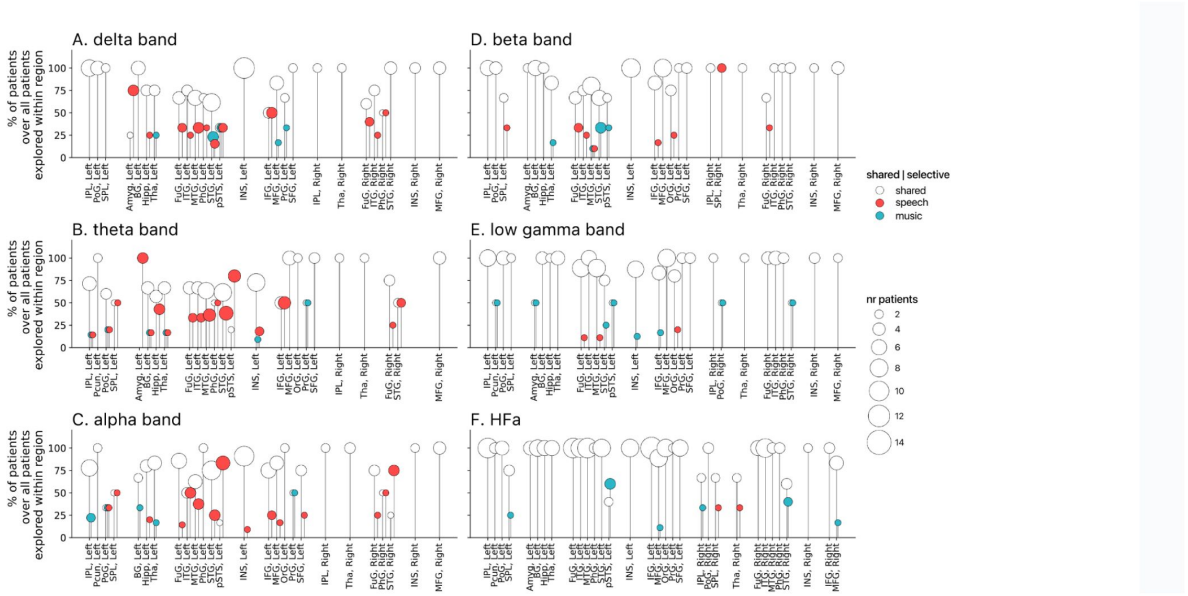


Figure 6.

Population prevalence for the power spectral analyses of deactivations (speech or music < tones; N = 18).

(A)-(F) Population prevalence of shared or selective responses for the 6 canonical frequency bands, per anatomical region. Only significant deactivations compared to the baseline condition (pure tones listening) are reported. Same conventions as in **Figure 5**.

Overall, these results reveal an absence of regional selectivity to speech or music under ecological conditions. Instead, selective responses coexist in space across different frequency bands. But, while selectivity may not be striking at the level of anatomical regional activity, it may still be present at the network level. To investigate this hypothesis, we explored the connectivity between the auditory cortex and the rest of the brain. And, to functionally define the auditory cortex for each patient, we first investigated the relation between the auditory signal itself and the brain response to identify which sEEG channels (spatial) best encode the dynamics of the auditory stimuli.

Low-frequency neural activity best encodes acoustic dynamics

We linearly modeled the neurophysiological responses to continuous speech and music using temporal response functions (TRF). Based on previous studies (Oganian & Chang, 2019; Zion Golumbic et al., 2013; Zuk et al., 2021), we compared four TRF models. From both stimuli, we extracted the continuous, broadband temporal envelope (henceforth 'envelope') and the discrete acoustic onset edges (henceforth 'peakRate'; see Methods) and we quantified how well these two acoustic features are encoded by either the low frequency band (LF, 1-9 Hz) or the high frequency amplitude (80-120 Hz) bands. For each model, we estimated the percentage of total channels for which a significant encoding was observed during speech and/or music listening. The model for which most channels significantly encoded speech and/or music acoustic features corresponded to the model in which LF neural activity encoded the *peakRates* (**Figure 7A**). In general, the LF activity encodes the acoustic features in significantly more channels than the HFa amplitude (*peakRate* & LF vs. & HFa amplitude comparison: t = 13.39, q < .0001; & LF vs. *envelope* & HFa amplitude comparison: t = 9.55, q < .0001). Note that this effect is not caused by the asymmetric comparison of bandpassed LF to HFa amplitude as model comparisons using the same extraction technique for both signals did not change the results (Figure S8). Then, while the are encoded by numerically more channels than the instantaneous envelope, this difference was not significant (*peakRate* & LF vs. & LF comparison: t = 1.93, q = .42).

Furthermore, we show that the *peakRates* are encoded by the LF neural activity throughout the cortex, for both speech and music (**Figure 7B-C**). More precisely, the regions wherein neural activity significantly encodes the acoustic structure of the stimuli go well beyond auditory regions and extend to the temporo-parietal junction, motor cortex, inferior frontal gyrus, and anterior and central sections of the superior and middle temporal gyrus. In particular, the strongest encoding values for speech are observed in the typical left-hemispheric language network, comprising the upper bank of the superior temporal gyrus, the posterior part of the inferior frontal gyrus, and the premotor cortex (Malik-Moraleda et al., 2022). Still, as expected, the best cortical tracking of the acoustic structure takes place in the auditory cortex, for both speech and music (**Figure 7D**). In other words, the best encoding channels are the same for speech and music and are those located closest to—or in—the primary auditory cortex. While the left hemisphere appears to be more strongly involved, this result is biased by the inclusion of a majority of patients with a left hemisphere exploration (see **Figure 1C-D** and Table S1). Proportionally, we found no difference in the number of significant channels between hemispheres (i.e. speech: 41% and 44% for left and right hemispheres respectively; music: 22% and 24% for left and right hemispheres respectively). Finally, the *peakRate* & LF model, i.e. the model that captures the largest proportion of significant channels during speech and/or music perception (**Figure 7A**), yields for both classes of stimuli a similar TRF shape (**Figure 7E**) as well as similar prediction accuracy scores (Pearson's r), of up to 0.55 (**Figure 7F**).

Connections of the auditory cortex are also mostly non-domain selective to speech or music

Seed-based connectivity analyses first revealed that, during speech or music perception, the auditory cortex is mostly connected to the rest of the brain through slow neural dynamics, with ~33% of the channels showing coherence values higher than the surrogate distribution at delta

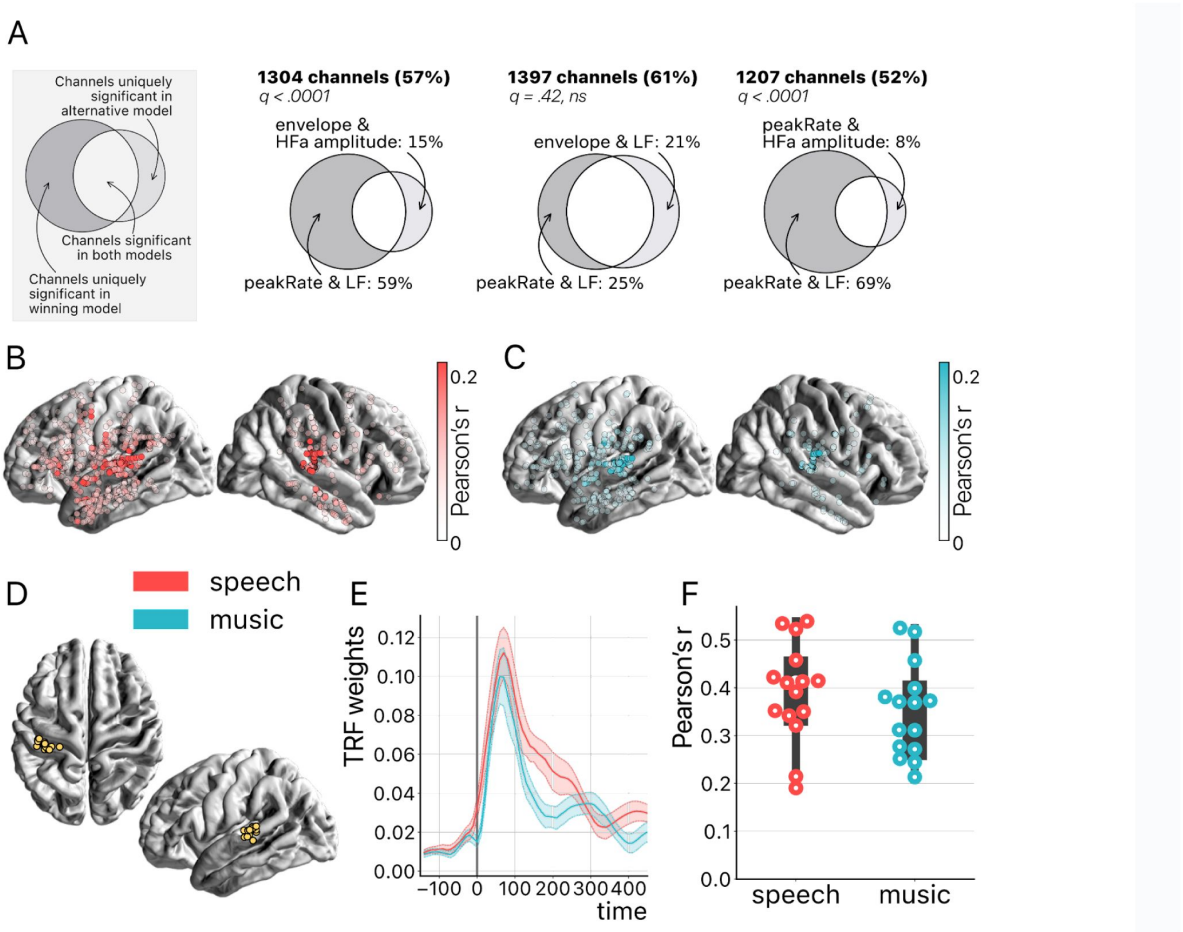


Figure 7.

Temporal Response Function (TRF) analyses.

(A) Model comparison. On the top left a toy model illustrates the use of Venn diagrams **comparing the winning model** (peakRate in LF) to each of the three other models for speech and music (pooled). Four TRF models were investigated to quantify the encoding of the instantaneous envelope and the discrete acoustic onset edges (peakRate) by either the low frequency band (LF) or the high frequency amplitude. The 'peakRate & LF' model significantly captures the largest proportion of channels, and is, therefore, considered the winning model. The percentages on the top (in bold) indicate the percentage of total channels for which a significant encoding was observed during speech and/or music listening in either of the two compared models. In the Venn diagram, we indicate, out of all significant channels, the percentage that responded in the winning model (left) or in the alternative model (right). The middle part indicates the percentage of channels shared between the winning and the alternative model (percentage not shown). Q-values indicate pairwise model comparisons (Wilcoxon sign rank test, FDR-corrected). **(B)** and **(C) peakRate & LF model**: Spatial distribution of sEEG channels wherein low frequency (LF) neural activity significantly encodes the speech (red) and music (blue) peakRates. Higher prediction accuracy (Pearson's r) is indicated by increasing color saturation. All results are significant at $q < 0.01$ (N=18). **(D) Anatomical localization of the best encoding channel within the left hemisphere for each patient** (N=15), as estimated by the 'peakRate and LF' model (averaged across speech and music). These channels are all within the auditory cortex and serve as seeds for subsequent connectivity analyses. **(E) TRFs averaged across the seed channels** (N=15), for speech and music. **(F) Prediction accuracy** (Pearson's r) of the neural activity of each seed channel, for speech and music.

rate, and only ~12% at HFa (**Figure 8**, see also Figure S10 for uncategorized, continuous results). Across frequencies, most of the significant connections are shared between the two cognitive domains (~70%), followed by preferred (~15%) and selective connections (~12%). Selectivity is nonetheless homogeneously present in all frequency bands (**Figure 8**). Importantly, selectivity is again frequency-specific (**Figure 9**). Estimating the cross-frequency channel selectivity, the percentage of total connections being selective to speech or music is at zero for all frequency bands except for the delta range (speech = 0.19%; music = 0.06%). Hence, selectivity is only visible at the level of frequency-specific distributed networks. Finally, here again no anatomical regional selectivity is observed, i.e. not a single cortical region is solely selective to speech or music. Rather, in every cortical region, the majority of patients show shared responses at the regional level, as estimated by the population prevalence analysis (**Figure 10**).

Discussion

In this study, we investigated the existence of domain-selectivity for speech and music under ecological conditions. We capitalized on the high spatiotemporal sensitivity of human stereotactic recordings (sEEG) to thoroughly evaluate the presence of selective neural responses—estimated both at the level of individual sEEG channels and anatomical cortical regions—when patients listened to a story or to instrumental music. More precisely, we statistically quantified the extent to which natural speech and music processing is performed by shared, preferred, or domain-selective neural populations. By combining sEEG investigations of high-frequency activity (HFa) with the analyses of other frequency bands (from delta to low-gamma), the neural encoding of acoustic dynamics and spectrally-resolved connectivity analyses, we obtained a thorough characterization of the neural dynamics at play during natural and continuous speech and music perception. Our results show that speech and music mostly rely on shared neural resources. Further, while selective responses seem absent at the level of atlas-based cortical regions, selectivity can be observed at the level of frequency-specific distributed networks in both power and connectivity analyses.

Previous work has reported that written or spoken language selectively activates a left-lateralized functional cortical network (Chen et al., 2023; Fedorenko et al., 2011; Fedorenko & Blank, 2020; Malik-Moraleda et al., 2022). In particular, in previous functional MRI studies, these strong and selective cortical responses were not visible during the presentation of short musical excerpts, and are hypothesized to index linguistic processes (Chen et al., 2023; Fedorenko et al., 2011). Moreover, in the superior temporal gyrus, specific and separate neural populations for speech, music, and song are visible (Boebinger et al., 2021; Norman-Haignere et al., 2022). These selective responses, not visible in primary cortical regions, seem independent of both low-level acoustic features and higher-order linguistic meaning (Norman-Haignere et al., 2015), and could subtend intermediate representations (Giordano et al., 2023) such as domain-dependent predictions (McCarty et al., 2023; Sankaran et al., 2023). Within this framework, the localizationism view applies to highly specialized processes (i.e. functional niches), while general cognitive domains are mostly spatially distributed. Recent studies have shown that some communicative signals (e.g. alarm, emotional, linguistic) can exploit distinct acoustic niches to target specific neural networks and trigger reactions adapted to the intent of the emitter (Albouy et al., 2020; Arnal et al., 2019). Using neurally relevant spectro-temporal representations (MPS), these studies show that different subspaces encode distinct information types: slow temporal modulations for meaning (speech), fast temporal modulations for alarms (screams), and spectral modulations for melodies (Albouy et al., 2020; Arnal et al., 2015, 2019; Flinker et al., 2019). Which acoustic features—and which neural mechanisms—are necessary and sufficient to route communicative sounds towards selective neural networks remains a promising field of investigation to explore.

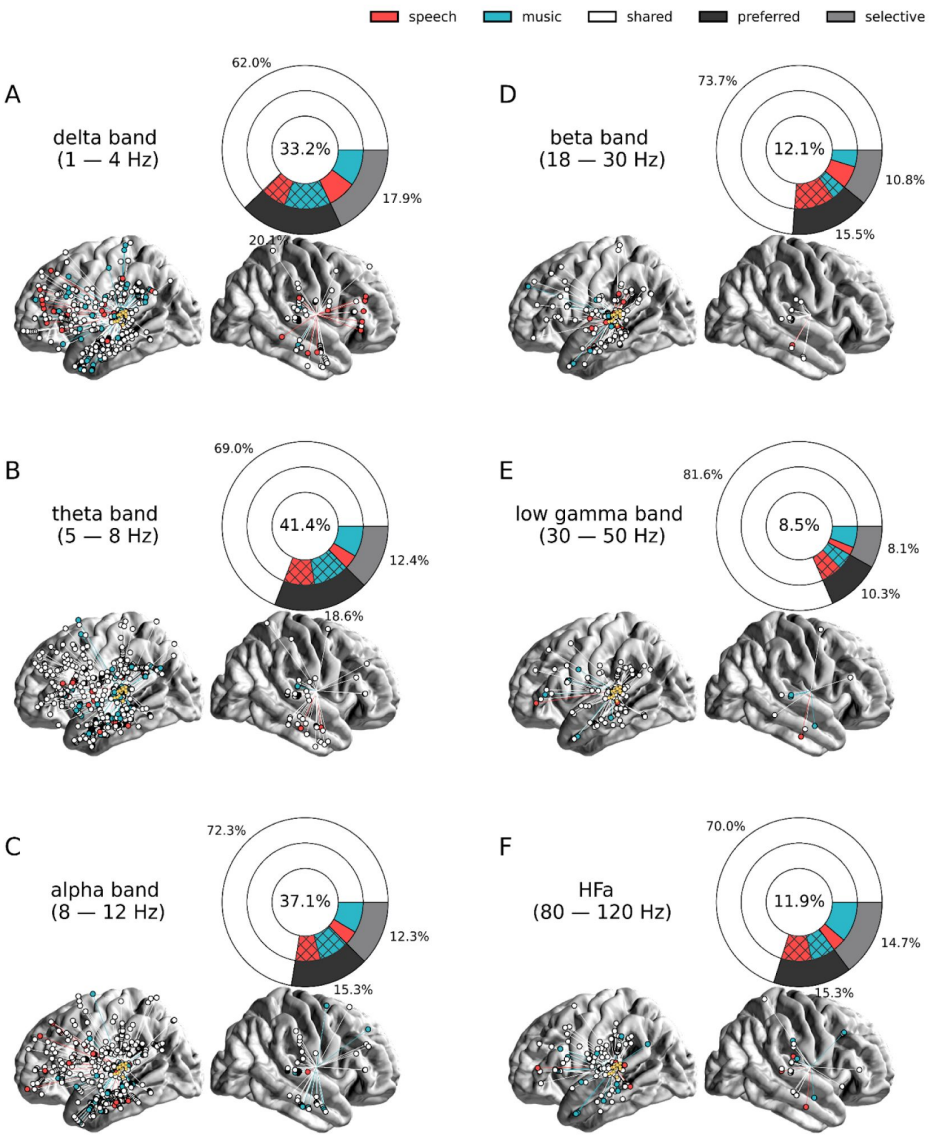


Figure 8.

