

## RESEARCH ARTICLE

# Discerning the functional networks behind processing of music and speech through human vocalizations

Arafat Angulo-Perkins<sup>1,2</sup>, Luis Concha<sup>1,3\*</sup>

**1** Instituto de Neurobiología, Universidad Nacional Autónoma de México, Querétaro, Querétaro, México,

**2** Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, Vienna, Austria,

**3** International Laboratory for Brain, Music and Sound (BRAMS), Montreal, Québec, Canada

\* [lconcha@unam.mx](mailto:lconcha@unam.mx)



## OPEN ACCESS

**Citation:** Angulo-Perkins A, Concha L (2019) Discerning the functional networks behind processing of music and speech through human vocalizations. PLoS ONE 14(10): e0222796. <https://doi.org/10.1371/journal.pone.0222796>

**Editor:** Sukru Torun, Anadolu University, TURKEY

**Received:** March 20, 2019

**Accepted:** September 6, 2019

**Published:** October 10, 2019

**Copyright:** © 2019 Angulo-Perkins, Concha. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying this study has been deposited to the OpenNeuro database (<https://openneuro.org/datasets/ds001482>). All other relevant data are within the paper and its Supporting Information files.

**Funding:** This study was sponsored by Conacyt (IE252-120295 and 181508) and UNAM-DGAPA (I1202811 and IN212811). We thank the Posgrado en Ciencias Biomédicas of the Universidad Nacional Autónoma de México (UNAM) and the CONACYT for Graduate Fellowship 233109 to A.P. A. The funders had no role in study design, data

## Abstract

A fundamental question regarding music processing is its degree of independence from speech processing, in terms of their underlying neuroanatomy and influence of cognitive traits and abilities. Although a straight answer to that question is still lacking, a large number of studies have described where in the brain and in which contexts (tasks, stimuli, populations) this independence is, or is not, observed. We examined the independence between music and speech processing using functional magnetic resonance imaging and a stimulation paradigm with different human vocal sounds produced by the same voice. The stimuli were grouped as **Speech** (spoken sentences), **Hum** (hummed melodies), and **Song** (sung sentences); the sentences used in **Speech** and **Song** categories were the same, as well as the melodies used in the two musical categories. Each category had a scrambled counterpart which allowed us to render speech and melody unintelligible, while preserving global amplitude and frequency characteristics. Finally, we included a group of musicians to evaluate the influence of musical expertise. Similar global patterns of cortical activity were related to all sound categories compared to baseline, but important differences were evident.

Regions more sensitive to musical sounds were located bilaterally in the anterior and posterior superior temporal gyrus (*planum polare* and *temporale*), the right supplementary and premotor areas, and the inferior frontal gyrus. However, only temporal areas and supplementary motor cortex remained music-selective after subtracting brain activity related to the scrambled stimuli. Speech-selective regions mainly affected by intelligibility of the stimuli were observed on the left *pars opercularis* and the anterior portion of the medial temporal gyrus. We did not find differences between musicians and non-musicians. Our results confirmed music-selective cortical regions in associative cortices, independent of previous musical training.

## Introduction

Despite continuous debate regarding the level of functional, anatomical and cognitive independence between speech and music processing [1–5], it is clear that there are several

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

differences regarding their basic acoustic properties (e.g., temporal, spectral or envelope features) and more importantly, they do not carry the same type of information (e.g., verbal aspects such as propositional meaning present in speech in comparison to music) [5–15]. Musical cognitive traits such as melodic, harmonic, timbral or rhythmic processing rely on basic analysis such as relative pitch, beat perception or metrical encoding of rhythm [13,16]. Interestingly, cortical activity (as measured through blood oxygenation level-dependent [BOLD] signal) is higher in secondary auditory cortices while listening to music as opposed to various types of non-musical human vocalizations—despite these regions being essential for speech processing [16,17]. The primary auditory cortex (i.e., Heschl's gyrus), on the other hand, does not show differential activation towards these two types of acoustic categories [18–22]. These results suggest that while music and speech processing share the same basic auditory pathway until the primary auditory cortex, different patterns of activation are observed in other brain regions, with some exhibiting hemisphere lateralization [19–24].

The immediate question when comparing music and speech refers to human vocalizations, given that they represent the most common sound in our environment, and the most ancient expression of music and language [25–29]. In addition to their evolutionary relevance, human vocalizations provide experimental advantages, as they allow researchers to test both acoustic signals at the same time (e.g., song), or to explore them separately (i.e., just linguistic or melodic information) while evaluating perception, discrimination, memory tasks, or vocal production [24,30–36]. Nevertheless, the level of independence between speech and vocal music reported in some of these studies is still difficult to evaluate, particularly because only few studies have included the three basic vocal expressions to compare music, speech and their combination (i.e., speech, hum and song) [8,24,33]. In addition, studies using passive listening paradigms are scarce [8,24,30,31,37], with most studies having subjects perform tasks while they listen to different stimuli, or requiring vocal production (overt and covert) [8,32,33,36,38].

Previous findings have shown that regardless of musical stimuli being human vocalizations (such as a syllabic hum or song) there are certain differences in brain activity when compared to that elicited by speech, particularly a right hemisphere lateralization for music processing which could be modulated depending on the content of lyrics or tasks involved [8,24,32,33,39].

Given this, our goal was to evaluate at which point in the hierarchy of the auditory pathway these two acoustic signals show divergent processing, and if these differences in brain activation are maintained after altering some of the basic acoustic properties, which are supposed to involve primary areas of the auditory cortex according to the hierarchical functional organization [40–42]. To achieve this, we designed an experimental paradigm including three vocal categories, speech stimuli, and two different forms of musical vocal sounds, namely one with lyrics (song) and one without (hum). To evaluate which regions are functionally more affected by the low-level acoustic manipulations, we included scrambled versions of each category (i.e., **Speech**, **Song** and **Hum**) affecting the intelligibility of the vocal sound but preserving the amplitude and spectral features. In the same way, we controlled for semantic content and included different sentences that were sung and spoken by a professional singer. Melodic content was the same for the hummed and sung stimuli.

Our main hypothesis was that music-selective regions would show increased activity during both musical categories, in comparison to speech regardless of: 1) being produced by the same instrument, 2) the presence or absence of lyrics (i.e., song and hum, respectively) and 3) musical expertise. In the same way, we predicted that regions mainly modulated by speech processing would not show differences in their activation in response to similar categories (i.e., **Song** and **Speech**) but would differ with respect to the music category without verbal content

(Hum). Based on previous results and numerous studies showing brain differences (anatomical and functional) in musicians as a result of training [19,43–50], we also hypothesized that musical training modulates cortical activity related to the three acoustic stimuli used. We therefore explored functional differences associated to musical training during vocal music perception between musicians and non-musicians.

## Results

All sound categories produced BOLD signals that were significantly higher than those obtained during the baseline condition (i.e., scanner noise; Fig 1). Analysis of trigger responses to target sounds showed that all subjects responded appropriately (i.e., they were attentive to the stimuli).

### Analysis 1: Natural versus scrambled stimuli

We searched for regions sensitive to spectral envelope modifications. All natural stimuli, when compared to their scrambled counterparts, elicited stronger BOLD activity around (but not in) Heschl's gyrus in the superior temporal gyrus (STG). However, particular areas were differentially activated by specific categories, as outlined below and visualized in Fig 2. Statistical maps for this and all other analyses are available at <https://neurovault.org/collections/DMKGWLFE/>.

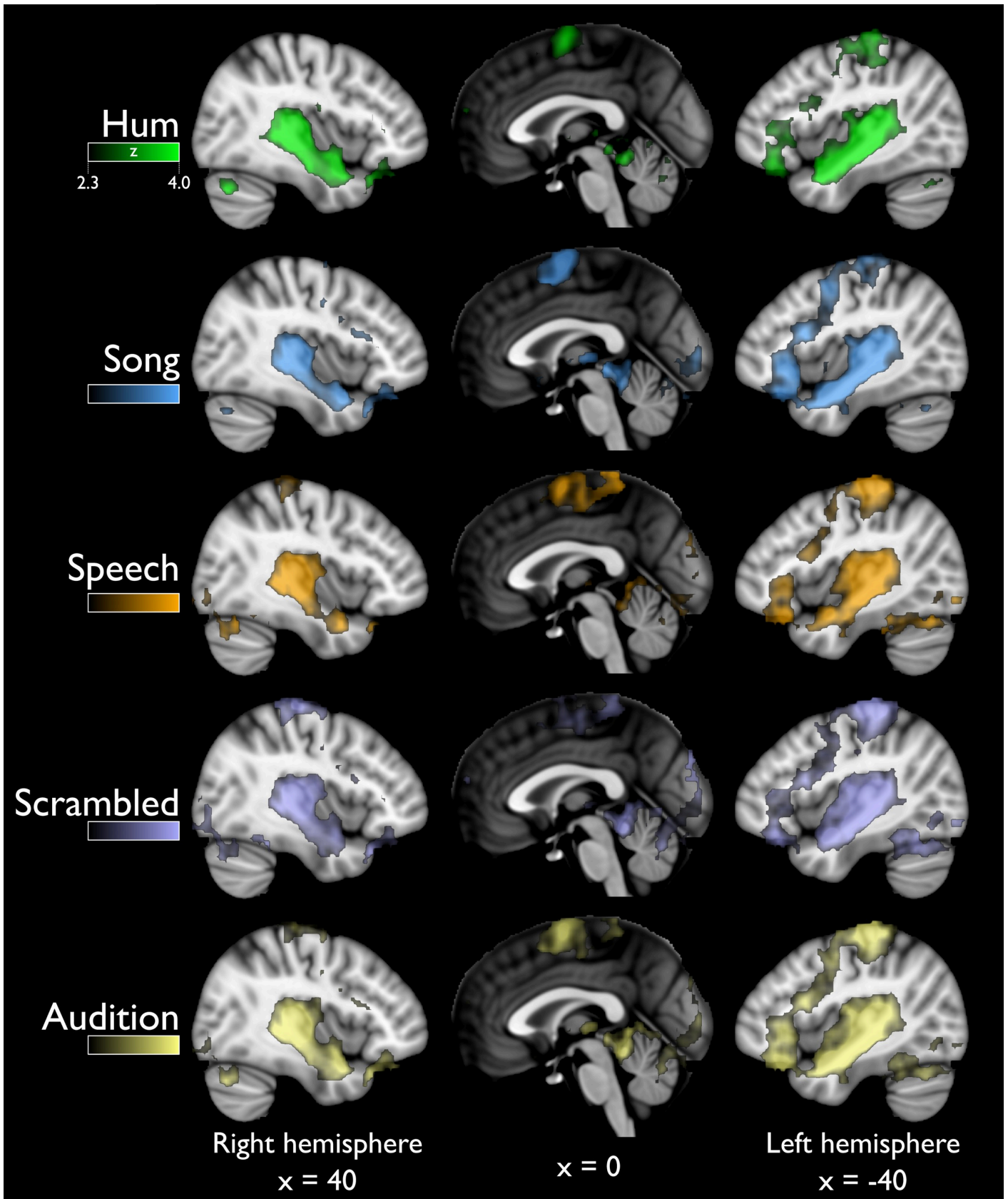
**1.1 Song versus Song-S.** The **Song > Song-S** comparison revealed 4 different clusters, two of them occupying the entire lateral aspect of the STG and the superior temporal sulcus (STS) bilaterally, with the notable exception of Heschl's gyrus; however, the most significant activation was located within the temporal pole (aSTG; blue colors in Fig 2). The left STG cluster extended into the inferior frontal gyrus (IFG) approximately in Brodmann areas 44 and 45. The third cluster had its most significant voxel in the right supplementary motor cortex (SMA) but it also covered the left counterpart, where it reached the dorsal premotor cortex (PMC). The opposite contrast (**Song-S > Song**) showed 4 clusters covering different regions: bilateral activation of primary auditory regions (Heschl's gyrus), the insular cortex and two bilateral activation of the middle and inferior occipital gyrus.

**1.2 Hum versus Hum-S.** The contrast testing for **Hum > Hum-S** showed bilateral activation along the STG with three peaks of maximal activation (Fig 2, green colors), two of which corresponded to the left and right temporal pole and the other one to the *planum temporale*. The fourth cluster was located in the supplementary motor area (SMA). No differences were found in Heschl's gyrus. The inverse contrast testing for **Hum-S > Hum** evidenced two clusters corresponding to Heschl's gyrus occupying the adjacent lateral face of the STG.

**1.3 Speech versus Speech-S.** Functional maps testing for **Speech > Speech-S** showed two large clusters distributed along the ventral lateral face of the STG reaching the dorsal area of the STS, in the left hemisphere the cluster presented a much larger volume (left 17,824 mm<sup>3</sup> and right 5,552 mm<sup>3</sup>; Fig 2, warm colors). The left cluster included the IFG, as did the **Song > Song-S** contrast. The analysis of the opposite comparison (**Speech-S > Speech**) also revealed two clusters located in the lateral aspect of Heschl's gyrus (in each hemisphere); cluster locations were almost identical to those found in **Hum-S > Hum** and in **Song-S > Song**.

### Analysis 2: Activation patterns for Song versus Speech

To search for differences between **Song** and **Speech** categories that are not explained solely by differences in their acoustic features, we performed a high-level statistical analysis comparing the parameter estimates derived from Analysis 1. The contrast (**Song > Song-S**) > (**Speech > Speech-S**) generated three different clusters. Two clusters were distributed along the right



**Fig 1. Global activity (above baseline) for all stimuli.** Sagittal views showing distributed activation in response to all sound categories, greater than to scanner noise. The last row (Audition) includes all the previous categories. R = right, L = left. Statistical maps are overlaid on the MNI-152 atlas (coordinates shown in mm). For all categories color scale corresponds to z values in the range (2.3–4.0), as in top panel.

<https://doi.org/10.1371/journal.pone.0222796.g001>

STG from the *planum temporale* to the *planum polare*, sparing Heschl's gyrus (Fig 3, cold colors); in the left hemisphere the activation was located on the aSTG covering part of the *planum polare*. Another cluster was located over the right supplementary motor area (SMA). The opposite contrast ([Speech>Speech-S] > [Song>Song-S]; Fig 3, warm colors) elicited focal activation located on the anterior portion of the right middle temporal gyrus (MTG) cortex, extending slightly into the superior temporal sulcus (STS). Differences between categories without controlling for their scrambled counterparts are shown in S1 Fig, which also extends Analyses 3 and 4.

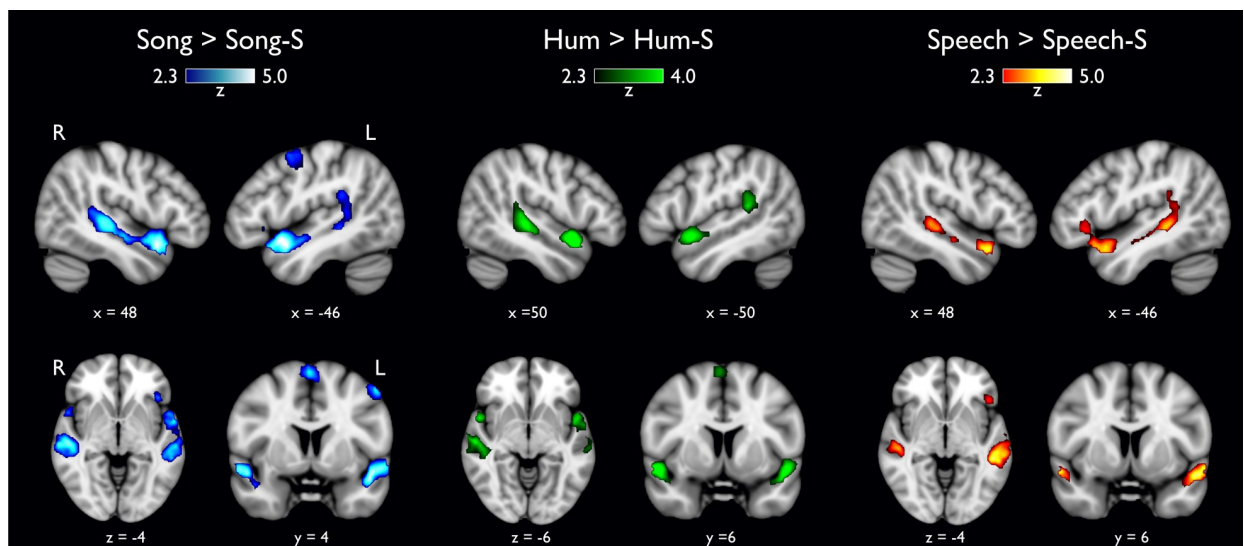
### Analysis 3: Activation patterns for Hum versus Speech

By excluding lyrics and leaving melodic information intact (i.e., **Hum**), we aimed to observe music-sensitive regions when compared with activations elicited by **Speech**. Similarly to Analysis 2, the contribution of acoustic features to brain activity was controlled by subtracting parameter estimates related to the corresponding scrambled acoustic categories.