Seed-based functional connectivity analyses.

(**A**)-(**F**) Significant coherence responses to speech and/or music for the 6 canonical frequency bands (see Figure S10 for uncategorized, continuous results). The seed was located in the left auditory cortex (see **Figure 7D**). Same conventions as in **Figure 2**, except for the center percentage in the nested pie charts which, here, reflects the percentage of channels significantly connected to the seed. Results are significant at $q < 0.01$ (N=15).

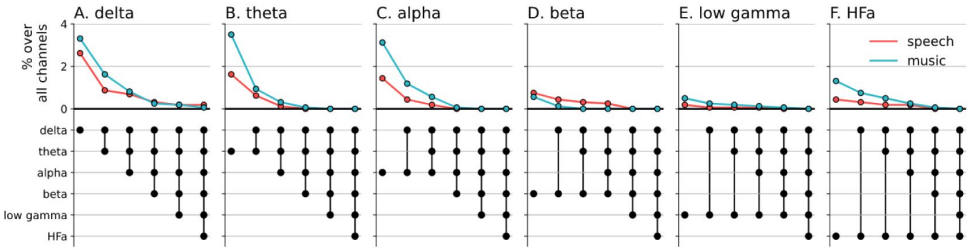


Figure 9.

Cross-frequency channel selectivity for the connectivity analyses.

Percentage of channels that showed selective coherence with the primary auditory cortex in speech (red) or music (blue) across different frequency bands. Same conventions as in **Figure 4**.

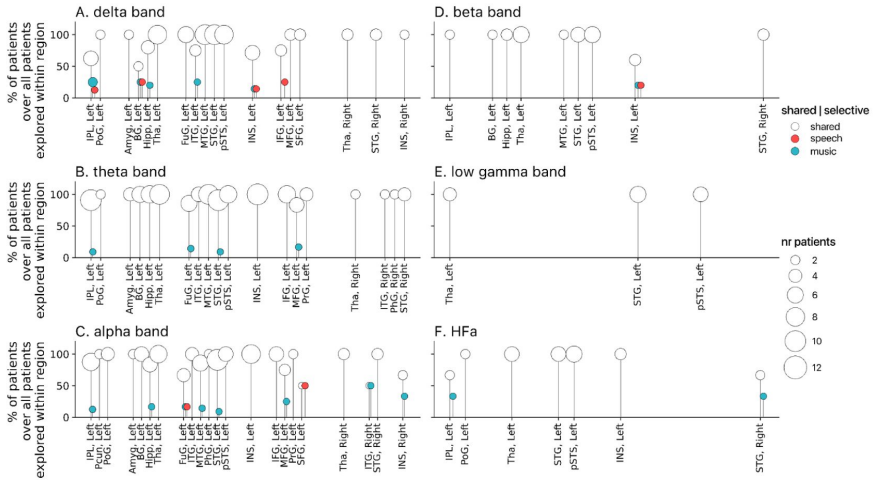


Figure 10.

Population prevalence for the connectivity analyses (N=15).

Same conventions as in **Figure 5**.

In this context, in the current study we did not observe a single anatomical region for which speech-selectivity was present, in any of our analyses. In other words, 10 minutes of instrumental music was enough to activate cortical regions classically labeled as speech (or language) –selective. On the contrary, we report spatially distributed and frequency-specific patterns of shared, preferred, or selective neural responses and connectivity fingerprints. This indicates that domain-selective brain regions should be considered as a set of functionally homogeneous but spatially distributed voxels, instead of anatomical landmarks. Several non-exclusive explanations may account for this finding. First, our results part with the simple selective versus shared dichotomy and adopt a more biologically valid and continuous framework (Buzsáki, 2019; Zatorre & Gandour, 2008) by adding a new category that is often neglected in the literature: *preferred* responses (**Figure 1A**). Indeed, responses in this category are usually reported as shared or selective and most often the statistical approach does not allow a more nuanced view (*cf* Chen et al., 2023). However preferred responses, namely responses that are stronger to a given class of stimuli but that are also present with other stimuli, are relevant and should not be collapsed into either the selective or shared categories. Introducing this intermediate category refines the epistemological and statistical approach on how to map cognitive and brain functions. It points toward the presence of gradients of neural activity across cognitive domains, instead of all-or-none responses. This framework is more compatible with the notion of distributed representations wherein specific regions are more-or-less recruited depending on their relative implication in a distributed manifold (Elman, 1991; Rissman & Wagner, 2012).

Second, most of the studies that reported regional-selectivity are grounded on functional MRI data that lack a precise temporal resolution. Furthermore, the few studies assessing selectivity with intracranial EEG recordings analyzed only the HFa amplitude (Bellier et al., 2022; Norman-Haignere et al., 2020; Oganian & Chang, 2019). However, while this latter reflects local (Kopell et al., 2000) and possibly feedforward activity (Bastos et al., 2015; Fontolan et al., 2014; Fries, 2015) other frequency bands are also constitutive of the cortical dynamics and involved in cognition. For instance, alpha/beta rhythms play a role in predicting upcoming stimuli and modulating sensory processing and associated spiking (Arnal & Giraud, 2012; Bastos et al., 2020; Morillon & Baillet, 2017; Saleh et al., 2010; van Kerkoerle et al., 2014). Also slower dynamics in the delta/theta range have been described to play a major role in cognitive processes and in particular for speech perception, contributing to speech tracking, segmentation and decoding (Ding et al., 2017; Doelling et al., 2014; Giraud & Poeppel, 2012; Gross et al., 2013; Keitel et al., 2017). Importantly, we here addressed both activations and deactivations that can co-occur in the same spatial location across different frequency bands (Pfurtscheller & Lopes da Silva, 1999, Proix et al., 2022) and indeed observed that the domain-selectivity observed within our restricted stimulus set is frequency-specific, meaning that domain-selectivity is marginal when considering the entire spectrum of activity of a given sEEG channel. Finally, most studies only investigated local neural activity and did not consider the brain as a distributed system, analyzed through the lens of functional connectivity analyses. While topological approaches are more complex, they also provide more nuanced and robust characterization of brain functions. Critically, our approach reveals the limitation of adopting a reductionist approach —either by considering the brain as a set of independent regions instead of distributed networks, or by overlooking the spectral complexity of the neural signal.

Third, the ecological auditory stimuli we used are longer and more complex than stimuli used in previous studies and hence more prone to elicit distributed and dynamical neural responses (Hasson et al., 2010; Sonkusare et al., 2019; Theunissen et al., 2000) and they require, in the case of music, for instance, more complex representations of melody and rhythm motifs contributing to stronger representations of meter, tonality, and groove (Boebinger et al., 2021). While listening to natural speech and music rests on cognitively relevant neural processes, our analytical approach, extending over a rather long period of time, does not allow to directly isolate specific brain operations. Computational models—which can be as diverse as acoustic (Chi et al., 2005), cognitive (Giordano et al., 2021), information-theoretic (Di Liberto et al., 2020;

Donhauser & Baillet, 2019), or self-supervised neural networks (Donhauser & Baillet, 2019; Millet et al., 2022; Sankaran et al., 2023) models—are hence necessary to further our understanding of the type of computations performed by our reported frequency-specific distributed networks. Moreover, incorporating models accounting for musical and linguistic structure can help us avoid misattributing differences between speech and music driven by unmatched sensitivity factors (e.g., arousal, emotion, or attention) as inherent speech or music selectivity (Mas-Herrero et al., 2013; Nantais & Schellenberg, 1999).

Our modeling approach, although lacking the modeling of melodic and linguistic features, was targeting the temporal dynamics of the speech and music stimuli. Beyond confirming that acoustic dynamics are strongly tracked by auditory neural dynamics, it revealed, investigating the entire cortex, that such neural tracking also occurs well outside of auditory regions—up to motor and inferior frontal areas (**Figure 7B**; see also Chalas et al., 2022; Zion Golumbic et al., 2013). Of note, this spatial map of speech dynamics encoding is very similar to former reports of the brain regions belonging to the language system (Diachek et al., 2020). But, here again, adopting an approach that investigates both low and high frequencies of the neural signal—an approach that is not enough embraced in intracranial EEG studies (Proix et al., 2022)—reveals that the low frequency activity clearly better encodes acoustic features than the HFa amplitude (**Figure 7A**).

In conclusion, our results point to a massive amount of shared neural response to speech and music, well beyond the auditory cortex. They also show the interest of considering shared, preferred, and selective responses when investigating domain selectivity. Importantly these three classes of responses should be considered in respect to 1) activation or deactivation patterns compared to a baseline, 2) different frequency bands and 3) power spectrum (activity) and connectivity approaches. Combining all these points of view, gives a richer although possibly more complex view of brain functions. While our data point to an absence of anatomical regional selectivity for speech and music, such a selectivity still exists at the level of a spatially distributed and frequency-specific network. Thus, the inconsistency with previous findings may be limited to the idea that some anatomical regions are selective to speech or music processing. However, the two points of view can be reconciled when considering a fine-grained network approach allowing selectivity to coexist for speech and music within the same anatomical region. Finally, in adopting here a comparative approach of speech and music—the two main auditory domains of human cognition—we only investigated one type of speech and of music during a passive listening task. Future work is needed to investigate for instance whether different sentences or melodies activate the same selective frequency-specific distributed networks and to what extent these results are related to the passive listening context compared to a more active and natural context (e.g. conversation).

Methods

Participants

18 patients (10 females, mean age 30 y, range 8 – 54 y) with pharmacoresistant epilepsy participated in the study. All patients were French native speakers. Neuropsychological assessments carried out before stereotactic EEG (sEEG) recordings indicated that all patients had intact language functions and met the criteria for normal hearing. In none of them were the auditory areas part of their epileptogenic zone as identified by experienced epileptologists. Recordings took place at the Hôpital de La Timone (Marseille, France). Patients provided informed consent prior to the experimental session, and the experimental protocol was approved by the Institutional Review board of the French Institute of Health (IRB00003888).

Data acquisition

The sEEG signal was recorded using depth electrodes shafts of 0.8 mm diameter containing 10 to 15 electrode contacts (Dixi Medical or Alcis, Besançon, France). The contacts were 2 mm long and were spaced from each other by 1.5 mm. The locations of the electrode implantations were determined solely on clinical grounds. Patients were included in the study if their implantation map covered at least partially the Heschl's gyrus (left or right). The cohort consists of 13 unilateral implantations (10 left, 3 right) and 5 bilateral implantations, yielding a total of 271 electrodes and 3371 contacts (see **Figure 1C-D** for electrodes localization).

Patients were recorded either in an insulated Faraday cage or in the bedroom. In the Faraday cage, they laid comfortably in a chair, the room was sound attenuated and data were recorded using a 256-channels amplifier (Brain Products), sampled at 1kHz and high-pass filtered at 0.016 Hz. In the bedroom, data were recorded using a 256-channels Natus amplifier (Deltamed system), sampled at 512 Hz and high-pass filtered at 0.16 Hz.

Experimental design

Patients completed three separate sessions. In one session they completed the main experimental paradigm and the two additional sessions served as baseline for the spectral analysis (see below).

In the main experimental session, patients passively listened to ~10 minutes of storytelling (Gripari, 2004); 577 secs, *La sorcière de la rue Mouffetard, (Gripari, 2004)* and ~10 minutes of instrumental music (580 secs, *Reflejos del Sur, (Oneness, 2006)* separated by 3 minutes of rest. The order of conditions was counterbalanced across patients (see Table S1). This session was conducted in the Faraday cage (N=6) or in the bedroom (N=12).

In the two baseline sessions, patients passively listened to two more basic types of auditory stimuli: 1) 30-ms–long pure tones, presented binaurally at 500 Hz or 1 kHz (with a linear rise and fall time of 0.3 ms) 110 times each, with an ISI of 1,030 (±200) ms; and 2) /ba/ or /pa/ syllables, pronounced by a French female speaker and presented binaurally 250 times each, with an ISI of 1,030 (±200) ms. These stimuli were designed for a clinical purpose in order to functionally map the auditory cortex. These two recording sessions (lasting ~ 2 and 4 minutes) were performed in the Faraday cage.

In the Faraday cage, a sound Blaster X-Fi Xtreme Audio, an amplifier Yamaha P2040 and Yamaha loudspeakers (NS 10M) were used for sound presentation. In the bedroom, stimuli were presented using a Sennheiser HD 25 headphone set. Sound stimuli were presented at 44.1 kHz sample rate and 16 bits resolution. Speech and music excerpts were presented at ~75 dBA (see **Figure 1B**).

General preprocessing related to electrodes localisation

To increase spatial sensitivity and reduce passive volume conduction from neighboring regions (Mercier et al., 2017, 2022), the signal was offline re-referenced using bipolar montage. That is, for a pair of adjacent electrode contacts, the referencing led to a virtual channel located at the midpoint locations of the original contacts. To precisely localize the channels, a procedure similar to the one used in the iELVis toolbox and in the fieldtrip toolbox was applied (Groppe et al., 2017; Stolk et al., 2018). First, we manually identified the location of each channel centroid on the post-implant CT scan using the Gardel software (Medina Villalon et al., 2018). Second, we performed volumetric segmentation and cortical reconstruction on the pre-implant MRI with the Freesurfer image analysis suite (documented and freely available for download online *http://surfer.nmr.mgh.harvard.edu/*). This segmentation of the pre-implant MRI with SPM12 provides us with both the tissue probability maps (i.e. gray, white, and cerebrospinal fluid (CSF) probabilities) and the indexed-binary representations (i.e., either gray, white, CSF, bone, or soft tissues). This

information allowed us to reject electrodes not located in the brain. Third, the post-implant CT scan was coregistered to the pre-implant MRI via a rigid affine transformation and the pre-implant MRI was registered to MNI152 space, via a linear and a non-linear transformation from SPM12 methods (Penny et al., 2011), through the FieldTrip toolbox (Oostenveld et al., 2011). Fourth, applying the corresponding transformations, we mapped channel locations to the pre-implant MRI brain that was labeled using the volume-based Human Brainnetome Atlas (Fan et al., 2016).

Based on the brain segmentation performed using SPM12 methods through the Fieldtrip toolbox, bipolar channels located outside of the brain were removed from the data (3%). The remaining data (**Figure 1C**) was then bandpass filtered between 0.1 Hz and 250 Hz, and, following a visual inspection of the power spectral density (PSD) profile of the data, when necessary, we additionally applied a notch filter at 50 Hz and harmonics up to 200 Hz to remove power line artifacts (N=12). Finally, the data were downsampled to 500 Hz.

Artifact rejection

To define artifacted channel we used both the broadband signal and the amplitude of the high frequency activity. This latter was obtained by computing, with the Hilbert transform, the analytic amplitude of four 10-Hz-wide sub-bands spanning from 80 to 120 Hz. Each sub-band was standardized by dividing it by its mean and, finally, all sub-bands were averaged together (Ossandón et al., 2012; Vidal et al., 2012). Channels with a variance greater than 2*IQR (interquartile range, i.e. a non-parametric estimate of the standard deviation)—on either the broadband or high frequency signals—were tagged as artifacted channels (on average 18% of the channels). Then the data were epoched in non-overlapping segments of 5 seconds (2500 samples). To exclude artifacted epochs, epochs wherein the maximum amplitude (over time) summed across non-excluded channels was greater than 2*IQR were tagged as artifacted epochs. Overall, 6% of the speech epochs and 7% of the music epochs were rejected. Channels and epochs defined as artifacted were excluded from subsequent analyses, except if specified otherwise (see TRF section).

Spectral analysis

Six canonical frequency bands were investigated: delta (1-4 Hz), theta (5-8 Hz), alpha (8-12 Hz), beta (18-30 Hz), low-gamma (30-50 Hz), and high-frequency activity (HFa; 80-120 Hz). To prevent edge artifacts, prior to extracting the power spectrum, epochs were zero-padded on both sides with 3.5-seconds segments which were later removed. For each patient, channel, epoch, and frequency band, the power of the neural signal was calculated using the Welch approach on Discrete Fourier Transform from the scipy-python library (Virtanen et al., 2020) and then averaged across the relevant frequencies to obtain these six canonical bands.

For each canonical band and each channel, we classified the time-averaged neural response as being selective, preferred, or shared across the two investigated cognitive domains (speech, music). We defined these categories by capitalizing on both the simple effects of —and contrast between— the neural responses to speech and music stimuli compared to a baseline condition (see **Figure 1A**). “Selective” responses are neural responses that are significantly different compared to the baseline for one domain (speech or music) but not the other, and with a significant difference between domains (i.e. speech or music is different from baseline + difference effect between the domains). “Preferred” responses correspond to neural responses that occur during both speech and music processing, but with a significantly stronger response for one domain over the other (i.e. both speech and music are significantly different from baseline + difference effect between the domains). Finally, “shared” responses occur when there are no significant differences between domains, and there is a significant neural response to at least one of the two stimuli (one or two simple effects + no difference). If none of the two domains produces a significant neural response, the difference is not assessed (case “neither” simple effect). In order to explore the full range of possible selective, preferred, or shared responses, we considered both responses greater

and smaller than the baseline. Indeed, as neural populations can synchronize or desynchronize in response to sensory stimulation, we estimated these categories separately for significant activations and significant deactivations compared to baseline.