Functional maps testing for (**Hum**>**Hum-S**) > (**Speech**>**Speech-S**) (Fig 4; left panel, green colors) revealed bilateral activation on the posterior STG (pSTG). In both hemispheres the activation was observed on the *planum temporale*. The opposite comparison ([**Speech**>**Speech-S**] > [**Hum**>**Hum-S**]) showed significant activity distributed along the left STS and dorsal MTG (Fig 4, left panel, warm colors), the cluster extended from the posterior regions until the anterior portion of the MTG (aMTG). A second cluster was located on the *pars opercularis* from the left inferior frontal gyrus (IFG).

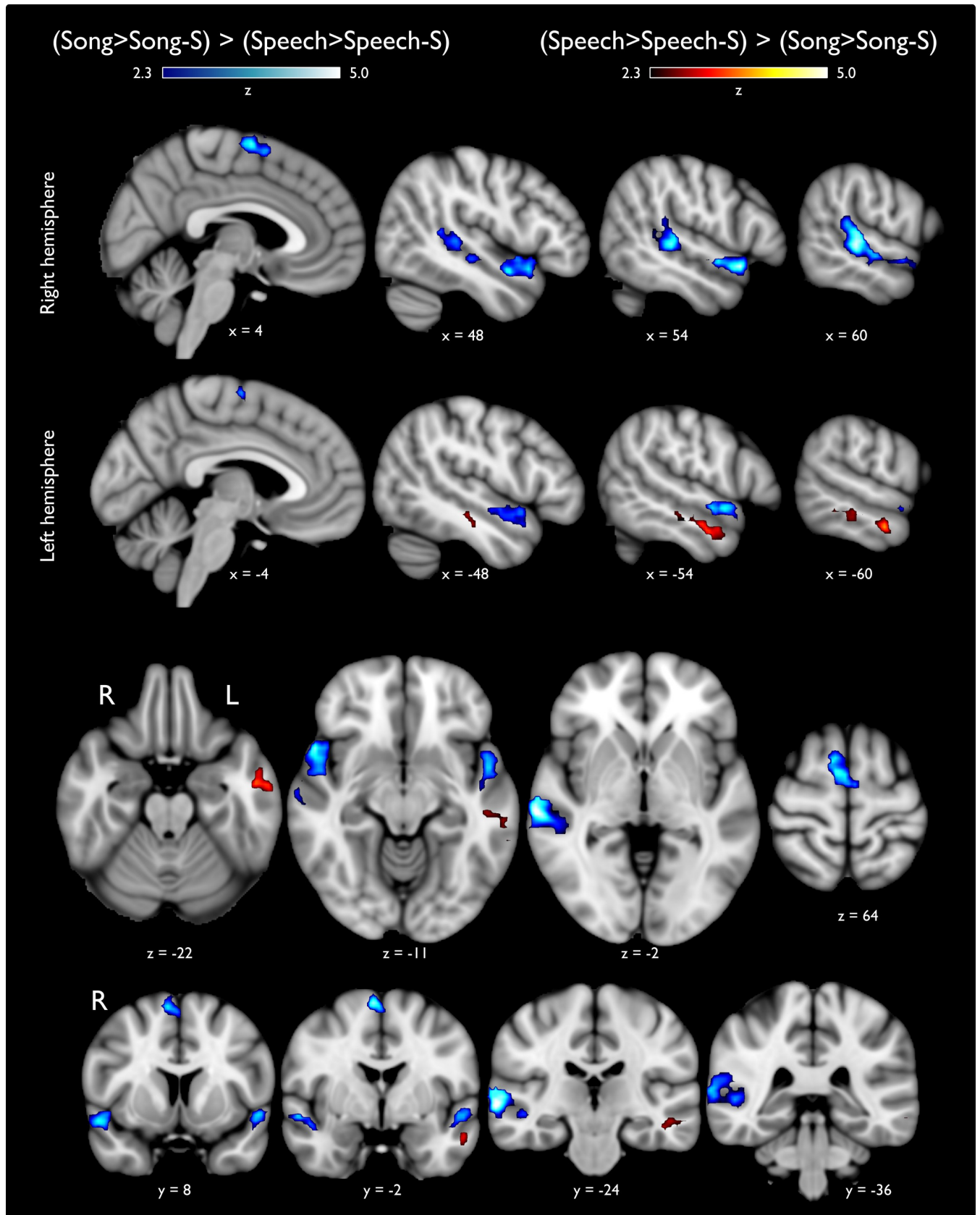
### Analysis 4: Comparing vocal-music with and without lyrics

This analysis compared the brain activation associated to **Song** stimuli (i.e., music with lyrics) with the hummed version, as a way to explore the question regarding sharing resources



**Fig 2. Natural stimuli versus scrambled stimuli.** Each category contrasted against its own scrambled version (-S). No differences were found in Heschl's gyrus; activation in the anterior portion of the superior temporal gyrus, specifically the *planum polare*, was found only in the musical categories (white arrowheads).

<https://doi.org/10.1371/journal.pone.0222796.g002>



**Fig 3. Statistical activation maps comparing vocal music and speech sounds.** Statistical activation maps for differences between music and speech sounds after subtraction of activity elicited by their scrambled versions (-S). Blue colors indicate those regions that were more active while listening to song stimuli as compared to speech; warm colors show the opposite comparison. Upper and middle panels show lateral progression in the sagittal plane, from medial to lateral regions in each hemisphere. Axial and coronal views are presented in the inferior two panels to facilitate visual inspection.

<https://doi.org/10.1371/journal.pone.0222796.g003>

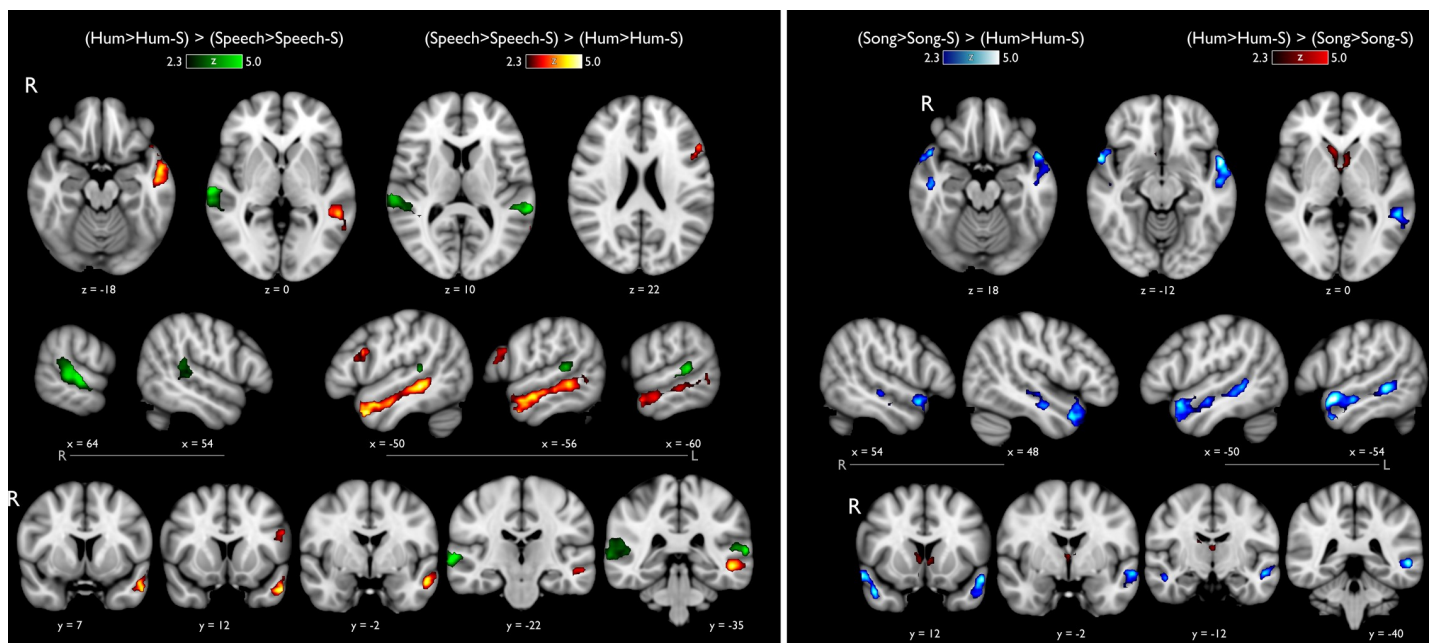
between music and speech. The results obtained from the contrast **(Song>Song-S) > (Hum>Hum-S)** (Fig 4; right panel, cold colors) showed activation covering the anterior and posterior aspects of the left STS, the cluster reached the dorsal portion of the MTG and the anterior portion and the most anterior portion of the STG (aSTG). In the right hemisphere we found the middle portion of the STS and the newly the aSTG. The reverse contrast **([Hum>Hum-S] > [Song>Song-S])** showed the caudate nucleus, bilaterally.

### Analysis 5: Influence of musical training

We evaluated whether musical training differentially modulated brain activity while listening to our experimental paradigm. All the contrasts described in Analyses 2–4 were tested for group differences, with none being statistically significant.

### Analysis 6: Functional characterization of the superior temporal gyrus

To further analyze BOLD signal changes derived from the different sound categories along the auditory cortices we used an unbiased approach to compare BOLD changes in the *planum polare*, *planum temporale* and Heschl’s gyrus. Fig 5 illustrates the percentage of BOLD signal change for each category (significance levels = \*  $p < 0.0028$ ; #  $p < 0.05$ ). We found statistically significant differences between **Song** and **Speech** in the left and right *planum polare*, and also



**Fig 4. Statistical activation maps for Hum versus Song and Speech.** Left panel: Hum *versus* Speech, after subtraction of activity elicited by their scrambled counterparts (-S). Listening to hummed songs elicited enhanced activity of the *planum temporale* bilaterally, as compared to that elicited by speech stimuli (green colors). The opposite contrast (warm colors) shows activation only in the left hemisphere including the IFG, MTG and STS. Right panel: Vocal-music with and without lyrics. Lateral, axial and coronal views showing in blue the cortical regions modulated preferentially to Song as compared to Hum stimuli, after removing the effect of their scrambled counterparts. Activation extended bilaterally over the aSTG and STS. The opposite contrast showed activation on the caudate nucleus.

<https://doi.org/10.1371/journal.pone.0222796.g004>

between **Song** and **Hum**; the lowest levels of activation within the *planum polare* were observed in the **Speech** category. BOLD responses in Heschl's gyrus (A1) were lower in response to **Hum** as compared to those elicited by **Song** or **Speech**, bilaterally; there were no differences in A1 activation between the last two categories. In parallel with this observation, acoustic feature analyses showed that **Hum** differs substantially from **Song/Speech** in several aspects, such as zero-cross, and spectral spread, brightness, centroid, kurtosis and flatness (S2 Fig).

The *planum temporale* showed different patterns between hemispheres. The right *planum temporale* was equally active during **Hum** and **Speech** conditions in contrast to **Song**; however, the left *planum temporale* exhibited similar pattern to that seen in A1, namely it was preferentially activated by **Song** and **Speech** categories (Fig 5; lower panel). From all comparisons among the three different regions in the STG, the highest activation were found in the left *planum temporale*.

## Discussion

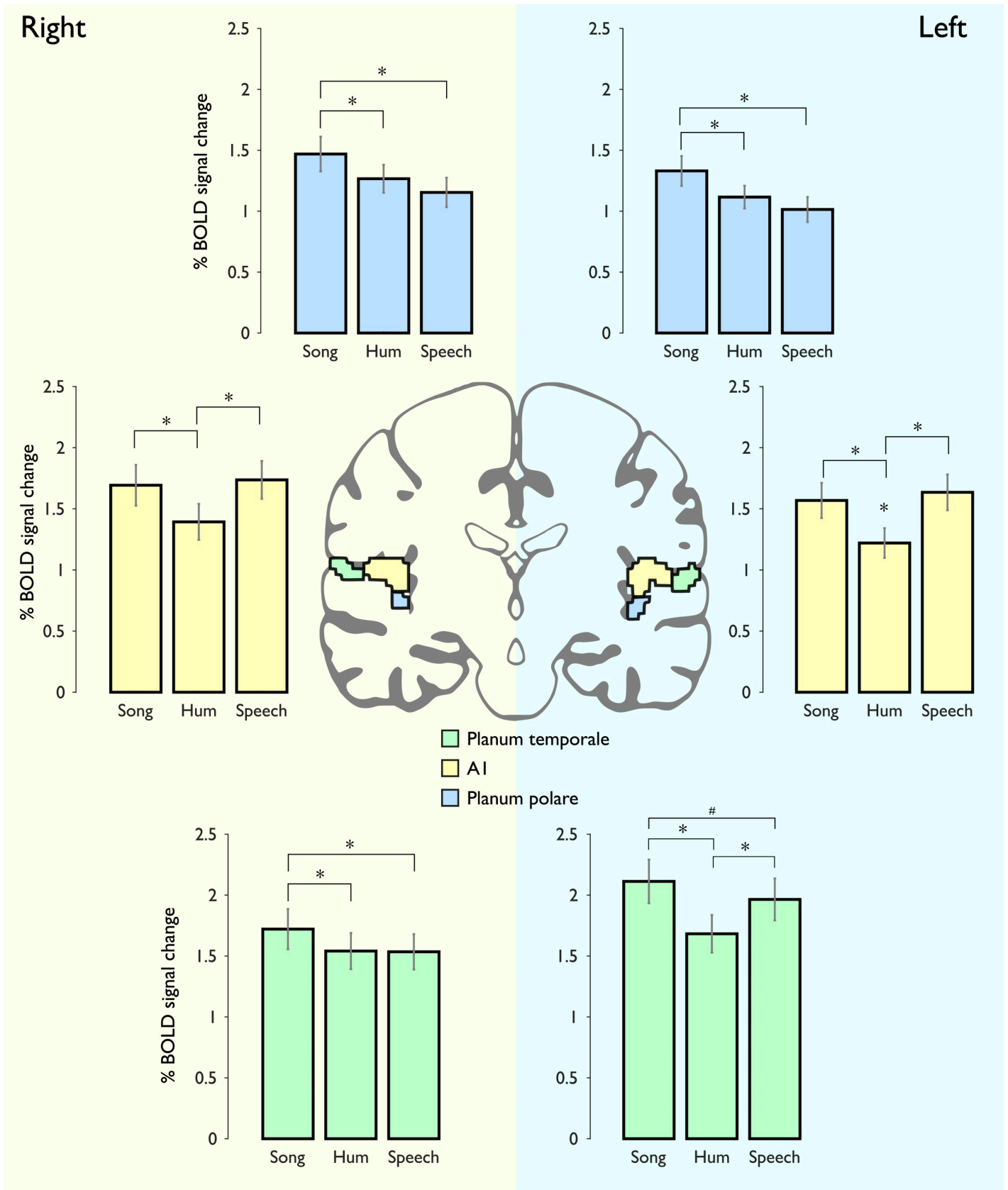
In this study we aimed to further explore brain regions modulated for music as compared to speech. In Particular, we investigated whether this selectivity changed when we combined the two domains (vocal music with lyrics), by comparing the **Song** category against **Speech**, or **Hum**, while we controlled for differences of timbre (using the same voice to produce all the stimuli), melodic, and semantic content (using the same sentences and same musical excerpts).

To identify brain regions sensitive to musical sounds at different processing levels, we first compared each natural vocal category against their scrambled versions. We observed that while all acoustic categories recruited different regions of the STG when compared them to their scrambled version, only the musical categories also revealed activation maps covering the SMA and PMC. This result confirms what has been previously described in speech processing showing that music is also hierarchically encoded involving brain regions beyond Heschl's gyrus [19,20,51,52]. All acoustic categories induced activations of the left posterior Sylvian region at the parietal-temporal junction (Spt), which has been suggested as a region integrating audio motor patterns from the vocal tract [53,54]. The contrast testing for **Song** > **Song-S** revealed bilateral activation of the *planum polare* and *temporale*, left PMC and SMA. This result suggests these areas are not exclusively modulated by low-level acoustic parameters, but also by the temporal structure affecting the semantic, syntactic and melodic-contour, as has been reported in other musical studies [22,55,56].

The opposite contrasts comparing the scrambled versions (e.g., **Song-S** > **Song**) revealed consistent higher BOLD modulations in Heschl's gyrus. The increased activity in response to scrambled acoustic stimuli, independently of acoustic category, is indicative of converging processing of music and speech processing at this level. The primary auditory cortex presents a tonotopical organization sensitive to frequency tuning and spectro-temporal modulations [57–61], which are present in all the categories we used. However, the primary auditory cortex also contributes to the construction of auditory objects, and it has been suggested that less meaningful sounds require more neural resources to extract information from them, in an attempt to elaborate predictions about their possible identity as an auditory object [60,62].