For each frequency band and channel, the statistical difference between conditions was estimated with paired sample permutation tests based on the t-statistic from the mne-python library (Gramfort et al., 2014) with 1000 permutations and the *tmax* method to control the family-wise error rate (Groppe et al., 2011; Nichols & Holmes, 2002). In *tmax* permutation testing, the null distribution is estimated by, for each channel (i.e. each comparison), swapping the condition labels (speech vs music or speech/music vs baseline) between epochs. After each permutation, the most extreme t-scores over channels (*tmax*) are selected for the null distribution. Finally, the t-scores of the observed data are computed and compared to the simulated *tmax* distribution, similar as in parametric hypothesis testing. Because with an increased number of comparisons, the chance of obtaining a large *tmax* (i.e. false discovery) also increases, the test automatically becomes more conservative when making more comparisons, as such correcting for the multiple comparison between channels.

Temporal Response Function (TRF) analysis

We used the Temporal Response Function (TRF) to estimate the encoding of acoustic features by All computations of the TRF used the pymTRF library (Steinkamp, 2019), a python adaption of the mTRF toolbox (Crosse et al., 2016). A TRF is a model that, via linear convolution, serves as a filter to quantify the relationship between two continuous signals, here stimulus features and neural activity. Hence, for this analysis, the entire duration of the recordings were preserved, i.e., no artifacted epochs were excluded. When applied in a forward manner, the TRF approach describes the mapping of stimulus features onto the neural response (henceforth 'encoding'; Crosse et al., 2016). Using ridge regression to avoid overfitting, we examined how well the two different acoustic features—envelope and peakRate—map onto low frequency activity (LF; 1-9 Hz) or the amplitude of the high frequency activity (80-120 Hz, see Artifact rejection section) (Ding et al., 2016; Zion Golumbic et al., 2013). Hence four encoding models were estimated: envelope/peakRate acoustic features * LF/amplitude of HFa neural activity. For each model and patient, the optimal ridge regularization parameter (λ) was estimated using cross-validation on the sEEG channels situated in the auditory cortex. We considered time lags from –150 to 1000 ms for the TRF estimations. 80% of the data was used to derive the TRFs and the remaining 20% was used as a validation set. The quality of the predicted neural response was assessed by computing Pearson's product moment correlations (Fisher-z-scored) between the predicted and actual neural data for each channel and model using the scipy-python library (p-values FDR-corrected).

Models were finally compared in terms of the percentage of channels that significantly encoded the acoustic structure of speech and/or music. This percentage was estimated at the single-subject level and combined with non-parametric Wilcoxon sign rank tests at the group level to define the winning model. In other words, the winning model is the model for which the percentage of channels significantly encoding speech and/or music acoustic features is the largest. Multiple comparison across pairs of models was controlled for with a FDR correction.

Connectivity analysis

We examined the frequency-specific functional connectivity maps in response to speech and music, between the entire brain and the auditory cortex using a seed-based approach (we dismissed the channels immediately neighboring the seed-channel). As seed, we selected, per patient, the channel that best encoded the speech and music acoustic features (see TRF analysis; **Figure 7D**). We used spectral coherence as a connectivity measure for all canonical bands (see above) and all analyses were performed using the mne-python library (Gramfort et al., 2014). Our rationale to use coherence as functional connectivity metric was three fold. First, coherence analysis considers both magnitude and phase information. While the absence of dissociation can

be criticized, signals with higher amplitude and/or SNR lead to better time-frequency estimates (which is not the case with a metric that would focus on phase only and therefore would be more likely to include estimates of various SNR). Second, we choose a metric that allows direct comparison between frequencies. As, at high frequencies phase angle changes more quickly, phase alignment/synchronization is less likely in comparison with lower frequencies. Third, we intend to align to previous work which, for the most part, used the measure of coherence most likely for the reasons explained above.

For each frequency band, we classified each channel into selective, preferred, or shared categories (see **Figure 1A**) by examining both the simple effects (i.e. which channels display a significantly coherent signal with the seed during speech and/or music processing) and the difference effects (i.e. is coherence significantly stronger for one domain over the other).

Statistical significance was assessed for each frequency band and channel using surrogate data with 1000 iterations, which were generated by modifying the temporal structure of the sEEG-signal recorded at the seeds (i.e. shuffling the epochs) prior to computing connectivity. This process led to a total of 1000 connectivity values, which were used as null-distribution to calculate the probability threshold associated with genuine connectivity.

Population prevalence

For both the spectral and the connectivity analyses, in order to make sure that the results are not driven by the heterogeneity of electrode locations across patients, we examined, for each region, the proportions of patients showing only shared or selective responses. That is, for both the spectral and connectivity results, we examined results representativeness as follows: for each anatomical region wherein at least two patients have at least two significantly responsive channels, we computed the percentage of patients that showed a pattern of selective (i.e. all channels selective to speech or music) or a shared (i.e. a mixture of channels responding to speech and/or music) responses. This approach is inspired by the population prevalence, where an equivalent metric is introduced (i.e., the Maximum A Posterior estimate see (Ince et al., 2021)).

Conflict of interests

The authors declare no competing interests.

Acknowledgements

We thank all patients for their willingful participation. We thank Patrick Marquis for helping with the data acquisition, and Anne-Catherine Tomei and all colleagues from the Institut de Neuroscience des Systèmes for useful discussions.

Funding sources

ANR-20-CE28-0007-01 (to B.M), ANR-21-CE28-0010 (to D.S), ANR-17-EURE-0029 (NeuroMarseille), and co-funded by the European Union (ERC, SPEEDY, ERC-CoG-101043344). This work, carried out within the Institute of Convergence ILCB, was also supported by grants from France 2030 (ANR-16-CONV-0002), the French government under the Programme «Investissements d'Avenir», and the Excellence Initiative of Aix-Marseille University (A*MIDEX, AMX-19-IET-004).

Author contributions

B.M. and D.S. designed research. A.T. and M.M. acquired data; N.t.R., M.M., B.M. and D.S. analyzed data; N.t.R., B.M., M.M., and D.S. wrote the paper; N.t.R., A.T., M.M., B.M., and D.S. edited the paper.

Data availability statement

The conditions of our ethics approval do not permit public archiving of anonymised study data. Readers seeking access to the data should contact Dr. Daniele Schön (daniele.schon@univ-amu.fr). Access will be granted to named individuals in accordance with ethical procedures governing the reuse of clinical data, including completion of a formal data sharing agreement.

Code availability statement

Data analyses were performed using custom scripts in Python, available on Github: github.com/noemietr/iSpeech

Supplementary Figures

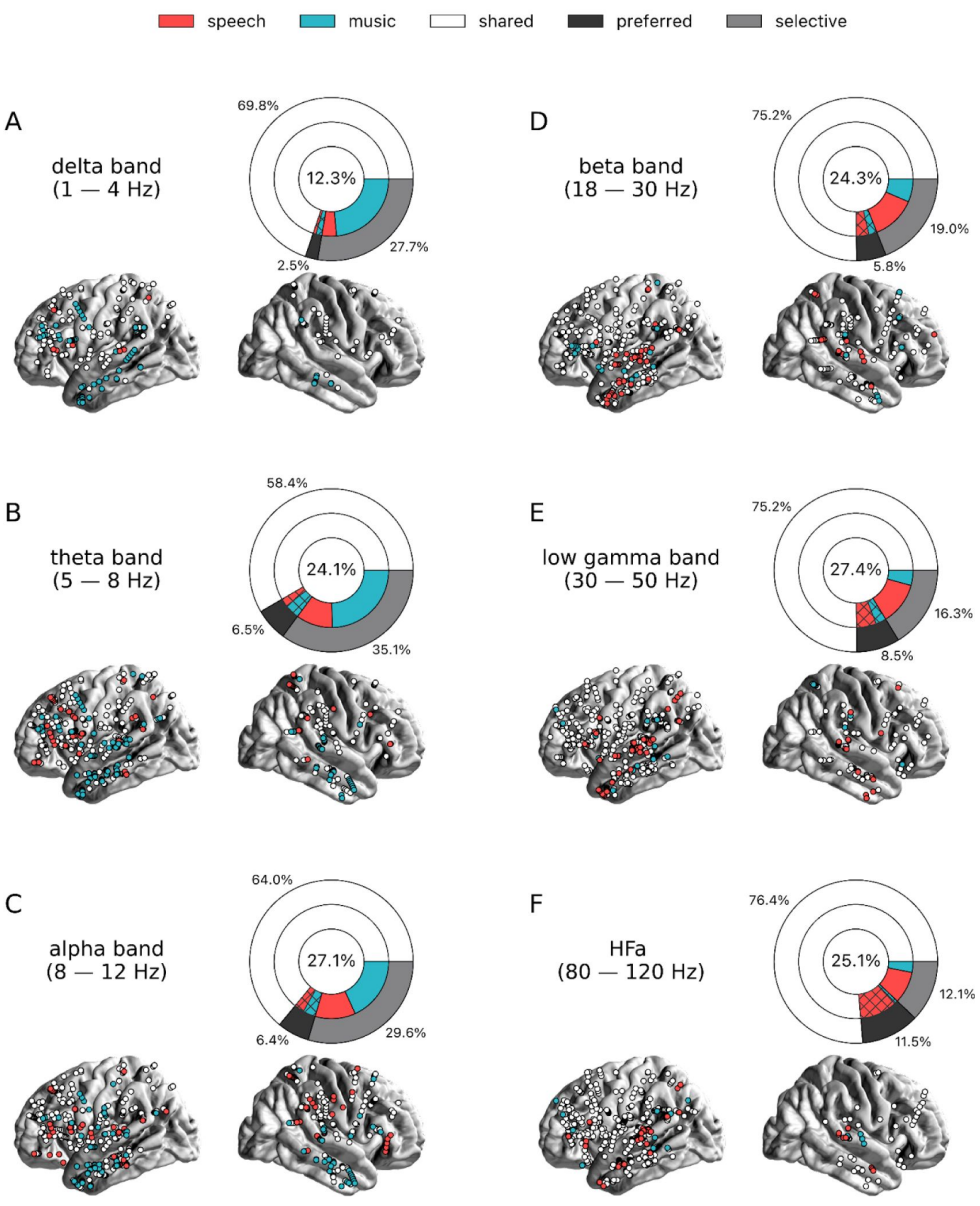


Figure S1.

Power spectrum analyses of activations (speech or music > syllables).

(A)-**(F)** Neural responses to speech and/or music for the 6 canonical frequency bands. Only significant activations compared to the baseline condition (syllables listening) are reported. Same conventions as in **Figure 2**. Results are significant at q < 0.01 (N=18).

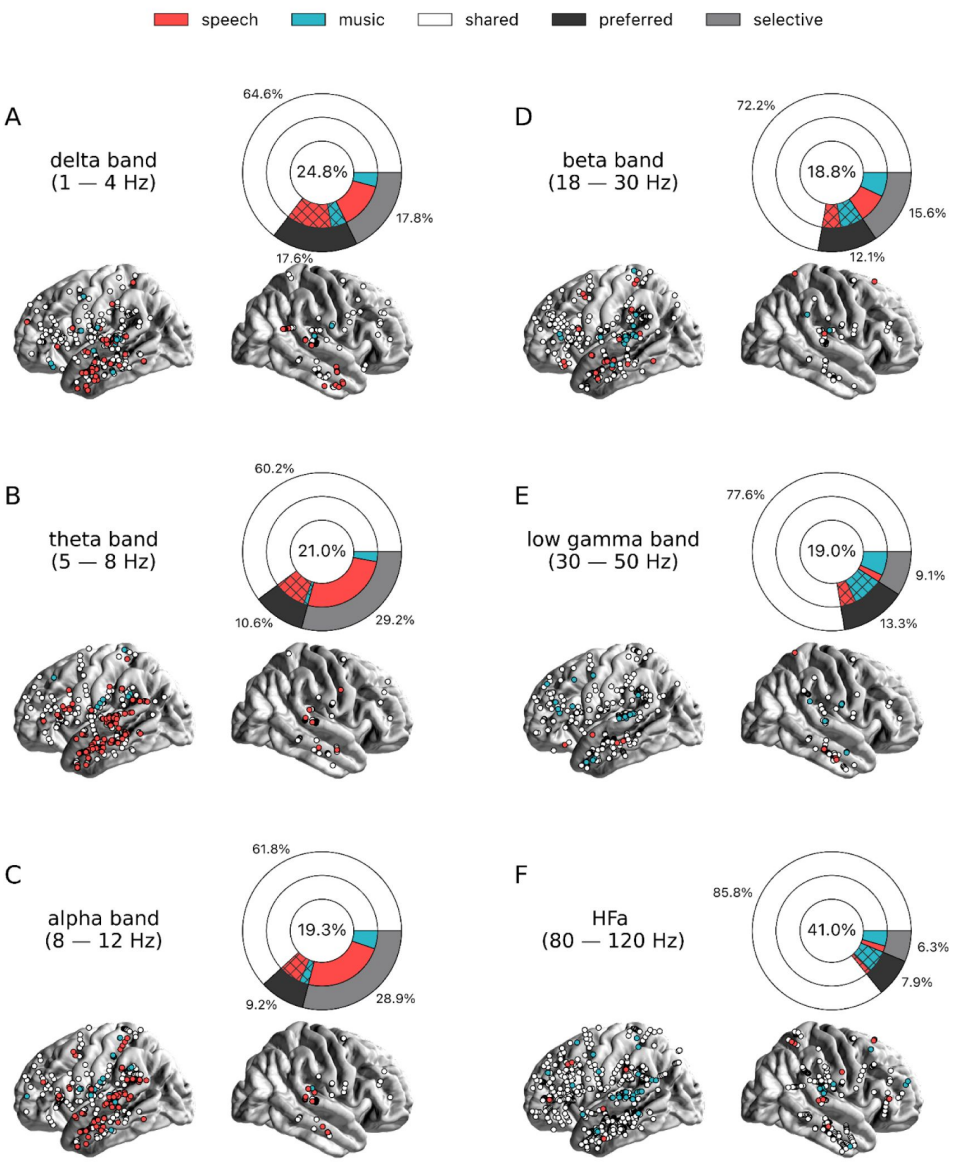


Figure S2.

Power spectrum analyses of deactivations (speech or music < syllables).

(**A**)-(**F**) Neural responses to speech and/or music for the 6 canonical frequency bands. Only significant deactivations compared to the baseline condition (syllables listening) are reported. Same conventions as in **Figure 2**. Results are significant at q < 0.01 (N=18).

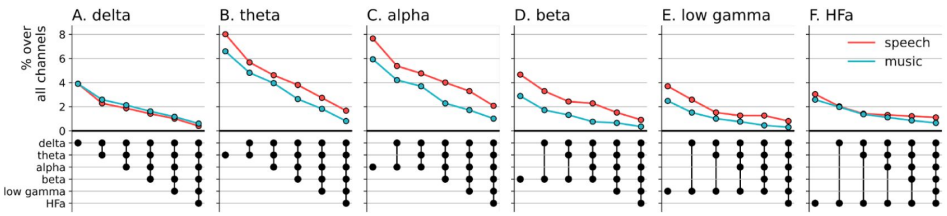


Figure S3.

Cross-frequency channel selectivity for the power spectrum analyses (syllables baseline).

Percentage of channels that selectively respond to speech (red) or music (blue) across different frequency bands, compared to the baseline condition (syllable listening). Same conventions as in **Figure 4**.

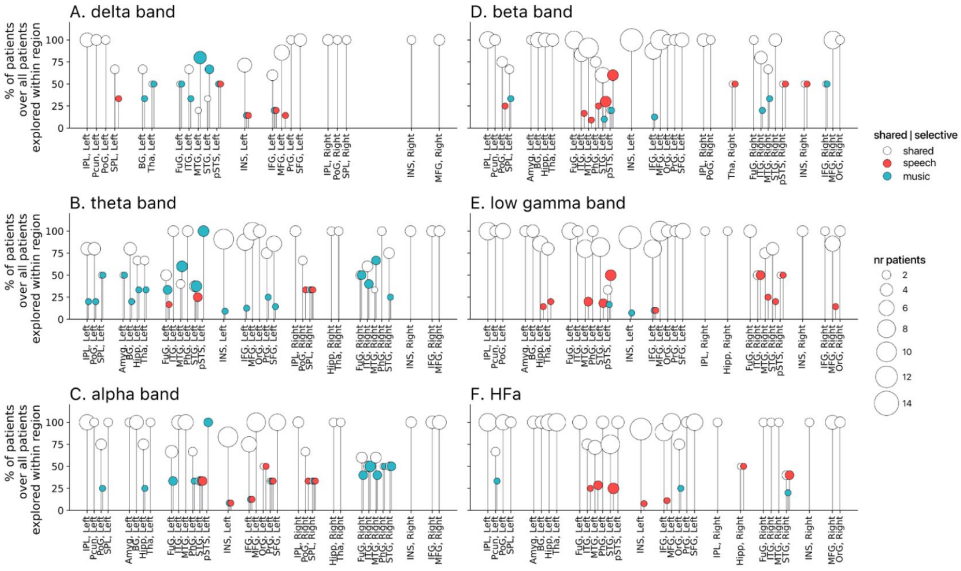


Figure S4.

Population prevalence for the power spectral analyses of activations (speech or music > syllables; N = 18).

(A)-**(F)** Population prevalence of shared or selective responses for the 6 canonical frequency bands, per anatomical region. Only significant activations compared to the baseline condition (syllables listening) are reported. Same conventions as in **Figure 5**.