Once we had identified the brain regions that are sensitive to the temporal structure of acoustic stimuli, we searched for the functional independence between music and speech processing. The contrast (**Song**>**Song-S**) > (**Speech**>**Speech-S**) (Fig 3) showed that the *planum polare* bilaterally, the right *planum temporale*, and SMA, are auditory regions particularly relevant for music processing in comparison to speech, independently of the similarities regarding semantic content. In addition, results from contrast comparing (**Hum**>**Hum-S**) >





**Fig 5. ROIs of Heschl's gyrus, planum polare and planum temporale.** Right and left hemispheres are shaded in different colors. Top panel: The percentage of BOLD signal modulation in the *planum polare* (blue color) showed similar patterns between the right and left hemisphere, the highest level of activation in this region corresponded to Song, followed by Hum and Speech sounds; statistical differences were found between Song and Speech. Middle panel: BOLD signal changes in the right and left Heschl's gyrus (yellow color), no differences between Song and Speech stimuli. Bottom panel: The left and right *planum temporale* (green color) exhibited different patterns of activation, in the right hemisphere Song stimuli elicited the highest values, while in the left hemisphere there were no differences between them. \* indicates significant difference from the rest of the stimuli ( $p < 0.0028$ ; Bonferroni correction); # stands for uncorrected  $p < 0.05$ .

<https://doi.org/10.1371/journal.pone.0222796.g005>

(**Speech**>**Speech-S**) (Fig 4), highlight the functional identity of the *planum temporale*, as it was the only region showing consistently increased activity in response to musical conditions. The *planum polare*, which is very active bilaterally in the **Song** condition (as compared to activity elicited by either **Speech** or **Hum**) was not more responsive to **Hum** as it was to **Speech**. Thus, the *planum polare* showed higher levels of activation in both musical conditions, particularly when intelligibility was not affected. However, stronger bilateral activation was observed when vocal music included lyrics (e.g., [**Song**>**Song-S**] > [**Hum**>**Hum-S**]), likely reflecting the simultaneous processing of speech and music present in **Song**. This finding was corroborated with an unbiased ROI analysis (Fig 5). In line with previous reports, our results suggest the *planum polare* may play an intermediate role between the primary auditory cortex and other associative cortices, possibly extracting information such as melodic patterns or pitch-interval ratios, required for further processing leading to perceptual evaluation of complex musical patterns [34,63]. The *planum polare* has been specifically reported in other studies evaluating vocal music while performing a same-different task [8,24] and during in over and cover vocal production tasks [32,33,64]. Similarly, the largest activation of the right *planum temporale* was elicited by **Song** stimuli, (whereas no significant differences were found between **Hum** and **Speech**), suggesting it has a shared functional identity during vocal music and speech processing, specially when they are combined (**Song**). On the other hand, the *left planum temporale* was equally modulated by **Speech** and **Song** categories, confirming its preference for stimuli with verbal content (whether musical or not), possibly associated with representations of lexical or semantic structures [51,65,66].

Our results showed music-sensitivity in cortical motor regions such as the SMA and the PMC, which have been related to beat and rhythm perception [67–72]. Particularly, SMA showed music preference independently of whether the music was hummed or sung and also regardless of musical training. These results support a particular role of SMA in music processing and its relation to action, as Lima and colleagues [73] suggested a possible role in motor facilitation to prepare and organize movement sequences.

It has been described that speech processing involves distributed cortical regions (e.g., primary auditory cortices, left STG, MTG, *planum temporale*, STS, IFG, post-central gyrus and the ventral division of M1) [74,75], which are recapitulated in our results (Fig 2 and S1 Fig). While **Speech** elicited larger activity than **Hum** in the entire left MTG (Fig 4), only the most anterior aspect of MTG was more sensitive to **Speech** than **Song** (Fig 3), suggesting a predominant role of the MTG in the processing of linguistic content.

In a previous study, we found differences between musicians and non-musicians involving greater bilateral activity of the *planum polare* and the right *planum temporale*, when listening to different types of instrumental music stimuli in comparison to speech [19]. In our current work, however, we did not find evidence that musical expertise modulates brain regions while processing vocal music or speech. We suggest this negative finding is indicative of the common exposure to (and perhaps similar production of) vocal music, as singing is commonplace in everyday situations since early age (e.g., birthdays, hymns, cheers); contrarily, learning to play an instrument is not a universal activity. Nonetheless, other factors may be responsible for the discrepancy with our previous report, such as the relatively smaller size of our current

sample (17 musicians), and the diversity of musical training in the group of musicians (S1 Table), which may induce functional plastic changes related to instrument-specific tuning [49,50]. The effect of musical training should therefore be explored further and more thoroughly (e.g., by studying professional singers).

### Limitations of the Study

Together, these data demonstrate different cortical regions that are preferentially modulated by particular sounds, whether they are music, speech or a mixture of the two. However, the anatomical resolution given by the technique does not allow us to distinguish finer anatomical details regarding the specific distribution of the statistical maps. For the same reason, we can not elaborate a more detailed analysis regarding the participation of the primary auditory cortex in **Speech** and **Song** conditions, for example. Temporal resolution is also limited in all fMRI studies, and there is a wealth of information from rapid temporal fluctuations present in music and language that are not easily addressed with this technique. As such, our results can only reveal relatively long-term changes of brain activity in response to listening to specific sound categories, and our conclusions will benefit greatly from other methods with high temporal resolution, such as (magneto-) electrophysiological recordings. Finally, although stimuli of all three categories were produced by the same singer, we acknowledge that other acoustic features were not homogenized to control for all acoustic parameters that could potentially differ between categories. While some differences in cortical activity may be explained by these low-level features, our main findings are better explained by higher-level, time-varying acoustic features that are characteristic of each category.

### Conclusions

Our data indicates that music selectivity in distributed brain regions independently of using vocal music with and without lyrics. Music-sensitive regions involved frontal and temporal cortical areas, specially in the right hemisphere. Our results indicate that the temporal structure of vocal music and speech is processed in a large temporal-frontal network, and that speech and music processing diverge, with specific regions (such as the *planum polare* and *temporale*, and SMA), being particularly sensitive to the temporal structure and acoustical properties of music subjected to further processing beyond the primary auditory cortices.

### Methods

#### Participants

Thirty-three healthy, right-handed volunteers, age  $28 \pm 8$  years (range: 20 to 42 years; 17 women), participated in this study. Seventeen volunteers were non-musicians (age  $27 \pm 6$  years; range: 20–45 years; 9 women), who had not received extra-curricular music education beyond mandatory courses in school. Musicians (16 volunteers; age  $28 \pm 7$  years; range: 20–42 years; 8 women) had received at least 3 years of formal training/studies in music (either instruments or singing) and were currently involved in musical activities on a daily basis (S1 Table). Groups did not differ in terms of age or gender. All volunteers were native Spanish speakers, self-reported normal hearing (which was confirmed during an audio test within the scanner), were free of contraindications for MRI scanning and gave written informed consent before the scanning session. The research protocol had approval from the Ethics Committee of the Institute of Neurobiology at the Universidad Nacional Autónoma de México and was conducted in accordance with the international standards of the Declaration of Helsinki of 1964.

## Experimental design

The vocal music paradigm used a pseudo-randomized block design; each block lasted 15 seconds and included 5–7 stimuli from the same category (Fig 6, panel A). Six different categories were included: **Hum**, **Song**, **Speech**, and their scrambled counterparts (5 blocks each along the paradigm). Additionally, 5 blocks of silence (each lasting 15 seconds) were interspersed throughout the stimulation paradigm. The stimulation protocol had a total duration of ~10 minutes.

The stimulation paradigm was presented with E-prime Study Software (version 2.0; Psychology Software Tools, Sharpsburg, PA) binaurally through MRI-compatible headphones (Nordic NeuroLab, Bergen, Norway) that attenuated acoustic interference (~20 dB) generated by the gradients. While not used as a formal test for normal hearing, we performed a short audio test (1 min) inside the scanner, using similar stimuli, to evaluate whether volunteers could hear and recognize the different sounds inside the scanner; the volume was deemed comfortable but sufficiently high to mask the noise generated by the imaging acquisitions.

Subjects were instructed to pay attention to the stimuli and to press a button with the index finger of their right hand every time they heard a pure tone (500 Hz, 500 ms duration), which was presented 5 times randomly throughout the paradigm. We used this strategy to ensure attention throughout the stimulation paradigm, and we used a pure tone as it is clearly different from the rest of our stimuli and therefore easily recognizable. Subjects kept their eyes open during scanning.