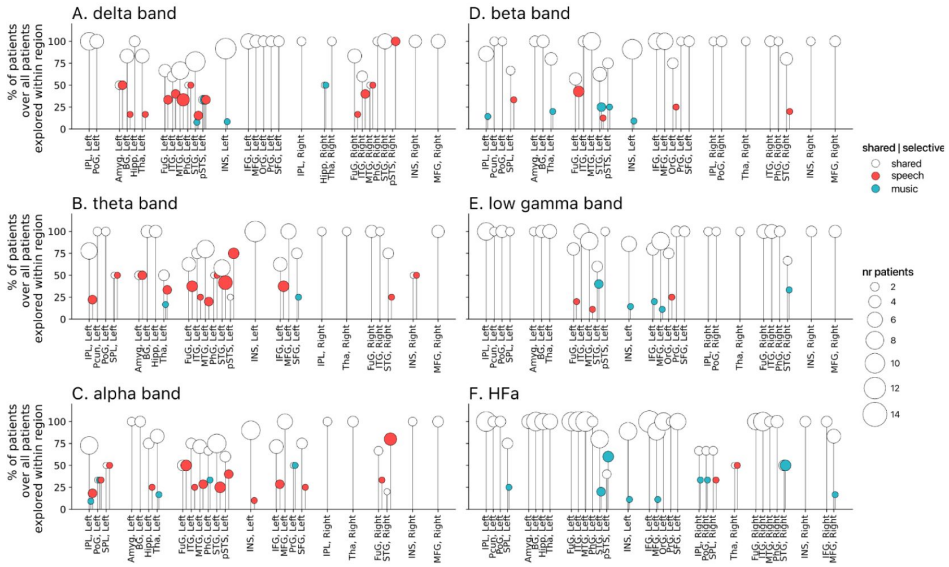


Figure S5.

Population prevalence for the power spectral analyses of deactivations (speech or music < syllables; N = 18).

(A)-**(F)** Population prevalence of shared or selective responses for the 6 canonical frequency bands, per anatomical region. Only significant deactivations compared to the baseline condition (syllables listening) are reported. Same conventions as in **Figure 5**.

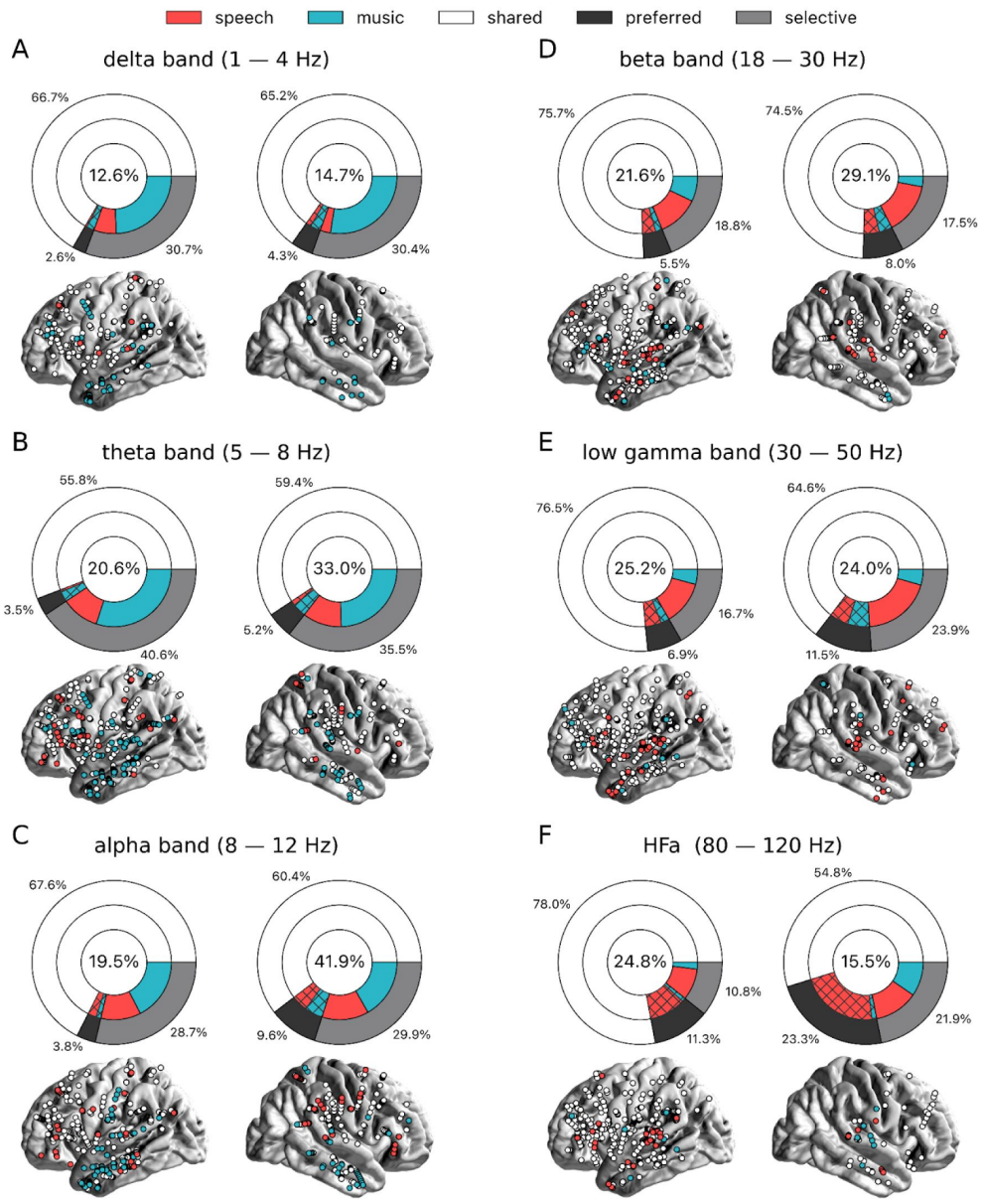


Figure S6.

Power spectrum analyses of activations (speech or music > tones) for each hemisphere separately.

Same conventions as in **Figure 2**. Results are significant at q < 0.01 (N=18).

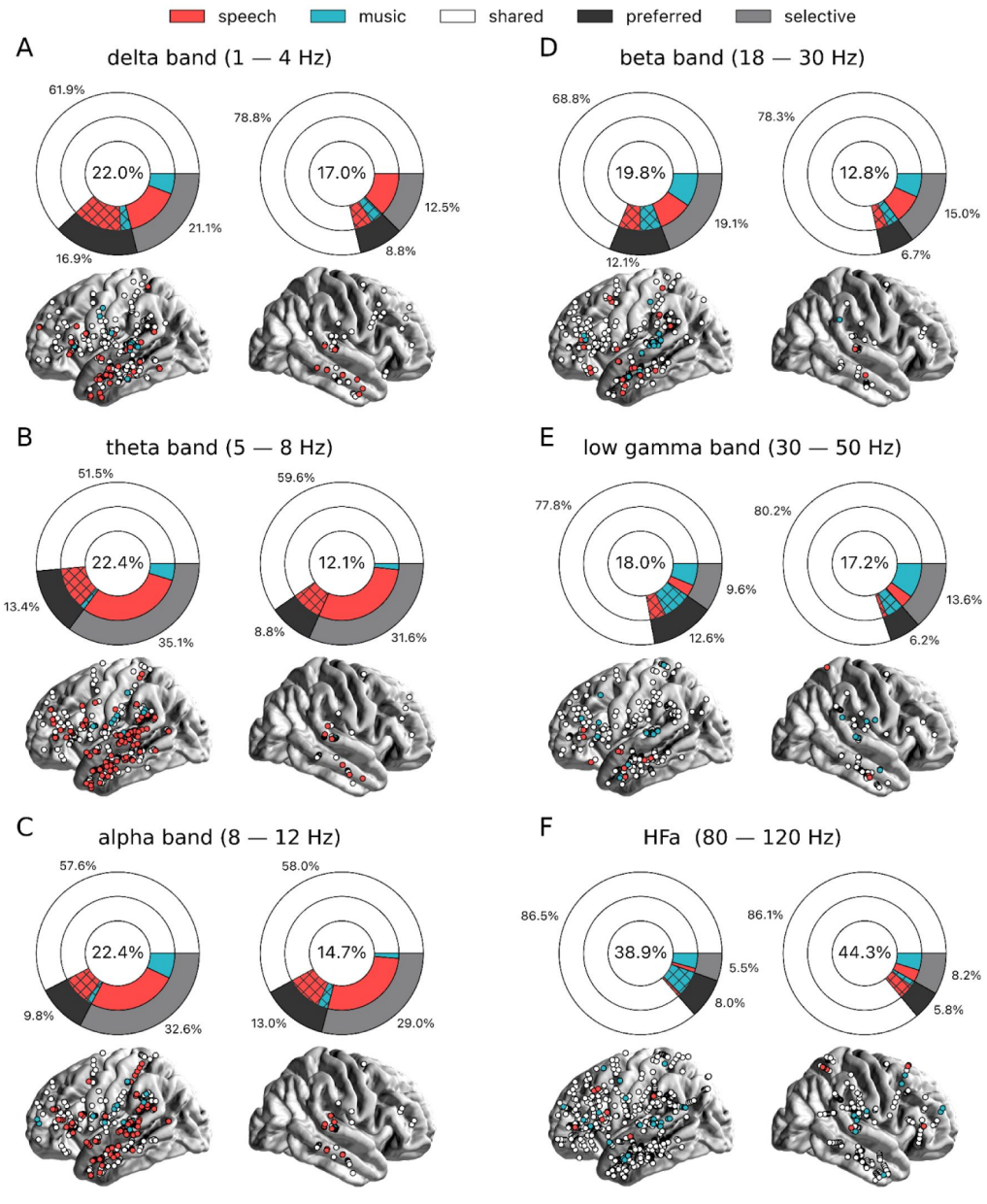


Figure S7.

Power spectrum analyses of deactivations (speech or music < tones) for each hemisphere separately.

Same conventions as in **Figure 2**. Results are significant at q < 0.01 (N=18).

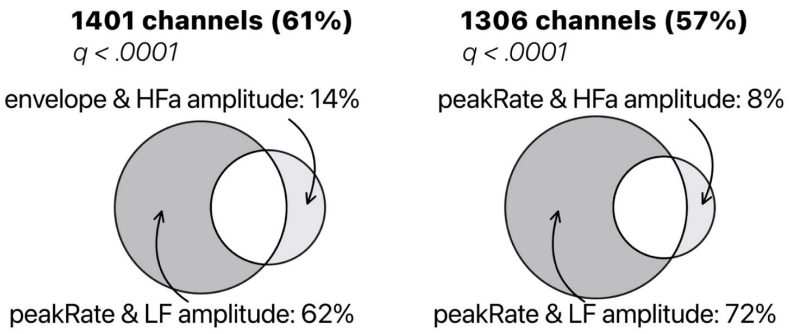


Figure S8.

TRF model comparison of low-frequency (LF) amplitude and high-frequency (HFa) amplitude.

Models were investigated to quantify the encoding of the instantaneous envelope and the discrete acoustic onset edges (peakRate) by either the low frequency (LF) amplitude or the high frequency (HFa) amplitude. The 'peakRate & LF amplitude' model significantly captures the largest proportion of channels, and is, therefore, considered the winning model. Same conventions as in **Figure 7A**.

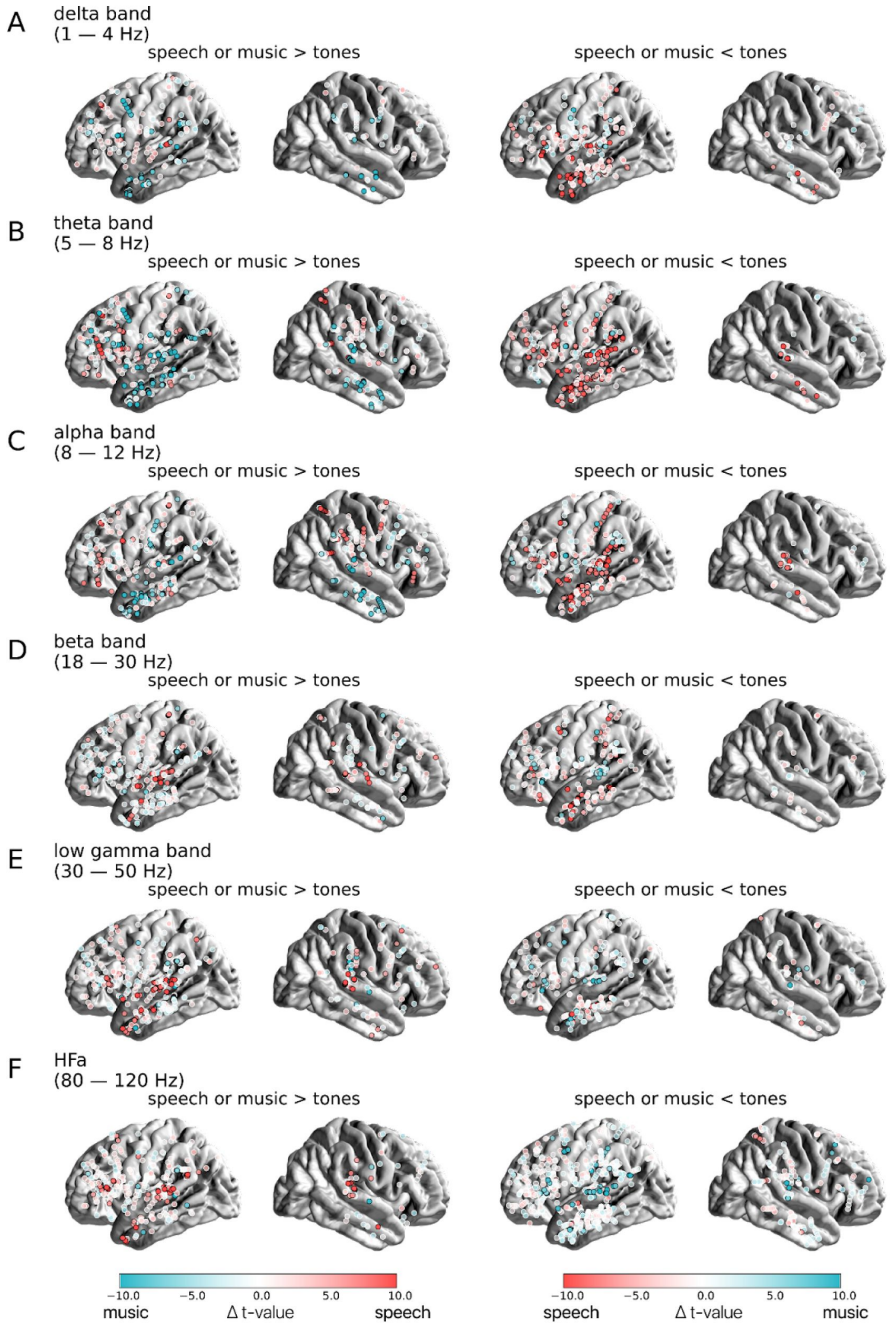


Figure S9.

Contrast between the neural responses to speech and music, for the 6 canonical frequency bands (tones baseline).

Distribution of uncategorized, continuous t-values from the speech vs. music permutation test, projected on the brain surface. Only sEEG channels for which speech and/or music is significantly stronger (left column) or weaker (right column) than the baseline condition (pure tones listening) are reported. Channels for which the speech vs. music contrast is significant are circled in black. Channels for which the speech vs. music contrast is not significant are circled in white. Results are significant at $q < 0.01$ (N=18).

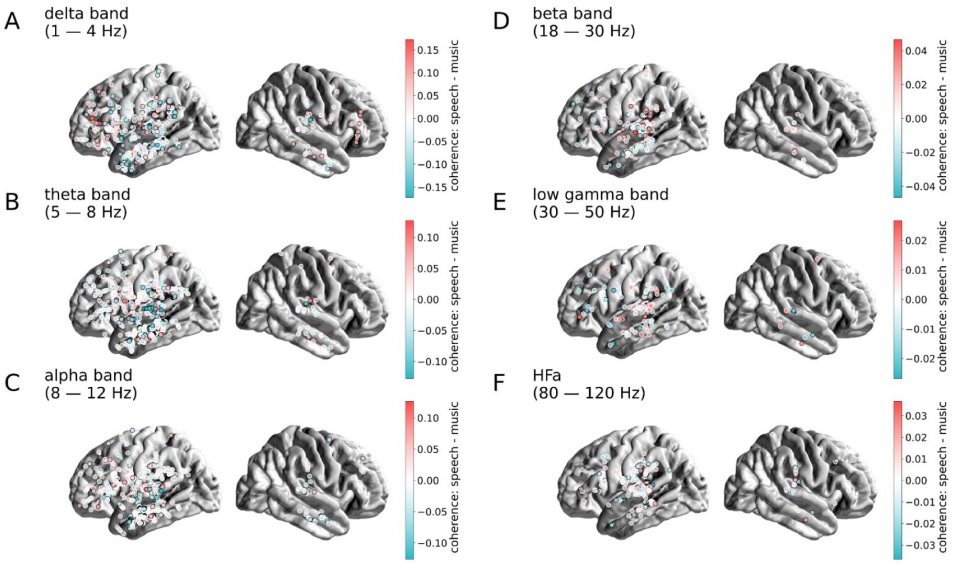


Figure S10.

Contrast between the coherence responses to speech and music, for the 6 canonical frequency bands.