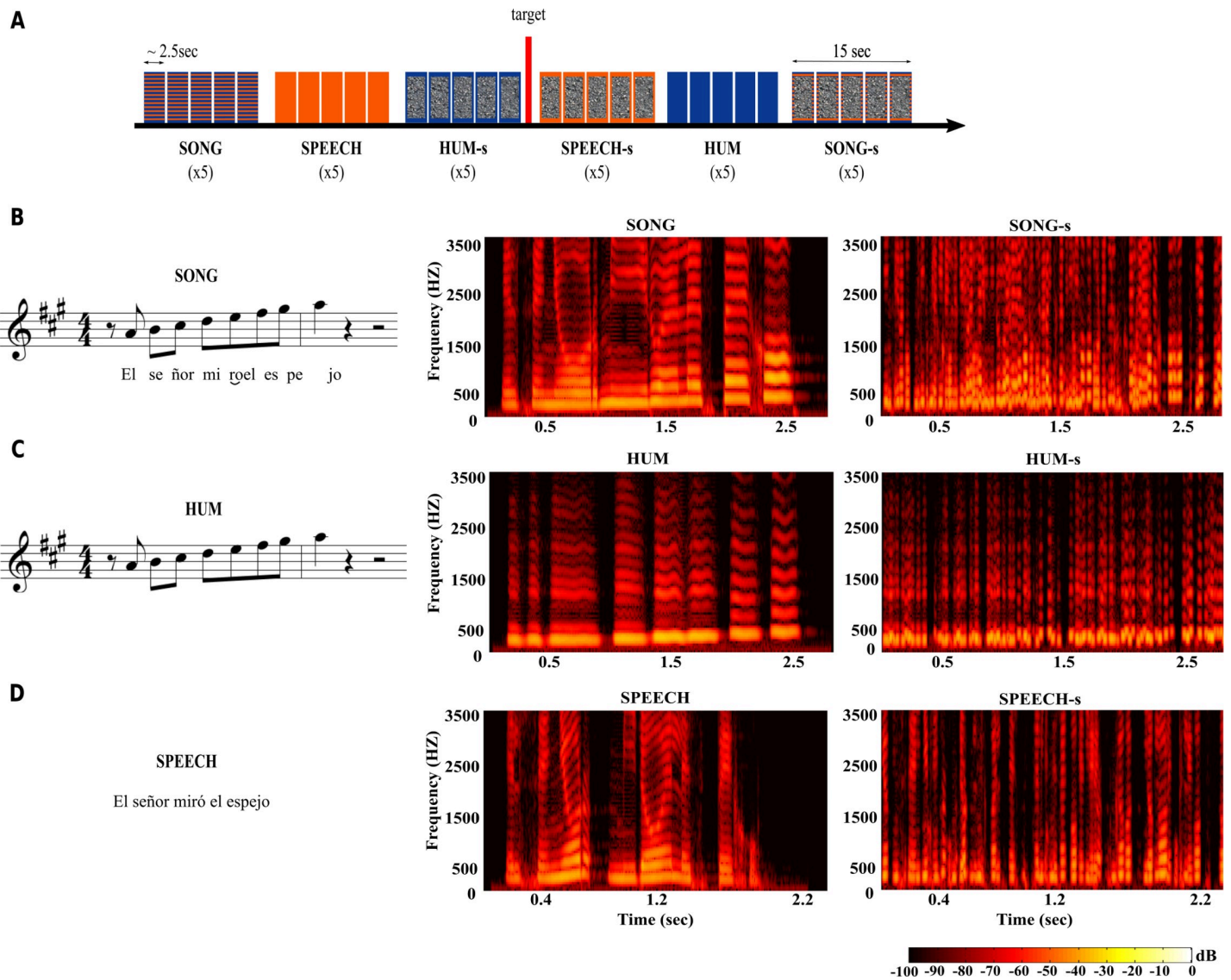
## Acoustic stimuli

**Vocal-music paradigm** (Fig 6, panel A): Auditory stimuli consisted of short excerpts of  $2.8 \pm 0.5$  seconds, normalized to -30 dB using Adobe Audition (Adobe Systems). MIRTtoolbox, implemented in MATLAB (Mathworks, Natick, MA)[76], was used for acoustic analysis.

All stimuli (divided into **Song**, **Speech** and **Hum** categories; Fig 6, panel B-D) were produced by a professional female singer after a period of training to avoid any emotional emphasis during production (e.g., affective prosody or emotional intonation). The sentences used in the **Song** and **Speech** categories were novel and carefully selected from a pool of 80 phrases in a pilot test, where 30 listeners (15 women), who did not participate in the main experiment, rated the emotional valence (e.g., from very emotional to neutral) and the complexity of each stimulus. Finally, we selected for use in the imaging experiments the 35 sentences considered the most neutral and simple, both in their grammatical structure and meaning (e.g., “La alfombra está en la sala”—“The rug is in the living room”).

1. **Hum**: This category included 25 novel musical sequences that we had previously used [19,77]; all melodies followed rules of Western tonal music. The singer hummed these melodies with her mouth closed (i.e., no syllable was used). Each block in this category consisted of five different melodies.
2. **Speech**: 35 Spanish sentences were included in this category. Given the slightly shorter duration of spoken sentences as compared to their sung versions, seven phrases were presented per block.
3. **Song**: The same 25 musical sequences from the **Hum** category were used as melodies to produce the sung versions of 25 sentences used in the speech condition. Five songs were included per block.

**Scrambled stimuli** (Fig 6, panel B-D): Scrambled versions for the three main categories. Small fragments (50 ms) with 50% overlap were randomly repositioned temporally within an interval of one second using a freely available Matlab toolbox (<http://www.ee.columbia.edu/~>



**Fig 6. Auditory stimulation paradigm.** Each block (~15 sec) included 5–7 different stimuli from the same category (A). The paradigm included 5 blocks of each of the 6 sound categories, for a total duration of 10 min. Blue indicates musical sounds, orange for speech stimuli, and scrambled gray indicates the scrambled version of each category (e.g., Hum-S). The target sound was presented between two randomly selected blocks on five occasions throughout the paradigm. (B–D) Example stimulus in three versions: Song (B), Hum (C) and Speech (D). All stimuli were produced by the same singer. Spectrograms show the frequency structure (y-axis) over time (x-axis), with colors representing the relative amplitude of each frequency band. Left column shows natural categories; right column shows the temporally scrambled condition. “El señor miró el espejo”: “The man looked at the mirror”.

<https://doi.org/10.1371/journal.pone.0222796.g006>

[dpwe/resources/matlab/scramble/](https://doi.org/10.1371/journal.pone.0222796.g006)). This procedure retained low level acoustical attributes (i.e., pitch, duration, loudness, and timbre) but rendered the stimuli unintelligible by disrupting their temporal organization (i.e., melody and rhythm) and therefore, their high-level perceptual and cognitive properties. The scrambled counterparts of each of the original sound categories are identified as: 4) Hum-S, 5) Speech-S and 6) Song-S.

### Image acquisition

All images were acquired at the National Laboratory for Magnetic Resonance Imaging using a 3T Discovery MR750 scanner (General Electric, Waukesha, Wisconsin) with a 32-channel

coil. Functional volumes consisted of 50 slices (3 mm thick), acquired with a gradient-echo, echo-planar imaging sequence with the following parameters: field of view (FOV) = 256×256 mm<sup>2</sup>, matrix size = 128×128 (voxel size = 2×2×3 mm<sup>3</sup>), TR = 3000 ms, TE = 40 ms. To improve image registration we also acquired a 3D T1-weighted volume with the following characteristics: voxel size = 1×1×1 mm<sup>3</sup>, TR = 2.3 s, TE = 3 ms. The total duration of the experiment was 18 minutes. All imaging data are freely accessible at <https://openneuro.org/datasets/ds001482>.

## Image processing and statistical analyses

Anatomical and functional images were preprocessed using *fsl* tools (version 5.0.9, fMRIB, Oxford UK). Images were corrected for movement and smoothed using a 5-mm FWHM Gaussian kernel; spatial normalization was performed using the MNI-152 standard template as reference. fMRI data analysis was conducted using FEAT (FMRI Expert Analysis Tool) version 6.00. Statistical analysis was performed using the general linear model. For each subject (first level analysis) the three original acoustic categories were modeled as explanatory variables (EV), along with their three scrambled counterparts (Analysis 1). Target stimuli were included as a nuisance regressor. Statistical maps of between-category differences for each subject were generated using a fixed-effects model, and the resulting contrasts were entered in a random-effects model for between-subject analyses using FLAME [78,79]; musical expertise was included as a group factor at this level. We used random field theory [80] to correct for multiple comparisons (voxel  $z > 2.3$ , cluster  $p < 0.05$ ) unless otherwise specified.

## Analyses

**Analysis 1 (natural versus scrambled stimuli):** By disrupting the global perception of the stimuli through temporal scrambling, while leaving low-level acoustic features untouched, we searched for brain areas that showed greater activation by natural stimuli compared to their scrambled counterparts in each category. The contrasts included were: 1.1 **Song vs. Song-S**; 1.2 **Hum vs. Hum-S** and 1.3 **Speech vs. Speech-S**.

**Analysis 2 (Song versus Speech):** We searched for differences in brain activity in response to listening to song and speech stimuli, both of which include semantic information but differ in temporal (e.g. rhythm) and spectral (e.g. pitch modulation) content. The contrasts included were: 2.1 **Song>Song-S vs. Speech>Speech-S**; 2.2 **Speech>Speech-S vs. Song>Song-S**.