Distribution of uncategorized, continuous differences in coherence values between speech and music, projected on the brain surface. Same conventions as in Figure S8. Results are significant at $q < 0.01$ (N=15).

patient	gender	age	order of presentation	recording site	hemispheric dominance (typical/atypical)	H: left/right	depth electrodes
P1	f	35	speech-music	room	Typical	Bilateral	11 left; 3 right
P2	m	29	speech-music	room	AtypicalB	Left	12 left; 2 right
P3	m	39	speech-music	lab	Typical	Left	14 left; 2 right
P4	m	8	speech-music	lab	Typical	Bilateral	10 left; 2 right
P5	f	36	speech-music	room	Typical	Left	13 left; 3 right
P6	m	44	music-speech	room	Typical	Left	15 left; 2 right
P7	f	37	speech-music	lab	AtypicalB	Bilateral	7 left; 9 right
P8	f	17	speech-music	lab	Typical	Bilateral	11 left; 4 right
P9	f	54	music-speech	room	Typical	Left	15 left; 5 right
P10	f	20	music-speech	lab	Typical	Right	1 left; 11 right
P11	f	37	music-speech	room	Typical	Left	11 left; 2 right
P12	f	37	music-speech	room	Typical	Left	14 left; 2 right
P13	m	24	music-speech	room	AtypicalB	Left	13 left; 2 right
P14	f	30	music-speech	room	Typical	Left	12 left; 1 right
P15	f	22	music-speech	room	Typical	Right	1 left; 11 right
P16	m	22	music-speech	room	Typical	Bilateral	12 left; 3 right
P17	m	32	music-speech	room	Typical	Left	10 left; 2 right
P18	m	17	speech-music	lab	Typical	Right	14 left; 1 right

Table S1

Patients description.

References

1. Albouy P., Benjamin L., Morillon B., Zatorre R. J (2020) **Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody** *Science* **367**:1043–1047

2. Arnal L. H., Flinker A., Kleinschmidt A., Giraud A.-L., Poeppel D (2015) **Human screams occupy a privileged niche in the communication soundscape** *Current Biology: CB* **25**:2051–2056

3. Arnal L. H., Giraud A.-L (2012) **Cortical oscillations and sensory predictions** *Trends in Cognitive Sciences* **16**:390–398

4. Arnal L. H., Kleinschmidt A., Spinelli L., Giraud A.-L., Mégevand P (2019) **The rough sound of salience enhances aversion through neural synchronisation** *Nature Communications* **10**

5. Bastos A. M., Lundqvist M., Waite A. S., Kopell N., Miller E. K (2020) **Layer and rhythm specificity for predictive routing** *Proceedings of the National Academy of Sciences of the United States of America* **117**:31459–31469

6. Bastos A. M., Vezoli J., Bosman C. A., Schoffelen J.-M., Oostenveld R., Dowdall J. R., De Weerd P., Kennedy H., Fries P. (2015) **Visual areas exert feedforward and feedback influences through distinct frequency channels** *Neuron* **85**:390–401

7. Bellier L., Llorens A., Marciano D., Schalk G., Brunner P., Knight R. T., Pasley B. N (2022) **Encoding and decoding analysis of music perception using intracranial EEG** *In bioRxiv* https://doi.org/10.1101/2022.01.27.478085

8. Bizley J. K., Cohen Y. E (2013) **The what, where and how of auditory-object perception** *Nature Reviews. Neuroscience* **14**:693–707

9. Boebinger D., Norman-Haignere S. V., McDermott J. H., Kanwisher N (2021) **Music-selective neural populations arise without musical training** *Journal of Neurophysiology* **125**:2237–2263

10. Buzsáki G (2019) **The Brain from Inside Out**

11. Buzsáki G., Vöröslakos M (2023) **Brain rhythms have come of age** *Neuron* **111**:922–926

12. Chalas N., Daube C., Kluger D. S., Abbasi O., Nitsch R., Gross J (2022) **Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex** *NeuroImage* **258**

13. Chen X., Affourtit J., Ryskin R., Regev T. I., Norman-Haignere S., Jouravlev O., Malik-Moraleda S., Kean H., Varley R., Fedorenko E (2023) **The human language system, including its inferior frontal component in "Broca's area," does not support music perception. Cerebral Cortex** *bhad* **87**

14. Chi T., Ru P., Shamma S. A (2005) **Multiresolution spectrotemporal analysis of complex sounds** *The Journal of the Acoustical Society of America* **118**:887–906

15. Crosse M. J., Di Liberto G. M., Bednar A., Lalor E. C. (2016) **The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli** *Frontiers in Human Neuroscience* **10**

16. Diachek E., Blank I., Siegelman M., Affourtit J., Fedorenko E (2020) **The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation** *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **40**:4536–4550

17. Di Liberto G. M., Pelofi C., Bianco R., Patel P., Mehta A. D., Herrero J. L., de Cheveigné A., Shamma S., Mesgarani N. (2020) **Cortical encoding of melodic expectations in human temporal cortex** *eLife* **9** https://doi.org/10.7554/eLife.51784

18. Ding N., Melloni L., Yang A., Wang Y., Zhang W., Poeppel D (2017) **Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG)** *Frontiers in Human Neuroscience* **11**

19. Ding N., Melloni L., Zhang H., Tian X., Poeppel D (2016) **Cortical tracking of hierarchical linguistic structures in connected speech** *Nature Neuroscience* **19**:158–164

20. Doelling K. B., Arnal L. H., Ghitza O., Poeppel D (2014) **Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing** *NeuroImage* **85**:761–768

21. Donhauser P. W., Baillet S (2019) **Two Distinct Neural Timescales for Predictive Speech Processing** *Neuron* https://doi.org/10.1016/j.neuron.2019.10.019

22. Elman J. L (1991) **Distributed representations, simple recurrent networks, and grammatical structure** *Machine Learning* **7**:195–225

23. Fadiga L., Craighero L., D'Ausilio A (2009) **Broca's area in language, action, and music** *Annals of the New York Academy of Sciences* **1169**:448–458

24. Fan L. *et al.* (2016) **The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture** *Cerebral Cortex* **26**:3508–3526

25. Fedorenko E., Behr M. K., Kanwisher N (2011) **Functional specificity for high-level linguistic processing in the human brain** *Proceedings of the National Academy of Sciences of the United States of America* **108**:16428–16433

26. Fedorenko E., Blank I. A (2020) **Broca's Area Is Not a Natural Kind** *Trends in Cognitive Sciences* **24**:270–284

27. Flaugnacco E., Lopez L., Terribili C., Montico M., Zoia S., Schön D (2015) **Music Training Increases Phonological Awareness and Reading Skills in Developmental Dyslexia: A Randomized Control Trial** *PloS One* **10**

28. Flinker A., Doyle W. K., Mehta A. D., Devinsky O., Poeppel D (2019) **Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries** *Nature Human Behaviour* **3**:393–405

29. Fontolan L., Morillon B., Liegeois-Chauvel C., Giraud A.-L (2014) **The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex** *Nature Communications* **5**

30. François C., Chobert J., Besson M., Schön D (2013) **Music training for the development of speech segmentation** *Cerebral Cortex* **23**:2038–2043

31. Friederici A. D (2020) **Hierarchy processing in human neurobiology: how specific is it?** *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **375**

32. Fries P. (2015) **Rhythms for cognition: communication through coherence**

33. Giordano B. L., Esposito M., Valente G., Formisano E (2023) **Intermediate acoustic-to-semantic representations link behavioral and neural responses to natural sounds** *Nature Neuroscience* **26**:664–672

34. Giordano B. L., Whiting C., Kriegeskorte N., Kotz S. A., Gross J., Belin P (2021) **The representational dynamics of perceived voice emotions evolve from categories to dimensions** *Nature Human Behaviour* **5**:1203–1213

35. Giraud A.-L., Poeppel D (2012) **Speech Perception from a Neurophysiological Perspective** *The Human Auditory Cortex* :225–260

36. Gramfort A., Luessi M., Larson E., Engemann D. A., Strohmeier D., Brodbeck C., Parkkonen L., Hämäläinen M. S (2014) **MNE software for processing MEG and EEG data** *NeuroImage* **86**:446–460

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

38. Gripari P. (2004) **La sorcière de la rue Mouffetard**

39. Groppe D. M., Bickel S., Dykstra A. R., Wang X., Mégevand P., Mercier M. R., Lado F. A., Mehta A. D., Honey C. J (2017) **iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data** *Journal of Neuroscience Methods* **281**:40–48

40. Groppe D. M., Urbach T. P., Kutas M (2011) **Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review** *Psychophysiology* **48**:1711–1725

41. Gross J., Hoogenboom N., Thut G., Schyns P., Panzeri S., Belin P., Garrod S (2013) **Speech rhythms and multiplexed oscillatory sensory coding in the human brain** *PLoS Biology* **11**

42. Hasson U., Malach R., Heeger D. J (2010) **Reliability of cortical activity during natural stimulation** *Trends in Cognitive Sciences* **14**:40–48

43. Inbar M., Grossman E., Landau A. N (2020) **Sequences of Intonation Units form a ~ 1 Hz rhythm** *Scientific Reports* **10**

44. Ince R. A., Paton A. T., Kay J. W., Schyns P. G (2021) **Bayesian inference of population prevalence** *eLife* **10** https://doi.org/10.7554/eLife.62461

45. Kayser C., Ince R. A. A., Panzeri S (2012) **Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices** *PLoS Computational Biology* **8**

46. Keitel A., Gross J., Kayser C (2017) **Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features** *In PLoS Biol. (Issue* **3** https://doi.org/10.1371/journal.pbio.2004473

47. Kivy P (1959) **Charles Darwin on Music** *Journal of the American Musicological Society* **12**:42–48

48. Koelsch S (2011) **Toward a neural basis of music perception – a review and updated model** *Frontiers in Psychology* **2**

49. Kopell N., Ermentrout G. B., Whittington M. A., Traub R. D (2000) **Gamma rhythms and beta rhythms have different synchronization properties** *Proceedings of the National Academy of Sciences of the United States of America* **97**:1867–1872

50. Kraus N., Chandrasekaran B (2010) **Music training for the development of auditory skills** *Nature Reviews. Neuroscience* **11**:599–605

51. Le Van Quyen M., Staba R., Bragin A., Dickson C., Valderrama M., Fried I., Engel J. (2010) **Large-scale microelectrode recordings of high-frequency gamma oscillations in human cortex during sleep** *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **30**:7770–7782

52. Malik-Moraleda S., Ayyash D., Gallée J., Affourtit J., Hoffmann M., Mineroff Z., Jouravlev O., Fedorenko E (2022) **An investigation across 45 languages and 12 language families reveals a universal language network** *Nature Neuroscience* **25**:1014–1019

53. Martin S., Millán J. D. R., Knight R. T., Pasley B. N (2019) **The use of intracranial recordings to decode human language: Challenges and opportunities** *Brain and Language* **193**:73–83

54. Mas-Herrero E., Marco-Pallares J., Lorenzo-Seva U., Zatorre R. J., Rodriguez-Fornells A (2013) **Individual differences in music reward experiences** *Music Perception* **31**:118–138

55. McCarty M. J., Murphy E., Scherschligt X., Woolnough O., Morse C. W., Snyder K., Mahon B. Z., Tandon N (2023) **Intraoperative cortical localization of music and language reveals signatures of structural complexity in posterior temporal cortex** *iScience* **26**

56. Medina Villalon S., Paz R., Roehri N., Lagarde S., Pizzo F., Colombet B., Bartolomei F., Carron R., Bénar C.-G. (2018) **EpiTools, A software suite for presurgical brain mapping in epilepsy: Intracerebral EEG** *Journal of Neuroscience Methods* **303**:7–15

57. Menon V., Levitin D. J., Smith B. K., Lembke A., Krasnow B. D., Glazer D., Glover G. H., McAdams S (2002) **Neural correlates of timbre change in harmonic sounds** *NeuroImage* **17**:1742–1754

58. Mercier M. R., Bickel S., Megevand P., Groppe D. M., Schroeder C. E., Mehta A. D., Lado F. A (2017) **Evaluation of cortical local field potential diffusion in stereotactic electro-encephalography recordings: A glimpse on white matter signal** *NeuroImage* **147**:219–232

59. Mercier M. R. *et al.* (2022) **Advances in human intracranial electroencephalography research, guidelines and good practices** *NeuroImage* **260**

60. Millet J., Caucheteux C., Orhan P., Boubenec Y., Gramfort A., Dunbar E., Pallier C., King J.-R. (2022) **Toward a realistic model of speech processing in the brain with self-supervised learning**

61. Morillon B., Baillet S (2017) **Motor origin of temporal predictions in auditory attention** *Proceedings of the National Academy of Sciences of the United States of America* **114**:E8913–E8921

62. Nantais K. M., Schellenberg E. G (1999) **The Mozart Effect: An Artifact of Preference** *Psychological Science* **10**:370–373

63. Nichols T. E., Holmes A. P (2002) **Nonparametric permutation tests for functional neuroimaging: a primer with examples** *Human Brain Mapping* **15**:1–25

64. Nieto-Castañón A., Fedorenko E (2012) **Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses** *NeuroImage* **63**:1646–1669

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

66. Norman-Haignere S. V., Feather J., Boebinger D., Brunner P., Ritaccio A., McDermott J. H., Schalk G., Kanwisher N (2020) **Intracranial recordings from human auditory cortex reveal a neural population selective for song** *In bioRxiv* **696161** https://doi.org/10.1101/696161

67. Norman-Haignere S. V., Feather J., Boebinger D., Brunner P., Ritaccio A., McDermott J. H., Schalk G., Kanwisher N (2022) **A neural population selective for song in human auditory cortex** *Current Biology: CB* https://doi.org/10.1016/j.cub.2022.01.069

68. Oganian Y., Chang E. F (2019) **A speech envelope landmark for syllable encoding in human superior temporal gyrus** *In Science Advances (Issue* **11** https://doi.org/10.1126/sciadv.aay6279

69. Oneness (2006) **Reflejos del Sur**

70. Oostenveld R., Fries P., Maris E., Schoffelen J.-M (2011) **FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data** *Computational Intelligence and Neuroscience* **2011**

71. Ossandón T., Vidal J. R., Ciumas C., Jerbi K., Hamamé C. M., Dalal S. S., Bertrand O., Minotti L., Kahane P., Lachaux J.-P (2012) **Efficient "Pop-Out" Visual Search Elicits Sustained Broadband Gamma Activity in the Dorsal Attention Network** *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **32**:3414–3421

72. Penny W. D., Friston K. J., Ashburner J. T., Kiebel S. J., Nichols T. E (2011) **Statistical Parametric Mapping: The Analysis of Functional Brain Images**

73. Pfurtscheller G., Lopes da Silva F. H. (1999) **Event-related EEG/MEG synchronization and desynchronization: basic principles** *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* **110**:1842–1857

74. Proix T. *et al.* (2022) **Imagined speech can be decoded from low– and cross-frequency intracranial EEG features** *Nature Communications* **13**

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

76. Rissman J., Wagner A. D (2012) **Distributed representations in memory: insights from functional brain imaging** *Annual Review of Psychology* **63**:101–128

77. Robert P., Zatorre R., Gupta A., Sein J., Anton J.-L., Belin P., Thoret E., Morillon B (2023) **Auditory hemispheric asymmetry as a specialization for actions and objects** *In bioRxiv* **2023**:4–19 https://doi.org/10.1101/2023.04.19.537361

78. Rousseau J.-J. (2009) **Essay on the Origin of Languages and Writings Related to Music**

79. Safaie M., Chang J. C., Park J., Miller L. E., Dudman J. T., Perich M. G., Gallego J. A. (2023) **Preserved neural dynamics across animals performing similar behaviour** *Nature* **623**:765–771

80. Saleh M., Reimer J., Penn R., Ojakangas C. L., Hatsopoulos N. G (2010) **Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues** *Neuron* **65**:461–471

81. Sankaran N., Leonard M. K., Theunissen F., Chang E. F (2023) **Encoding of melody in the human auditory cortex** *bioRxiv: The Preprint Server for Biology* https://doi.org/10.1101/2023.10.17.562771

82. Schön D., Gordon R., Campagne A., Magne C., Astésano C., Anton J.-L., Besson M (2010) **Similar cerebral networks in language, music and song perception** *NeuroImage* **51**:450–461

83. Schön D., Magne C., Besson M (2004) **The music of speech: music training facilitates pitch processing in both music and language** *Psychophysiology* **41**:341–349

84. Siegel M., Donner T. H., Engel A. K (2012) **Spectral fingerprints of large-scale neuronal interactions** *Nature Reviews. Neuroscience* **13**:121–134

85. Sonkusare S., Breakspear M., Guo C (2019) **Naturalistic Stimuli in Neuroscience: Critically Acclaimed** *Trends in Cognitive Sciences* **23**:699–714

86. Steinkamp S. R. (2019) **pymtrf: Translation of the mtrf-Toolbox for Matlab**

87. Stolk A., Griffin S., van der Meij R., Dewar C., Saez I., Lin J. J., Piantoni G., Schoffelen J.-M., Knight R. T., Oostenveld R. (2018) **Integrated analysis of anatomical and electrophysiological human intracranial data** *Nature Protocols* **13**:1699–1723

88. Theunissen F. E., Sen K., Doupe A. J (2000) **Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds** *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **20**:2315–2331

89. van Kerkoerle T., Self M. W., Dagnino B., Gariel-Mathis M.-A., Poort J., van der Togt C., Roelfsema P. R. (2014) **Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex** *Proceedings of the National Academy of Sciences of the United States of America* **111**:14332–14341

90. Vidal J. R., Freyermuth S., Jerbi K., Hamamé C. M., Ossandon T., Bertrand O., Minotti L., Kahane P., Berthoz A., Lachaux J.-P (2012) **Long-Distance Amplitude Correlations in the High Gamma Band Reveal Segregation and Integration within the Reading Network** *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **32**:6421–6434

91. Virtanen P. *et al.* (2020) **SciPy 1.0: fundamental algorithms for scientific computing in Python** *Nature Methods* **17**:261–272

92. Zatorre R. J., Chen J. L., Penhune V. B (2007) **When the brain plays music: auditory–motor interactions in music perception and production** *Nature Reviews. Neuroscience* **8**:547–558

93. Zatorre R. J., Gandour J. T (2008) **Neural specializations for speech and pitch: moving beyond the dichotomies** *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **363**:1087–1104

94. Zion Golumbic E. M. *et al.* (2013) **Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party."** *Neuron* **77**:980–991

95. Zuk N. J., Murphy J. W., Reilly R. B., Lalor E. C (2021) **Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies** *PLoS Computational Biology* **17**

Editors

Reviewing Editor
Huan Luo
Peking University, Beijing, China

Senior Editor
Andrew King
University of Oxford, Oxford, United Kingdom

Reviewer #1 (Public Review):

Summary:

In this study, the authors examined the extent to which processing of speech and music depends on neural networks that are either specific to a domain or general in nature. They conducted comprehensive intracranial EEG recordings on 18 epilepsy patients as they listened to natural, continuous forms of speech and music. This enabled an exploration of brain activity at both the frequency-specific and network levels across a broad spectrum. Utilizing statistical methods, the researchers classified neural responses to auditory stimuli into categories of shared, preferred, and domain-selective types. It was observed that a significant portion of both focal and network-level brain activity is commonly shared between the processing of speech and music. However, neural responses that are selectively responsive to speech or music are confined to distributed, frequency-specific areas. The authors highlight the crucial role of using natural auditory stimuli in research and the need to explore the extensive spectral characteristics inherent in the processing of speech and music.