**Analysis 3 (Hum versus Speech):** These two categories differ in terms of temporal, spectral and semantic content. The contrasts included were: 3.1 **(Hum>Hum-S) vs. (Speech>Speech-S)**; 3.2 **(Speech>Speech-S vs. (Hum>Hum-S))**.

**Analysis 4 (Hum versus Song):** This analysis aimed to find differences in brain activity in response to listening to two categories of melodic sounds that differed in semantic content. The contrasts included were: 4.1 **(Hum>Hum-S) vs. (Song>Song-S)**; 4.2 **(Song>Song-S) vs. (Hum>Hum-S)**.

**Analysis 5 (musicians versus non-musicians):** we evaluated differences between groups to identify whether musical training changes the patterns of activation during the perception of different types of human vocalizations. We included the following comparisons: 5.1 **(Song>Song-S) vs. (Speech>Speech-S) | musicians > non-musicians**; 5.2 **(Speech>Speech-S) vs. (Song>Song-S) | musicians > non-musicians**; 5.3 **(Hum>Hum-S) vs. (Speech>Speech-S) | musicians > non-musicians**; 5.4 **(Speech>Speech-S) vs. (Hum>Hum-S) | musicians > non-musicians**; 5.5 **(Song>Song-S) vs. (Hum>Hum-S) | musicians > non-musicians**; 5.6 **(Song vs. Hum) | musicians > non-musicians**.

**Analysis 6 (Functional characterization of the Superior Temporal Gyrus):** Finally, we also performed an independent ROI (region of interest) analysis of the *planum polare*, *planum temporale* and Heschl's gyrus. ROIs were derived from the Harvard-Oxford Probabilistic Anatomical Atlas thresholded at 33%; statistical significance threshold was set at  $p < 0.0028$ , considering eighteen comparisons were performed (i.e., three categories in six ROIs). This analysis explored BOLD signal modulations to melodic and non-melodic vocal sounds in three different auditory regions.

## Supporting information

**S1 Fig. Simple contrasts for the three original categories (Song, Hum and Speech).** Left panel shows **Song vs Speech**; Middle panel **Hum vs Speech**; and Right panel shows **Song vs Hum**. Color codes are similar to Figs 2, 3 and 4. Statistical maps are overlaid on the MNI-152 atlas. MNI coordinates of each slice are expressed in mm.

(TIF)

**S2 Fig. Box plot of acoustic characteristics across categories.** Horizontal bars represent the median; black boxes show the interquartile range; vertical lines show the data range excluding outliers (dots).

(TIF)

**S1 Table. Musical training information.**

(PDF)

## Acknowledgments

We are grateful to all the volunteers who made this study possible. We also thank Pézel Flores for helping us to select, evaluate and record all the vocal stimuli; Daniel Ramírez Peña for his valuable help with the validation and preprocessing of all the acoustic stimuli; Edgar Morales Ramírez for his permanent technical assistance; Leopoldo González-Santos, Erick Pasaye, Juan Ortíz-Retana and the support team from the Magnetic Resonance Unit (UNAM, Juriquilla), for technical assistance; and Dorothy Pless for proof-reading and editing our manuscript. Finally, we thank the authorities of the Music Conservatory “José Guadalupe Velázquez”, the Querétaro School of Violin Making, and the Querétaro Fine Arts Institute for their help with the recruitment of musicians. The authors declare no competing financial interests.

## Author Contributions

**Conceptualization:** Arafat Angulo-Perkins, Luis Concha.

**Data curation:** Arafat Angulo-Perkins, Luis Concha.

**Formal analysis:** Arafat Angulo-Perkins, Luis Concha.

**Funding acquisition:** Arafat Angulo-Perkins, Luis Concha.

**Investigation:** Arafat Angulo-Perkins, Luis Concha.

**Methodology:** Arafat Angulo-Perkins, Luis Concha.

**Project administration:** Arafat Angulo-Perkins, Luis Concha.

**Resources:** Arafat Angulo-Perkins, Luis Concha.

**Supervision:** Luis Concha.

**Validation:** Arafat Angulo-Perkins, Luis Concha.

**Visualization:** Arafat Angulo-Perkins, Luis Concha.

**Writing – original draft:** Arafat Angulo-Perkins, Luis Concha.

**Writing – review & editing:** Arafat Angulo-Perkins, Luis Concha.

## References

1. Patel AD. Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Front Psychol* [Internet]. 2011 [cited 2019 Feb 28];2. Available from: <http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00142/abstract>
2. Patel AD. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear Res*. 2014 Feb; 308:98–108. <https://doi.org/10.1016/j.heares.2013.08.011> PMID: 24055761
3. Peretz I, Vuvan D, Lagrois M-É, Armony JL. Neural overlap in processing music and speech. *Philos Trans R Soc Lond B Biol Sci*. 2015 Mar 19; 370(1664):20140090. <https://doi.org/10.1098/rstb.2014.0090> PMID: 25646513
4. Pinker S. *How the Mind Works*. W. W. Norton & Company; 1999. 672 p.
5. Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G, Friederici AD. Bach speaks: a cortical “language-network” serves the processing of music. *NeuroImage*. 2002 Oct; 17(2):956–66. PMID: 12377169
6. Merker B. Synchronous chorusing and the origins of music. *Music Sci*. 1999 Sep; 3(1\_suppl):59–73.
7. Schmithorst VJ. Separate cortical networks involved in music perception: preliminary functional MRI evidence for modularity of music processing. *NeuroImage*. 2005 Apr 1; 25(2):444–51. <https://doi.org/10.1016/j.neuroimage.2004.12.006> PMID: 15784423
8. Schön D, Gordon R, Campagne A, Magne C, Astésano C, Anton J-L, et al. Similar cerebral networks in language, music and song perception. *NeuroImage*. 2010 May 15; 51(1):450–61. <https://doi.org/10.1016/j.neuroimage.2010.02.023> PMID: 20156575
9. Schön D, Magne C, Besson M. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*. 2004; 41(3):341–9. <https://doi.org/10.1111/1469-8986.00172.x> PMID: 15102118
10. Zatorre RJ, Gandour JT. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci*. 2008 Mar 12; 363(1493):1087–104. <https://doi.org/10.1098/rstb.2007.2161> PMID: 17890188
11. Fitch WT. Four principles of bio-musicology. *Philos Trans R Soc B Biol Sci* [Internet]. 2015 Mar 19 [cited 2018 Sep 6]; 370(1664). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4321132/>
12. Fitch WT. Dance, Music, Meter and Groove: A Forgotten Partnership. *Front Hum Neurosci* [Internet]. 2016 [cited 2018 Apr 10]; 10. Available from: <https://www.frontiersin.org/articles/10.3389/fnhum.2016.00064/full>
13. Honing H. Without it no music: beat induction as a fundamental musical trait. *Ann N Y Acad Sci*. 2012 Apr 1; 1252(1):85–91.
14. Sergent J, Zuck E, Terriah S, MacDonald B. Distributed neural network underlying musical sight-reading and keyboard performance. *Science*. 1992 Jul 3; 257(5066):106–9. <https://doi.org/10.1126/science.1621084> PMID: 1621084
15. Corbeil M, Trehub SE, Peretz I. Speech vs. singing: infants choose happier sounds. *Front Psychol*. 2013; 4:372. <https://doi.org/10.3389/fpsyg.2013.00372> PMID: 23805119
16. Bispham JC. Music’s “design features”: Musical motivation, musical pulse, and musical pitch. *Music Sci*. 2009 Sep 1; 13(2\_suppl):41–61.
17. Tillmann B. PITCH PROCESSING IN MUSIC AND SPEECH. *Acoust Aust*. 7.
18. Aichert I, Späth M, Ziegler W. The role of metrical information in apraxia of speech. Perceptual and acoustic analyses of word stress. *Neuropsychologia*. 2016 Feb 1; 82:171–8. <https://doi.org/10.1016/j.neuropsychologia.2016.01.009> PMID: 26792367
19. Angulo-Perkins A, Aubé W, Peretz I, Barrios FA, Armony JL, Concha L. Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex J Devoted Study Nerv Syst Behav*. 2014 Aug 12; 59C:126–37.