Strengths:

The study's strengths include its high-quality sEEG data from a substantial number of patients, covering a majority of brain regions. This extensive cortical coverage grants the authors the ability to address their research questions with high spatial resolution, marking an advantage over previous studies. They performed thorough analyses across the entire cortical coverage and a wide frequency range of neural signals. The primary analyses, including spectral analysis, temporal response function calculation, and connectivity analysis, are presented straightforwardly. These analyses, as well as figures, innovatively display how neural responses, in each frequency band and region/electrode, are 'selective' (according to the authors' definition) to speech or music stimuli. The findings are summarized in a manner that efficiently communicates information to readers. This research offers valuable insights into the cortical selectivity of speech and music processing, making it a noteworthy reference for those interested in this field. Overall, this research offers a valuable dataset and carries out extensive yet clear analyses, amounting to an impressive empirical investigation into the cortical selectivity of speech and music. It is recommended for readers who are keen on understanding the nuances of selectivity and generality in the processing of speech and music to refer to this study's data and its summarized findings.

Weaknesses:

(1) The study employed longer speech and music stimuli, thereby promising improved ecological validity as compared to prior research, a point emphasized by the authors. However, it failed to differentiate between neural responses to the diverse content or local structures within speech and music. The authors considered the potential limitation of treating these extensive speech and music stimuli as stationary signals, neglecting their complex musical or linguistic structural details and temporal variations across local structures such as sentences and phrases. This balanced perspective offered by the authors aids readers in better understanding the context of the study and highlights potential areas for expansion and further considerations.

(2) In contrast to previous studies that employed short stimulus segments along with various control stimuli to ensure that observed selectivity for speech or music was not merely due to low-level acoustic properties, this study used longer, ecological stimuli. However, the control stimuli used in this study, such as tone or syllable sequences, do not align with the low-level acoustic properties of the speech and music stimuli. This mismatch raises concerns that the differences or selectivity between speech and music observed in this study might be attributable to these basic acoustic characteristics rather than to more complex processing factors specific to speech or music. However, this should not deter readers from recognizing the study's strengths, namely, the use of iEEG recordings that offer high spatial resolution and extensive cortical coverage.

(3) The concept of selectivity - shared, preferred, and domain-selective - may not present sufficient theoretical accuracy. It is appreciated that the authors put effort into clearly defining their operational measurement on 'selectivity'. Later, the authors further mentioned the specific indication of their analyses. However, the authors' categorization of neural sites/regions as shared, preferred, or domain-selective regarding speech and music processing essentially resembles a traditional ANOVA test with posthoc analysis. While this categorization gives meaningful context to the results, the mere presence of significant differences among control stimuli, a segment of speech, and a piece of music does not present a strong case that a region is specifically selective to a type of stimulus like speech. The narrative of the manuscript could potentially lead to an overgeneralized interpretation of their findings as being broadly applicable to speech or music, if a reader does not delve into the details.

(4) The authors' approach, akin to mapping a 'receptive field' by correlating stimulus properties with neural responses to ascertain functional selectivity for speech and music, presents potential issues. If cortical regions exhibit heightened responses to one type of stimulus over another, it doesn't automatically imply selectivity or preference for that stimulus. The explanation could lie in functional aspects, such as a region's sensitivity to temporal units of a specific duration, be it music, speech, or even movie segments, and its role in chunking such units (e.g., around 500 ms), which might be more prevalent in music than in speech, or vice versa in the current study. This study does not delve into the functional mechanisms of how speech and music are processed across different musical or linguistic hierarchical levels but merely demonstrates differences in neural responses to various stimuli over a 10-minute span.

https://doi.org/10.7554/eLife.94509.2.sa2

Reviewer #2 (Public Review):

Summary:

The study investigates whether speech and music processing involve specific or shared brain networks. Using intracranial EEG recordings from 18 epilepsy patients, it examines neural

responses to speech and music. The authors found that most neural activity is shared between speech and music processing, without specific regional brain selectivity. Furthermore, domain-selective responses to speech or music are limited to frequency-specific coherent oscillations. The findings challenge the notion of anatomically distinct regions for different cognitive functions in the auditory process.

Strengths:

(1) This study uses a relatively large corpus of intracranial EEG data, which provides high spatiotemporal resolution neural recordings, allowing for more precise and dynamic analysis of brain responses. The use of continuous speech and music enhances ecological validity compared to artificial or segmented stimuli.

(2) This study uses multiple frequency bands in addition to just high-frequency activity (HFA), which has been the focus of many existing studies in the literature. This allows for a more comprehensive analysis of neural processing across the entire spectrum. The heterogeneity across different frequency bands also indicates that different frequency components of the neural activity may reflect different underlying neural computations.

(3) This study also adds empirical evidence towards distributed representation versus domain-specificity. It challenges the traditional view of highly specialized, anatomically distinct regions for different cognitive functions. Instead, the study suggests a more integrated and overlapping neural network for processing complex stimuli like speech and music.

Weaknesses:

While this study is overall convincing, there are still some weaknesses in the methods and analyses that limit the implication of the work.

The study's main approach, focusing primarily on the grand comparison of response amplitudes between speech and music, may overlook intricate details in neural coding. Speech and music are not entirely orthogonal with each other at different levels of analysis: at the high-level abstraction, these are two different categories of cognitive processes; at the low-level acoustics, they overlap a lot; at intermediate levels, they may also share similar features. For example, the study doesn't adequately address whether purely melodic elements in music correlate with intonations in speech at the neural level. A more granular analysis, dissecting stimuli into distinct features like pitch, phonetics, timbre, and linguistic elements, could unveil more nuanced shared, and unique neural processes between speech and music. Prior research indicates potential overlap in neural coding for certain intermediate features in speech and music (Sankaran et al. 2023), suggesting that a simple averaged response comparison might not fully capture the complexity of neural encoding. Further delineation of phonetic, melodic, linguistic, and other coding, along with an analysis of how different informational aspects (phonetic, linguistic, melodic, etc) are represented in shared neural activities, could enhance our understanding of these processes and strengthen the study's conclusions.

While classifying electrodes into 3 categories provides valuable insights, it may not fully capture the complexity of the neural response distribution to speech and music. A more nuanced and continuous approach could reveal subtler gradations in neural response, rather than imposing categorical boundaries. This could be done by computing continuous metrics, like unique variances explained by each category or by each acoustic feature, etc. Incorporating such a continuum could enhance our understanding of the neural representation of speech and music, providing a more detailed and comprehensive picture of cortical processing. This goes back to my first comment that the selected set of stimuli may not fully exploit the entire space of speech and music, and there are possible exemplars that violate the preference map here. For example, this study only considered a specific set of

multi-instrumental music, it is not clear to me if other types of music would result in different response profiles in individual channels. It is also not clear if a foreign language that the listeners cannot comprehend would evoke similar response profiles. On the contrary, breaking down into the neural coding of more fundamental feature representations that constitute speech and music, and analyzing the unique contribution of each feature would give a more comprehensive understanding.

The paper's emphasis on shared and overlapping neural activity, as observed through sEEG electrodes, provides valuable insights. It is probably true that domain-specificity for speech and music does not exist at such a macro scale. However, it's important to consider that each electrode records from a large neuronal population, encompassing thousands of neurons. This broad recording scope might mask more granular, non-overlapping feature representations at the single neuron level. Thus, while the study suggests shared neural underpinnings for speech and music perception at a macroscopic level, it cannot definitively rule out the possibility of distinct, non-overlapping neural representations at the microscale of local neuronal circuits for features that are distinctly associated with speech and music. This distinction is crucial for fully understanding the neural mechanisms underlying speech and music perception that merit future endeavors with more advanced large-scale neuronal recordings.

https://doi.org/10.7554/eLife.94509.2.sa1

Reviewer #3 (Public Review):

Summary:

Te Rietmolen et al., investigated the selectivity of cortical responses to speech and music stimuli using neurosurgical stereo EEG in humans. The authors address two basic questions: 1. Are speech and music responses localized in the brain or distributed; 2. Are these responses selective and domain specific or rather domain general and shared. To investigate this, the study proposes a nomenclature of shared responses (speech and music responses are not significantly different), domain selective (one domain is significant from baseline and the other is not), domain preferred (both are significant from baseline but one is larger than the other and significantly different from each other). The authors employ this framework using neural responses across the spectrum (rather than focusing on high gamma), providing evidence for a low level of selectivity across spectral signatures. To investigate the nature of the underlying representations they use encoding models to predict neural responses (low and high frequency) given a feature space of the stimulus envelope or peak rate (by time delay) and find stronger encoding for both in the low frequency neural responses. The top encoding electrodes are used as seeds for a pair-wise connectivity (coherence) in order to repeat the shared/selective/preferred analysis across the spectra, suggesting low selectivity. Spectral power and connectivity are also analyzed on the level of regional patient population to rule out (and depict) any effects driven by a select few patients. Across analyses the authors consistently show a paucity of domain selective responses and when evident these selective responses were not represented across the entire cortical region. The authors argue that speech and music mostly rely on shared neural resources.

Strengths:

I found this manuscript to be rigorous providing compelling and clear evidence towards shared neural signatures for speech and music. The use of intracranial recordings provides an important spatial and temporal resolution that lends itself to the power, connectivity and encoding analyses. The statistics and methods employed are rigorous and reliable, estimated based on permutation approaches and cross-validation/regularization was employed and reported properly. The analysis of measures across the entire spectra in both power,

coherence and encoding models provides a comprehensive view of responses that no doubt will benefit the community as an invaluable resource. Analysis on the level of patient population (feasible with their high N) per region also supports the generalizability of the conclusions across a relatively large cohort of patients. Last but not least, I believe the framework of selective, preferred, and shared is a welcome lens through which to investigate cortical function.

Weaknesses:

I did not find methodological weaknesses in the current version of the manuscript. I do believe that it is important to highlight that the data is limited to passively listening to naturalistic speech and music. The speech and music stimuli are not completely controlled with varying key acoustic features (inherent to the different domains). Overall, I found the differences in stimulus and lack of attentional controls (passive listening) to be minor weaknesses that would not dramatically change the results or conclusions.

https://doi.org/10.7554/eLife.94509.2.sa0

Author response:

The following is the authors' response to the original reviews.

We have specifically addressed the points of uncertainty highlighted in eLife's editorial assessment, which concerned the lack of low-level acoustics control, limitations of experimental design, and in-depth analysis. Regarding "the lack of low-level acoustics control, limitations of experimental design", in response to Reviewer #1, we clarify that our study aimed to provide a broad perspective —which includes both auditory and higher-level processes— on the similarities and distinctions in processing natural speech and music within an ecological context. Regarding "the lack of in-depth analysis", in response to Reviewer #1 and #2, we have clarified that while model-based analyzes are valuable, they pose fundamental challenges when comparing speech and music. Non-acoustic features inherently differ between speech and music (such as phonemes and pitch), making direct comparisons reliant on somewhat arbitrary choices. Our approach mitigates this challenge by analyzing the entire neural signal, thereby avoiding potential pitfalls associated with encoding models of non-comparable features. Finally, we provide some additional analyzes suggested by the Reviewers.

We sincerely appreciate your thoughtful and thorough consideration throughout the review process.

> ***eLife assessment***
>
> *This study presents valuable intracranial findings on how two important types of natural auditory stimuli - speech and music - are processed in the human brain, and demonstrates that speech and music largely share network-level brain activities, thus challenging the domain-specific processing view. The evidence supporting the claims of the authors is solid but somewhat incomplete since although the data analysis is thorough, the results are robust and the stimuli have ecological validity, important considerations such as low-level acoustics control, limitations of experimental design, and in-depth analysis, are lacking. The work will be of broad interest to speech and music researchers as well as cognitive scientists in general.*
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*

In this study, the authors examined the extent to which the processing of speech and music depends on neural networks that are either specific to a domain or general in nature. They conducted comprehensive intracranial EEG recordings on 18 epilepsy patients as they listened to natural, continuous forms of speech and music. This enabled an exploration of brain activity at both the frequency-specific and network levels across a broad spectrum. Utilizing statistical methods, the researchers classified neural responses to auditory stimuli into categories of shared, preferred, and domain-selective types. It was observed that a significant portion of both focal and network-level brain activity is commonly shared between the processing of speech and music. However, neural responses that are selectively responsive to speech or music are confined to distributed, frequency-specific areas. The authors highlight the crucial role of using natural auditory stimuli in research and the need to explore the extensive spectral characteristics inherent in the processing of speech and music.

Strengths:

The study's strengths include its high-quality sEEG data from a substantial number of patients, covering a majority of brain regions. This extensive cortical coverage grants the authors the ability to address their research questions with high spatial resolution, marking an advantage over previous studies. They performed thorough analyses across the entire cortical coverage and a wide frequency range of neural signals. The primary analyses, including spectral analysis, temporal response function calculation, and connectivity analysis, are presented straightforwardly. These analyses, as well as figures, innovatively display how neural responses, in each frequency band and region/electrode, are 'selective' (according to the authors' definition) to speech or music stimuli. The findings are summarized in a manner that efficiently communicates information to readers. This research offers valuable insights into the cortical selectivity of speech and music processing, making it a noteworthy reference for those interested in this field. Overall, this research offers a valuable dataset and carries out extensive yet clear analyses, amounting to an impressive empirical investigation into the cortical selectivity of speech and music. It is recommended for readers who are keen on understanding the nuances of selectivity and generality in the processing of speech and music to refer to this study's data and its summarized findings.

Weaknesses:

The weakness of this study, in my view, lies in its experimental design and reasoning:

(1) Despite using longer stimuli, the study does not significantly enhance ecological validity compared to previous research. The analyses treat these long speech and music stimuli as stationary signals, overlooking their intricate musical or linguistic structural details and temporal variation across local structures like sentences and phrases. In previous studies, short, less ecological segments of music were used, maintaining consistency in content and structure. However, this study, despite employing longer stimuli, does not distinguish between neural responses to the varied contents or structures within speech and music. Understanding the implications of long-term analyses, such as spectral and connectivity analyses over extended periods of around 10 minutes, becomes challenging when they do not account for the variable, sometimes quasi-periodical or even non-periodical, elements present in natural speech and music. When contrasting this study with prior research and highlighting its advantages, a more balanced perspective would have been beneficial in the manuscript.

Regarding ecological validity, we respectfully hold a differing perspective from the reviewer. In our view, a one-second music stimulus lacks ecological validity, as real-world music always extends much beyond such a brief duration. While we acknowledge the trade-off in selecting

longer stimuli, limiting the diversity of musical styles, we maintain that only long stimuli afford participants an authentic musical listening experience. Conversely, shorter stimuli may lead participants to merely "skip through" musical excerpts rather than engage in genuine listening.

Regarding the critique that we "did not distinguish between neural responses to the varied contents or structures within speech and music," we partly concur. Our TRF (temporal response function) analyzes incorporate acoustic content, particularly the acoustic envelope, thereby addressing this concern to some extent. However, it is accurate to note that we did not model non-acoustic features. In acknowledging this limitation, we would like to share an additional thought with the reviewer regarding model comparison for speech and music. Specifically, comparing results from a phonetic (or syntactic) model of speech to a pitch-melodic (or harmonic) model for music is not straightforward, as these models operate on fundamentally different dimensions. In other words, while assuming equivalence between phonemes and pitches may be a reasonable assumption, it in essence relies on a somewhat arbitrary choice. Consequently, comparing and interpreting neuronal population coding for one or the other model remains problematic. In summary, because the models for speech and music are different (except for acoustic models), direct comparison is challenging, although still commendable and of interest.

Finally, we did take into account the reviewer's remark and did our best to give a more balanced perspective of our approach and previous studies in the discussion.

"While listening to natural speech and music rests on cognitively relevant neural processes, our analytical approach, extending over a rather long period of time, does not allow to directly isolate specific brain operations. Computational models -which can be as diverse as acoustic (Chi et al., 2005), cognitive (Giordano et al., 2021), information-theoretic (Di Liberto et al., 2020), or self-supervised neural network (Donhauser & Baillet, 2019 ; Millet et al., 2022) models- are hence necessary to further our understanding of the type of computations performed by our reported frequency-specific distributed networks. Moreover, incorporating models accounting for musical and linguistic structure can help us avoid misattributing differences between speech and music driven by unmatched sensitivity factors (e.g., arousal, emotion, or attention) as inherent speech or music selectivity (Mas-Herrero et al., 2013; Nantais & Schellenberg, 1999)."

> *(2) In contrast to previous studies that employed short stimulus segments along with various control stimuli to ensure that observed selectivity for speech or music was not merely due to low-level acoustic properties, this study used longer, ecological stimuli. However, the control stimuli used in this study, such as tone or syllable sequences, do not align with the low-level acoustic properties of the speech and music stimuli. This mismatch raises concerns that the differences or selectivity between speech and music observed in this study might be attributable to these basic acoustic characteristics rather than to more complex processing factors specific to speech or music.*

We acknowledge the reviewer's concern. Indeed, speech and music differ on various levels, including acoustic and cognitive aspects, and our analyzes do not explicitly distinguish them. The aim of this study was to provide an overview of the similarities and differences between natural speech and music processing, in ecological context. Future work is needed to explore further the different hierarchical levels or networks composing such listening experiences. Of note, however, we report whole-brain results with high spatial resolution (thanks to iEEG recordings), enabling the distinction between auditory, superior temporal gyrus (STG), and higher-level responses. Our findings clearly highlight that both auditory and higher-level regions predominantly exhibit shared responses, challenging the interpretation that our results can be attributed solely to differences in 'basic acoustic characteristics'.

We have now more clearly pointed out this reasoning in the results section:

“The spatial distribution of the spectrally-resolved responses corresponds to the network typically involved in speech and music perception. This network encompasses both ventral and dorsal auditory pathways, extending well beyond the auditory cortex and, hence, beyond auditory processing that may result from differences in the acoustic properties of our baseline and experimental stimuli.“

> *(3) The concept of selectivity - shared, preferred, and domain-selective - increases the risks of potentially overgeneralized interpretations and theoretical inaccuracies. The authors' categorization of neural sites/regions as shared, preferred, or domain-selective regarding speech and music processing essentially resembles a traditional ANOVA test with post hoc analysis. While this categorization gives meaningful context to the results, the mere presence of significant differences among control stimuli, a segment of speech, and a piece of music does not necessarily imply that a region is specifically selective to a type of stimulus like speech. The manuscript's narrative might lead to an overgeneralized interpretation that their findings apply broadly to speech or music. However, identifying differences in neural responses to a few sets of specific stimuli in one brain region does not robustly support such a generalization. This is because speech and music are inherently diverse, and specificity often relates more to the underlying functions than to observed neural responses to a limited number of examples of a stimulus type. See the next point.*

Exactly! Here, we present a precise operational definition of these terms, implemented with clear and rigorous statistical methods. It is important to note that in many cognitive neuroscience studies, the term "selective" is often used without a clear definition. By establishing operational definitions, we identified three distinct categories based on statistical testing of differences from baseline and between conditions. This approach provides a framework for more accurate interpretation of experimental findings, as now better outlined in the introduction:

“Finally, we suggest that terms should be operationally defined based on statistical tests, which results in a clear distinction between shared, selective, and preferred activity. That is, be A and B two investigated cognitive functions, “shared” would be a neural population that (compared to a baseline) significantly and equally contributes to the processing of both A and B; “selective” would be a neural population that exclusively contributes to the processing of A or B (e.g. significant for A but not B); and “preferred” would be a neural population that significantly contributes to the processing of both A and B, but more prominently for A or B (Figure 1A).”

Regarding the risk of over-generalization, we want to clarify that our manuscript does not claim that a specific region or frequency band is selective to speech or music. As indeed we focus on testing excerpts of speech and music, we employ the reverse logical reasoning: "if 10 minutes of instrumental music activates a region traditionally associated with speech selectivity, we can conclude that this region is NOT speech-selective." Our conclusions revolve around the absence of selectivity rather than the presence of selective areas or frequency bands. In essence, "one counterexample is enough to disprove a theory." We now further elaborated on this point in the discussion section:

“In this context, in the current study we did not observe a single anatomical region for which speech-selectivity was present, in any of our analyzes. In other words, 10 minutes of instrumental music was enough to activate cortical regions classically labeled as speech (or language) -selective. On the contrary, we report spatially distributed and frequency-specific patterns of shared, preferred, or selective neural responses and connectivity fingerprints.

This indicates that domain-selective brain regions should be considered as a set of functionally homogeneous but spatially distributed voxels, instead of anatomical landmarks."

> *(4) The authors' approach, akin to mapping a 'receptive field' by correlating stimulus properties with neural responses to ascertain functional selectivity for speech and music, presents issues. For instance, in the cochlea, different stimuli activate different parts of the basilar membrane due to the distinct spectral contents of speech and music, with each part being selective to certain frequencies. However, this phenomenon reflects the frequency selectivity of the basilar membrane - an important function, not an inherent selectivity for speech or music. Similarly, if cortical regions exhibit heightened responses to one type of stimulus over another, it doesn't automatically imply selectivity or preference for that stimulus. The explanation could lie in functional aspects, such as a region's sensitivity to temporal units of a specific duration, be it music, speech, or even movie segments, and its role in chunking such units (e.g., around 500 ms), which might be more prevalent in music than in speech, or vice versa in the current study. This study does not delve into the functional mechanisms of how speech and music are processed across different musical or linguistic hierarchical levels but merely demonstrates differences in neural responses to various stimuli over a 10-minute span.*

We completely agree with the last statement, as our primary goal was not to investigate the functional mechanisms underlying speech and music processing. However, the finding of a substantial portion of the cortical network as being shared between the two domains constrains our understanding of the underlying common operations. Regarding the initial part of the comment, we would like to clarify that in the framework we propose, if cortical regions show heightened responses to one type of stimulus over another, this falls into the 'preferred' category. The 'selective' (exclusive) category, on the other hand, would require that the region be unresponsive to one of the two stimuli.

> ***Reviewer #2 (Public Review):***
>
> *Summary:*
>
> *The study investigates whether speech and music processing involve specific or shared brain networks. Using intracranial EEG recordings from 18 epilepsy patients, it examines neural responses to speech and music. The authors found that most neural activity is shared between speech and music processing, without specific regional brain selectivity. Furthermore, domain-selective responses to speech or music are limited to frequency-specific coherent oscillations. The findings challenge the notion of anatomically distinct regions for different cognitive functions in the auditory process.*
>
> *Strengths:*
>
> *(1) This study uses a relatively large corpus of intracranial EEG data, which provides high spatiotemporal resolution neural recordings, allowing for more precise and dynamic analysis of brain responses. The use of continuous speech and music enhances ecological validity compared to artificial or segmented stimuli.*
>
> *(2) This study uses multiple frequency bands in addition to just high-frequency activity (HFA), which has been the focus of many existing studies in the literature. This allows for a more comprehensive analysis of neural processing across the entire spectrum. The heterogeneity across different frequency bands also indicates that different frequency components of the neural activity may reflect different underlying neural computations.*
>
> *(3) This study also adds empirical evidence towards distributed representation versus domain-specificity. It challenges the traditional view of highly specialized, anatomically distinct regions for different cognitive functions. Instead, the study suggests a more*

> *integrated and overlapping neural network for processing complex stimuli like speech and music.*
>
> *Weaknesses:*
>
> *While this study is overall convincing, there are still some weaknesses in the methods and analyses that limit the implication of the work.*
>
> *The study's main approach, focusing primarily on the grand comparison of response amplitudes between speech and music, may overlook intricate details in neural coding. Speech and music are not entirely orthogonal with each other at different levels of analysis: at the high-level abstraction, these are two different categories of cognitive processes; at the low-level acoustics, they overlap a lot; at intermediate levels, they may also share similar features. The selected musical stimuli, incorporating both vocals and multiple instrumental sounds, raise questions about the specificity of neural activation. For instance, it's unclear if the vocal elements in music and speech engage identical neural circuits. Additionally, the study doesn't adequately address whether purely melodic elements in music correlate with intonations in speech at a neural level. A more granular analysis, dissecting stimuli into distinct features like pitch, phonetics, timbre, and linguistic elements, could unveil more nuanced shared, and unique neural processes between speech and music. Prior research indicates potential overlap in neural coding for certain intermediate features in speech and music (Sankaran et al. 2023), suggesting that a simple averaged response comparison might not fully capture the complexity of neural encoding. Further delineation of phonetic, melodic, linguistic, and other coding, along with an analysis of how different informational aspects (phonetic, linguistic, melodic, etc) are represented in shared neural activities, could enhance our understanding of these processes and strengthen the study's conclusions.*

We appreciate the reviewer's acknowledgment that delving into the intricate details of neural coding of speech and music was beyond the scope of this work. To address some of the more precise issues raised, we have clarified in the manuscript that our musical stimuli do not contain vocals and are purely instrumental. We apologize if this was not clear initially.

“In the main experimental session, patients passively listened to ~10 minutes of storytelling (Gripari, 2004); 577 secs, La sorcière de la rue Mouffetard, (Gripari, 2004) and ~10 minutes of instrumental music (580 secs, Reflejos del Sur, (Oneness, 2006) separated by 3 minutes of rest.”

Furthermore, we now acknowledge the importance of modeling melodic, phonetic, or linguistic features in the discussion, and we have referenced the work of Sankaran et al. (2024) and McCarty et al. (2023) in this regard. However, we would like to share an additional thought with the reviewer regarding model comparison for speech and music. Specifically, comparing results from a phonetic (or syntactic) model of speech to a pitch-melodic (or harmonic) model for music is not straightforward, as these models operate on fundamentally different dimensions. In other words, while assuming equivalence between phonemes and pitches may be a reasonable assumption, it in essence relies on a somewhat arbitrary choice. Consequently, comparing and interpreting neuronal population coding for one or the other model remains problematic. In summary, because the models for speech and music are different (except for acoustic models), direct comparison is challenging, although still commendable and of interest.

“These selective responses, not visible in primary cortical regions, seem independent of both low-level acoustic features and higher-order linguistic meaning (Norman-Haignere et al., 2015), and could subtend intermediate representations (Giordano et al., 2023) such as domain-dependent predictions (McCarty et al., 2023; Sankaran et al., 2023).”

References:

McCarty, M. J., Murphy, E., Scherschligt, X., Woolnough, O., Morse, C. W., Snyder, K., Mahon, B. Z., & Tandon, N. (2023). Intraoperative cortical localization of music and language reveals signatures of structural complexity in posterior temporal cortex. *iScience*, *26(7)*, 107223.

Sankaran, N., Leonard, M. K., Theunissen, F., & Chang, E. F. (2023). Encoding of melody in the human auditory cortex. *bioRxiv*. https://doi.org/10.1101/2023.10.17.562771

> *The paper's emphasis on shared and overlapping neural activity, as observed through sEEG electrodes, provides valuable insights. It is probably true that domain-specificity for speech and music does not exist at such a macro scale. However, it's important to consider that each electrode records from a large neuronal population, encompassing thousands of neurons. This broad recording scope might mask more granular, non-overlapping feature representations at the single neuron level. Thus, while the study suggests shared neural underpinnings for speech and music perception at a macroscopic level, it cannot definitively rule out the possibility of distinct, non-overlapping neural representations at the microscale of local neuronal circuits for features that are distinctly associated with speech and music. This distinction is crucial for fully understanding the neural mechanisms underlying speech and music perception that merit future endeavors with more advanced large-scale neuronal recordings.*

We appreciate the reviewer's concern, but we do not view this as a weakness for our study's purpose. Every method inherently has limitations, and intracranial recordings currently offer the best possible spatial specificity and temporal resolution for studying the human brain. Studying cell assemblies thoroughly in humans is ethically challenging, and examining speech and music in non-human primates or rats raises questions about cross-species analogy. Therefore, despite its limitations, we believe intracranial recording remains the best option for addressing these questions in humans.

Regarding the granularity of neural representation, while understanding how computations occur in the central nervous system is crucial, we question whether the single neuron scale provides the most informative insights. The single neuron approach seem more versatile (e.g., in term of cell type or layer affiliation) than the local circuitry they contribute to, which appears to be the brain's building blocks (e.g., like the laminar organization; see Mendoza-Halliday et al.,2024). Additionally, the population dynamics of these functional modules appear crucial for cognition and behavior (Safaie et al. 2023; Buzsáki and Vöröslakos, 2023). Therefore, we emphasize the need for multi-scale research, as we believe that a variety of approaches will complement each other's weaknesses when taken individually. We clarified this in the introduction:

“This approach rests on the idea that the canonical computations that underlie cognition and behavior are anchored in population dynamics of interacting functional modules (Safaie et al. 2023; Buzsáki and Vöröslakos, 2023) and bound to spectral fingerprints consisting of network- and frequency-specific coherent oscillations (Siegel et al., 2012).”

Importantly, we focus on the macro-scale and conclude that, at the anatomical region level, no speech or music selectivity can be observed during natural stimulation. This is stated in the discussion, as follow:

“In this context, in the current study we did not observe a single anatomical region for which speech-selectivity was present, in any of our analyses. In other words, 10 minutes of instrumental music was enough to activate cortical regions classically labeled as speech (or language) -selective. On the contrary, we report spatially distributed and frequency-specific patterns of shared, preferred, or selective neural responses and connectivity fingerprints.

This indicates that domain-selective brain regions should be considered as a set of functionally homogeneous but spatially distributed voxels, instead of anatomical landmarks."

References :

Mendoza-Halliday, D., Major, A.J., Lee, N. *et al.* A ubiquitous spectrolaminar motif of local field potential power across the primate cortex. *Nat Neurosci* (2024).

Safaie, M., Chang, J.C., Park, J. *et al.* Preserved neural dynamics across animals performing similar behaviour. *Nature* **623**, 765–771 (2023).

Buzsáki, G., & Vöröslakos, M. (2023). Brain rhythms have come of age. *Neuron*, *111*(7), 922-926.

> *While classifying electrodes into 3 categories provides valuable insights, it may not fully capture the complexity of the neural response distribution to speech and music. A more nuanced and continuous approach could reveal subtler gradations in neural response, rather than imposing categorical boundaries. This could be done by computing continuous metrics, like unique variances explained by each category, or ratio-based statistics, etc. Incorporating such a continuum could enhance our understanding of the neural representation of speech and music, providing a more detailed and comprehensive picture of cortical processing.*

To clarify, the metrics we are investigating (coherence, power, linear correlations) are continuous. Additionally, we conduct a comprehensive statistical analysis of these results. The statistical testing, which includes assessing differences from baseline and between the speech and music conditions using a statistical threshold, yields three categories. Of note, ratio-based statistics (a continuous metric) are provided in Figures S9 and S10 (Figures S8 and S9 in the original version of the manuscript).

> ***Reviewer #3 (Public Review):***
>
> *Summary:*
>
> *Te Rietmolen et al., investigated the selectivity of cortical responses to speech and music stimuli using neurosurgical stereo EEG in humans. The authors address two basic questions: 1. Are speech and music responses localized in the brain or distributed; 2. Are these responses selective and domain-specific or rather domain-general and shared? To investigate this, the study proposes a nomenclature of shared responses (speech and music responses are not significantly different), domain selective (one domain is significant from baseline and the other is not), domain preferred (both are significant from baseline but one is larger than the other and significantly different from each other). The authors employ this framework using neural responses across the spectrum (rather than focusing on high gamma), providing evidence for a low level of selectivity across spectral signatures. To investigate the nature of the underlying representations they use encoding models to predict neural responses (low and high frequency) given a feature space of the stimulus envelope or peak rate (by time delay) and find stronger encoding for both in the low-frequency neural responses. The top encoding electrodes are used as seeds for a pair-wise connectivity (coherence) in order to repeat the shared/selective/preferred analysis across the spectra, suggesting low selectivity. Spectral power and connectivity are also analyzed on the level of the regional patient population to rule out (and depict) any effects driven by a select few patients. Across analyses the authors consistently show a paucity of domain selective responses and when evident these selective responses were not represented across the entire cortical region. The authors argue that speech and music mostly rely on shared neural resources.*
>
> *Strengths:*

> *I found this manuscript to be rigorous providing compelling and clear evidence of shared neural signatures for speech and music. The use of intracranial recordings provides an important spatial and temporal resolution that lends itself to the power, connectivity, and encoding analyses. The statistics and methods employed are rigorous and reliable, estimated based on permutation approaches, and cross-validation/regularization was employed and reported properly. The analysis of measures across the entire spectra in both power, coherence, and encoding models provides a comprehensive view of responses that no doubt will benefit the community as an invaluable resource. Analysis of the level of patient population (feasible with their high N) per region also supports the generalizability of the conclusions across a relatively large cohort of patients. Last but not least, I believe the framework of selective, preferred, and shared is a welcome lens through which to investigate cortical function.*
>
> *Weaknesses:*
>
> *I did not find methodological weaknesses in the current version of the manuscript. I do believe that it is important to highlight that the data is limited to passively listening to naturalistic speech and music. The speech and music stimuli are not completely controlled with varying key acoustic features (inherent to the different domains). Overall, I found the differences in stimulus and lack of attentional controls (passive listening) to be minor weaknesses that would not dramatically change the results or conclusions.*

Thank you for this positive review of our work. We added these points as limitations and future directions in the discussion section:

“Finally, in adopting here a comparative approach of speech and music – the two main auditory domains of human cognition – we only investigated one type of speech and of music also using a passive listening task. Future work is needed to investigate for instance whether different sentences or melodies activate the same selective frequency-specific distributed networks and to what extent these results are related to the passive listening context compared to a more active and natural context (e.g. conversation).”

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(1) The concepts of activation and deactivation within the study's context of selectivity are not straightforward to comprehend. It would be beneficial for the authors to provide more detailed explanations of how these phenomena relate to the selectivity of neural responses to speech and music. Such elaboration would aid readers in better understanding the nuances of how certain brain regions are selectively activated or deactivated in response to different auditory stimuli.*

The reviewer is right that the reported results are quite complex to interpret. The concepts of activation and deactivation are generally complex to comprehend as they are in part defined by an approach (e.g., method and/or metric) and the scale of observation (Pfurtscheller et al., 1999). The power (or the magnitude) of time-frequency estimate is by definition a positive value. Deactivation (or desynchronization) is therefore related to the comparison used (e.g., baseline, control, condition). This is further complexified by the scale of the measurement, for instance, when it comes to a simple limb movement, some brain areas in sensory motor cortex are going to be activated, yet this phenomenon is accompanied at a finer scale by some desynchonization of the mu-activity, and such desynchronization is a relative measure (e.g., before/after motor movement). At a broader scale it is not rare to see some form of balance between brain networks, some being ‘inhibited’ to let some others be activated like the default mode network versus sensory-motor networks. In our case, when estimating selective

responses, it is the strength of the signal that matters. The type of selectivity is then defined by the sign/direction of the comparison/subtraction. We now provide additional details about the sign of selectivity between domains and frequencies in the Methods and Results section:

Methods:

“In order to explore the full range of possible selective, preferred, or shared responses, we considered both responses greater and smaller than the baseline. Indeed, as neural populations can synchronize or desynchronize in response to sensory stimulation, we estimated these categories separately for significant activations and significant deactivations compared to baseline.”

Results:

“We classified, for each canonical frequency band, each channel into one of the categories mentioned above, i.e. shared, selective, or preferred (Figure 1A), by examining whether speech and/or music differ from baseline and whether they differ from each other. We also considered both activations and deactivations, compared to baseline, as both index a modulation of neural population activity, and have been linked with cognitive processes (Pfurtscheller & Lopes da Silva, 1999; Proix et al., 2022). However, because our aim was not to interpret specific increase or decrease with respect to the baseline, we here simply consider significant deviations from the baseline. In other words, when estimating selectivity, it is the strength of the response that matters, not its direction (activation, deactivation).”

“Both domains displayed a comparable percentage of selective responses across frequency bands (Figure 4, first values of each plot). When considering separately activation (Figure 2) and deactivation (Figure 3) responses, speech and music showed complementary patterns: for low frequencies (<15 Hz) speech selective (and preferred) responses were mostly deactivations and music responses activations compared to baseline, and this pattern reversed for high frequencies (>15 Hz).”

References :

J.P. Lachaux, J. Jung, N. Mainy, J.C. Dreher, O. Bertrand, M. Baciu, L. Minotti, D. Hoffmann, P. Kahane,Silence Is Golden: Transient Neural Deactivation in the Prefrontal Cortex during Attentive Reading, *Cerebral Cortex*, Volume 18, Issue 2, February 2008, Pages 443–450

Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical neurophysiology, 110(11), 1842-1857

> *(2) The manuscript doesn't easily provide information about the control conditions, yet the conclusion significantly depends on these conditions as a baseline. It would be beneficial if the authors could clarify this information for readers earlier and discuss how their choice of control stimuli influences their conclusions.*

We added information in the Results section about the baseline conditions:

“[...] with respect to two baseline conditions, in which patients passively listened to more basic auditory stimuli: one in which patients passively listened to pure tones (each 30 ms in duration), the other in which patients passively listened to isolated syllables (/ba/ or /pa/, see Methods).”

Of note, while the choice of different ‘basic auditory stimuli’ as baseline can change the reported results in regions involved in low-level acoustical analyzes (auditory cortex), it will have no impact on the results observed in higher-level regions, which predominantly also exhibit shared responses. We have now more clearly pointed out this reasoning in the results section:

"The spatial distribution of the spectrally-resolved responses corresponds to the network typically involved in speech and music perception. This network encompasses both ventral and dorsal auditory pathways, extending well beyond the auditory cortex and, hence, beyond auditory processing that may result from differences in the acoustic properties of our baseline and experimental stimuli."

> *(3) The spectral analyses section doesn't clearly explain how the authors performed multiwise correction. The authors' selectivity categorization appears similar to ANOVAs with posthoc tests, implying the need for certain corrections in the p values or categorization. Could the authors clarify this aspect?*

We apologize that this was not in the original version of the manuscript. In the spectral analyzes, the selectivity categorization depended on both (1) the difference effects between the domains and the baseline, and (2) the difference effect between domains. Channels were marked as selective when there was (1) a significant difference between domains and (2) only one domain significantly differed from the baseline. All difference effects were estimated using the paired sample permutation tests based on the t-statistic from the mne-python library (Gramfort et al., 2014) with 1000 permutations and the build-in *tmax* method to correct for the multiple comparisons over channels (Nichols & Holmes, 2002; Groppe et al. 2011). We have now more clearly explained how we controlled family-wise error in the Methods section:

"For each frequency band and channel, the statistical difference between conditions was estimated with paired sample permutation tests based on the t-statistic from the mne-python library (Gramfort et al., 2014) with 1000 permutations and the tmax method to control the family-wise error rate (Nichols and Holmes 2002; Groppe et al. 2011). In tmax permutation testing, the null distribution is estimated by, for each channel (i.e. each comparison), swapping the condition labels (speech vs music or speech/music vs baseline) between epochs. After each permutation, the most extreme t-scores over channels (tmax) are selected for the null distribution. Finally, the t-scores of the observed data are computed and compared to the simulated tmax distribution, similar as in parametric hypothesis testing. Because with an increased number of comparisons, the chance of obtaining a large tmax (i.e. false discovery) also increases, the test automatically becomes more conservative when making more comparisons, as such correcting for the multiple comparison between channels."

References :

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. NeuroImage, 86, 446–460.

Groppe, D. M., Bickel, S., Dykstra, A. R., Wang, X., Mégevand, P., Mercier, M. R., Lado, F. A., Mehta, A. D., & Honey, C. J. (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. Journal of Neuroscience Methods, 281, 40–48.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human Brain Mapping, 15(1), 1–25.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *Other suggestions:*
>
> *(1) The authors need to provide more details on how the sEEG electrodes were localized and selected. Are all electrodes included or only the ones located in the gray matter? If all electrodes were used, how to localize and label the ones that are outside of gray matter?*

> *In Figures 1C & 1D it seems that a lot of the electrodes were located in depth locations, how were the anatomical labels assigned for these electrodes*

We apologize that this was not clear in the original version of the manuscript. Our electrode localization procedure was based on several steps described in detail in Mercier et al., 2022. Once electrodes were localized in a post-implant CT-scan and the coordinates projected onto the pre-implant MRI, we were able to obtain the necessary information regarding brain tissues and anatomical region. That is, first, the segmentation of the pre-impant MRI with SPM12 provided both the tissue probability maps (i.e. gray, white, and cerebrospinal fluid (csf) probabilities) and the indexed-binary representations (i.e., either gray, white, csf, bone, or soft tissues) that allowed us to dismiss electrodes outside of the brain and select those in the gray matter. Second, the individual's brain was co-registered to a template brain, which allowed us to back project atlas parcels onto individual's brain and assign anatomical labels to each electrode. The result of this procedure allowed us to group channels by anatomical parcels as defined by the Brainnetome atlas (Figure 1D), which informed the analyses presented in section Population Prevalence (Methods, Figures 4, 9-10, S4-5). Because this study relies on stereotactic EEG, and not Electro-Cortico-Graphy, recording sites include both gyri and sulci, while depth structures were not retained.

We have now updated the "General preprocessing related to electrodes localisation" section in the Methods. The relevant part now states:

"To precisely localize the channels, a procedure similar to the one used in the iELVis toolbox and in the fieldtrip toolbox was applied (Groppe et al., 2017; Stolk et al., 2018). First, we manually identified the location of each channel centroid on the post-implant CT scan using the Gardel software (Medina Villalon et al., 2018). Second, we performed volumetric segmentation and cortical reconstruction on the pre-implant MRI with the Freesurfer image analysis suite (documented and freely available for download online http://surfer.nmr.mgh.harvard.edu/). This segmentation of the pre-implant MRI with SPM12 provides us with both the tissue probability maps (i.e. gray, white, and cerebrospinal fluid (CSF) probabilities) and the indexed-binary representations (i.e., either gray, white, CSF, bone, or soft tissues). This information allowed us to reject electrodes not located in the brain. Third, the post-implant CT scan was coregistered to the pre-implant MRI via a rigid affine transformation and the pre-implant MRI was registered to MNI152 space, via a linear and a non-linear transformation from SPM12 methods (Penny et al., 2011), through the FieldTrip toolbox (Oostenveld et al., 2011). Fourth, applying the corresponding transformations, we mapped channel locations to the pre-implant MRI brain that was labeled using the volume-based Human Brainnetome Atlas (Fan et al., 2016)."

Reference:

Mercier, M. R., Dubarry, A.-S., Tadel, F., Avanzini, P., Axmacher, N., Cellier, D., Vecchio, M. D., Hamilton, L. S., Hermes, D., Kahana, M. J., Knight, R. T., Llorens, A., Megevand, P., Melloni, L., Miller, K. J., Piai, V., Puce, A., Ramsey, N. F., Schwiedrzik, C. M., … Oostenveld, R. (2022). Advances in human intracranial electroencephalography research, guidelines and good practices. NeuroImage, 260, 119438.

> *(2) From Figures 5 and 6 (and also S4, S5), is it true that aside from the shared response, lower frequency bands show more music selectivity (blue dots), while higher frequency bands show more speech selectivity (red dots)? I am curious how the authors interpret this.*

The reviewer is right in noticing the asymmetric selective response to music and speech in lower and higher frequency bands. However, while this effect is apparent in the analyzes wherein we inspected stronger synchronization (activation) compared to baseline (Figures 2

and S1), the pattern appears to reverse when examining deactivation compared to baseline (Figures 3 and S2). In other words, there seems to be an overall stronger deactivation for speech in the lower frequency bands and a relatively stronger deactivation for music in the higher frequency bands.

We now provide additional details about the sign of selectivity between domains and frequencies in the Results section:

“Both domains displayed a comparable percentage of selective responses across frequency bands (Figure 4, first values of each plot). When considering separately activation (Figure 2) and deactivation (Figure 3) responses, speech and music showed complementary patterns: for low frequencies (<15 Hz) speech selective (and preferred) responses were mostly deactivations and music responses activations compared to baseline, and this pattern reversed for high frequencies (>15 Hz).”

Note, however, that this pattern of results depends on only a select number of patients, i.e. when ignoring regional selective responses that are driven by as few as 2 to 4 patients, the pattern disappears (Figures 5-6). More precisely, ignoring regions explored by a small number of patients almost completely clears the selective responses for both speech and music. For this reason, we do not feel confident interpreting the possible asymmetry in low vs high frequency bands differently encoding (activation or deactivation) speech and music.

> *Minor:*
>
> *(1) P9 L234: Why only consider whether these channels were unresponsive to the other domain in the other frequency bands? What about the responsiveness to the target domain?*

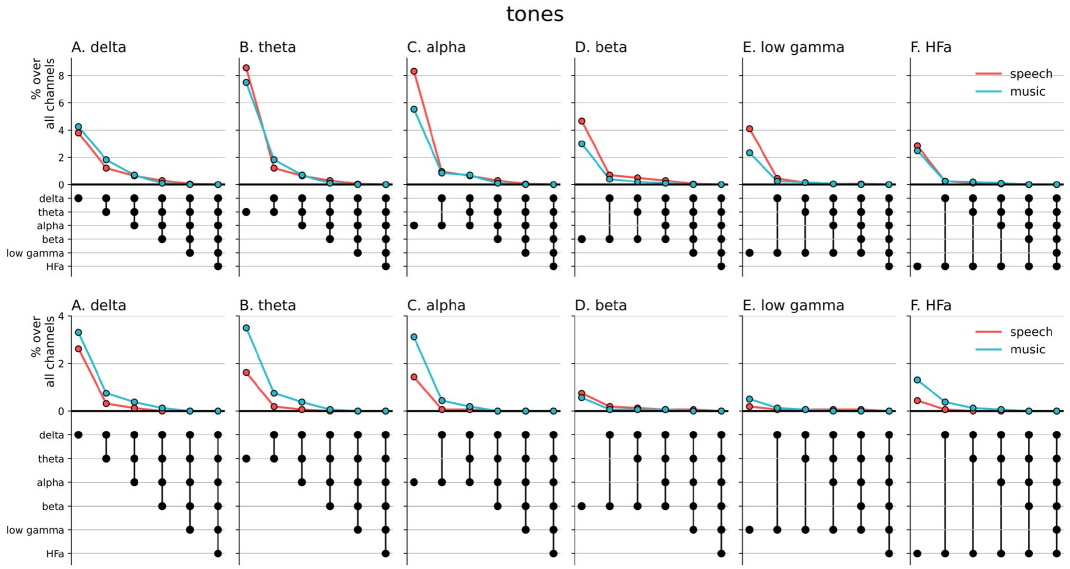
We thank the reviewer for their interesting suggestion. The primary objective of the cross-frequency analyzes was to determine whether domain-selective channels for a given frequency band remain unresponsive (i.e. exclusive) to the other domain across frequency bands, or whether the observed selectivity is confined to specific frequency ranges (i.e.frequency-specific). In other words, does a given channel exclusively respond to one domain and never—in whichever frequency band—to the other domain? The idea behind this question is that, for a channel to be selectively involved in the encoding of one domain, it does not necessarily need to be sensitive to all timescales underlying that domain as long as it remains unresponsive to any timescale in the other domain. However, if the channel is sensitive to information that unfolds slowly in one domain and faster in the other domain, then the channel is no longer globally domain selective, but the selectivity is frequency-specific to each domain.

The proposed analyzes answer a slightly different, albeit also meaningful, question: how many frequencies (or frequency bands) do selective responses span? From the results presented below, the reviewer can appreciate the overall steep decline in selective response beyond the single frequency band with only few channels remaining selectively responsive across maximally four frequency bands. That is, selective responses globally span one frequency band.

Author response image 1.

Cross-frequency channel selective responses. The top figure shows the results for the spectral analyzes (baselined against the tones condition, including both activation and deactivation). The bottom figure shows the results for the connectivity analyzes. For each plot, the first (leftmost) value corresponds to the percentage (%) of channels displaying a selective response in a specific frequency band. In the next value, we remove the channels that no longer respond selectively to the target domain for the following frequency band. The black dots at

the bottom of the graph indicate which frequency bands were successively included in the analysis.



(2) P21 L623: "Population prevalence." The subsection title should be in bold.

Done.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *The authors chose to use pure tone and syllables as baseline, I wonder if they also tried the rest period between tasks and if they could comment on how it differed and why they chose pure tones, (above and beyond a more active auditory baseline).*

This is an interesting suggestion. The reason for not using the baseline between speech and music listening (or right after) is that it will be strongly influenced by the previous stimulus. Indeed, after listening to the story it is likely that patients keep thinking about the story for a while. Similarly after listening to some music, the music remains in “our head” for some time.

This is why we did not use rest but other auditory stimulation paradigms. Concerning the choice of pure tones and syllables, these happen to be used for clinical purposes to assess functioning of auditory regions. They also corresponded to a passive listening paradigm, simply with more basic auditory stimuli. We clarified this in the Results section:

“[...] with respect to two baseline conditions, in which patients passively listened to more basic auditory stimuli: one in which patients passively listened to pure tones (each 30 ms in duration), the other in which patients passively listened to isolated syllables (/ba/ or /pa/, see Methods).”

> *Discussion - you might want to address phase information in contrast to power. Your encoding models map onto low-frequency (bandpassed) activity which includes power and phase. However, the high-frequency model includes only power. The model comparison is not completely fair and may drive part of the effects in Figure 7a. I would recommend discussing this, or alternatively ruling out the effect with modeling power separately for the low frequency.*

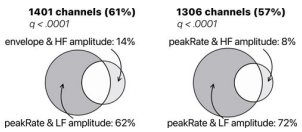
We thank the reviewer for their recommendation. First, we would like to emphasize that the chosen signal extraction techniques that we used are those most frequently reported in previous papers (e.g. Ding et al., 2012; Di Liberto et al., 2015; Mesgarani and Chang, 2012).

Low-frequency (LF) phase and high-frequency (HFa) amplitude are also known to track acoustic rhythms in the speech signal in a joint manner (Zion-Golumbic et al., 2013; Ding et al., 2016). This is possibly due to the fact that HFa amplitude and LF phase dynamics have a somewhat similar temporal structure (see Lakatos et al., 2005 ; Canolty and Knight, 2010).

Still, the reviewer is correct in pointing out the somewhat unfair model comparison and we appreciate the suggestion to rule out a potential confound. We now report in Supplementary Figure S8, a model comparison for LF amplitude vs. HFa amplitude to complement the findings displayed in Figure 7A. Overall, the reviewer can appreciate that using LF amplitude or phase does not change the results: LF (amplitude or phase) always better captures acoustic features than HFa amplitude.

Author response image 2.

TRF model comparison of low-frequency (LF) amplitude and high-frequency (HFa) amplitude. Models were investigated to quantify the encoding of the instantaneous envelope and the discrete acoustic onset edges (peakRate) by either the low frequency (LF) amplitude or the high frequency (HFa) amplitude. The 'peakRate & LF amplitude' model significantly captures the largest proportion of channels, and is, therefore, considered the winning model. Same conventions as in Figure 7A.



References:

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. Trends in Cognitive Sciences, 14(11), 506–515.

Di Liberto, G. M., O'sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. Current Biology, 25(19), 2457-2465.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences, 109(29), 11854-11859.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience, 19(1), 158–164.

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron, 77(5), 980-991.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. Journal of Neurophysiology, 94(3), 1904–1911.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature, 485(7397), 233-236.

> *Similarly, the Coherence analysis is affected by both power and phase and is not dissociated. i.e. if the authors wished they could repeat the coherence analysis with phase coherence (normalizing by the amplitude). Alternatively, this issue could be addressed in the discussion above*

We agree with the Reviewer. We have now better clarified our choice in the Methods section:

“Our rationale to use coherence as functional connectivity metric was three fold. First, coherence analysis considers both magnitude and phase information. While the absence of dissociation can be criticized, signals with higher amplitude and/or SNR lead to better time-frequency estimates (which is not the case with a metric that would focus on phase only and therefore would be more likely to include estimates of various SNR). Second, we choose a metric that allows direct comparison between frequencies. As, at high frequencies phase angle changes more quickly, phase alignment/synchronization is less likely in comparison with lower frequencies. Third, we intend to align to previous work which, for the most part, used the measure of coherence most likely for the reasons explained above.“

https://doi.org/10.7554/eLife.94509.2.sa4