20. Norman-Haignere S, Kanwisher NG, McDermott JH. Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*. 2015 Dec; 88(6):1281–96. <https://doi.org/10.1016/j.neuron.2015.11.035> PMID: 26687225
21. Rogalsky C, Rong F, Saberi K, Hickok G. Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J Neurosci Off J Soc Neurosci*. 2011 Mar 9; 31(10):3843–52.
22. Leaver AM, Rauschecker JP. Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. *J Neurosci*. 2010 Jun 2; 30(22):7604–12. <https://doi.org/10.1523/JNEUROSCI.0296-10.2010> PMID: 20519535
23. Abrams DA, Bhatara A, Ryali S, Balaban E, Levitin DJ, Menon V. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb Cortex*. 2011 Jul; 21(7):1507–18. <https://doi.org/10.1093/cercor/bhq198> PMID: 21071617
24. Merrill J, Sammler D, Bangert M, Goldhahn D, Lohmann G, Turner R, et al. Perception of Words and Pitch Patterns in Song and Speech. *Front Psychol [Internet]*. 2012 [cited 2019 Mar 12]; 3. Available from: <https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00076/full>
25. Fitch WT. The biology and evolution of music: a comparative perspective. *Cognition*. 2006 May; 100(1):173–215. <https://doi.org/10.1016/j.cognition.2005.11.009> PMID: 16412411
26. MacLarnon AM, Hewitt GP. The evolution of human speech: the role of enhanced breathing control. *Am J Phys Anthropol*. 1999 Jul; 109(3):341–63. [https://doi.org/10.1002/\(SICI\)1096-8644\(199907\)109:3<341::AID-AJPA5>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2) PMID: 10407464
27. d'Errico F, Henshilwood C, Lawson G, Vanhaeren M, Tillier A-M, Soressi M, et al. Archaeological evidence for the emergence of language, symbolism, and music—An alternative multidisciplinary perspective. *J World Prehistory*. 2003; 17(1):1–70.
28. Souza JD. Voice and Instrument at the Origins of Music. *Curr Musicol*. 16.
29. Molino J. Toward an evolutionary theory of music and language. In: *The origins of music*. Cambridge, MA, US: The MIT Press; 2000. p. 165–76.
30. Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res*. 2002 Feb; 13(1):17–26. PMID: 11867247
31. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature*. 2000 Jan 20; 403(6767):309–12. <https://doi.org/10.1038/35002078> PMID: 10659849
32. Callan DE, Tsytarev V, Hanakawa T, Callan AM, Katsuhara M, Fukuyama H, et al. Song and speech: Brain regions involved with perception and covert production. *NeuroImage*. 2006 Jul; 31(3):1327–42. <https://doi.org/10.1016/j.neuroimage.2006.01.036> PMID: 16546406
33. Ozdemir E, Norton A, Schlaug G. Shared and distinct neural correlates of singing and speaking. *NeuroImage*. 2006 Nov 1; 33(2):628–35. <https://doi.org/10.1016/j.neuroimage.2006.07.013> PMID: 16956772
34. Brown S, Martinez MJ, Hodges DA, Fox PT, Parsons LM. The song system of the human brain. *Cogn Brain Res*. 2004 Aug; 20(3):363–75.
35. Sammler D, Grosbras M-H, Anwender A, Bestelmeyer PEG, Belin P. Dorsal and Ventral Pathways for Prosody. *Curr Biol CB*. 2015 Dec 7; 25(23):3079–85. <https://doi.org/10.1016/j.cub.2015.10.009> PMID: 26549262
36. Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci*. 2003 Jul 1; 15(5):673–82. <https://doi.org/10.1162/089892903322307393> PMID: 12965041
37. Sammler D, Baird A, Valabrègue R, Clément S, Dupont S, Belin P, et al. The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J Neurosci Off J Soc Neurosci*. 2010 Mar 10; 30(10):3572–8.
38. Ito T, Tiede M, Ostry DJ. Somatosensory function in speech perception. *Proc Natl Acad Sci U S A*. 2009 Jan 27; 106(4):1245–8. <https://doi.org/10.1073/pnas.0810063106> PMID: 19164569
39. Tierney A, Dick F, Deutsch D, Sereno M. Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb Cortex N Y N* 1991. 2013 Feb; 23(2):249–54.
40. Kaas JH, Hackett TA. Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A*. 2000 Oct 24; 97(22):11793–9. <https://doi.org/10.1073/pnas.97.22.11793> PMID: 11050211
41. Okada K, Rong F, Venezia J, Matchin W, Hsieh I-H, Saberi K, et al. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex*. 2010 Oct; 20(10):2486–95. <https://doi.org/10.1093/cercor/bhp318> PMID: 20100898

42. Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci*. 2001 Jan 1; 13(1):1–7. PMID: [11224904](#)
43. Pantev C, Roberts LE, Schulz M, Engelien A, Ross B. Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*. 2001 Jan 22; 12(1):169. <https://doi.org/10.1097/00001756-200101220-00041> PMID: [11201080](#)
44. Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. Increased auditory cortical representation in musicians. *Nature*. 1998 Apr; 392(6678):811–4. <https://doi.org/10.1038/33918> PMID: [9572139](#)
45. Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of Neuroplastic P2 and N1c Auditory Evoked Potentials in Musicians. *J Neurosci*. 2003 Jul 2; 23(13):5545–52. <https://doi.org/10.1523/JNEUROSCI.23-13-05545.2003> PMID: [12843255](#)
46. Shahin AJ, Roberts LE, Chau W, Trainor LJ, Miller LM. Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*. 2008 May 15; 41(1):113–22. <https://doi.org/10.1016/j.neuroimage.2008.01.067> PMID: [18375147](#)
47. Fauvel B, Groussard M, Chételat G, Fouquet M, Landeau B, Eustache F, et al. Morphological brain plasticity induced by musical expertise is accompanied by modulation of functional connectivity at rest. *NeuroImage*. 2014 Apr 15; 90:179–88. <https://doi.org/10.1016/j.neuroimage.2013.12.065> PMID: [24418502](#)
48. Grahn JA, Rowe JB. Feeling the Beat: Premotor and Striatal Interactions in Musicians and Nonmusicians during Beat Perception. *J Neurosci*. 2009 Jun 10; 29(23):7540–8. <https://doi.org/10.1523/JNEUROSCI.2018-08.2009> PMID: [19515922](#)
49. Strait DL, Chan K, Ashley R, Kraus N. Specialization among the specialized: Auditory brainstem function is tuned in to timbre. *Cortex J Devoted Study Nerv Syst Behav*. 2012; 48(3):360–2.
50. Herholz SC, Zatorre RJ. Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*. 2012 Nov 8; 76(3):486–502. <https://doi.org/10.1016/j.neuron.2012.10.011> PMID: [23141061](#)
51. Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, et al. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex N Y N* 1991. 2000 May; 10(5):512–28.
52. Hickok, Poeppel. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci*. 2000 Apr; 4(4):131–8. PMID: [10740277](#)
53. Buchsbaum BR, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn Sci*. 2001; 25(5):663–78.
54. Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci*. 2003 Jul 1; 15(5):673–82. <https://doi.org/10.1162/089892903322307393> PMID: [12965041](#)
55. Levitin DJ, Menon V. Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*. 2003 Dec; 20(4):2142–52. PMID: [14683718](#)
56. Roskies AL, Fiez JA, Balota DA, Raichle ME, Petersen SE. Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *J Cogn Neurosci*. 2001 Aug 15; 13(6):829–43. <https://doi.org/10.1162/08989290152541485> PMID: [11564326](#)
57. Morosan P, Schleicher A, Amunts K, Zilles K. Multimodal architectonic mapping of human superior temporal gyrus. *Anat Embryol (Berl)*. 2005 Dec; 210(5–6):401–6. <https://doi.org/10.1007/s00429-005-0029-1> PMID: [16170539](#)
58. Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, et al. Functional properties of human auditory cortical fields. *Front Syst Neurosci*. 2010; 4:155. <https://doi.org/10.3389/fnsys.2010.00155> PMID: [21160558](#)
59. Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M. Human primary auditory cortex follows the shape of Heschl’s gyrus. *J Neurosci Off J Soc Neurosci*. 2011 Oct 5; 31(40):14067–75.
60. Nudds M. What Are Auditory Objects? *Rev Philos Psychol*. 2007; 1(1):105–122.
61. Humphries C, Liebenthal E, Binder JR. Tonotopic organization of human auditory cortex. *NeuroImage*. 2010 Apr 15; 50(3):1202–11. <https://doi.org/10.1016/j.neuroimage.2010.01.046> PMID: [20096790](#)
62. Skipper J. Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philos Trans R Soc Lond B Biol Sci*. 2014 Sep 19; 369.
63. Griffiths TD, Büchel C, Frackowiak RS, Patterson RD. Analysis of temporal structure in sound by the human brain. *Nat Neurosci*. 1998 Sep; 1(5):422–7. <https://doi.org/10.1038/1637> PMID: [10196534](#)

64. Kleber B, Veit R, Birbaumer N, Gruzelier J, Lotze M. The brain of opera singers: experience-dependent changes in functional activation. *Cereb Cortex N Y N 1991*. 2010 May; 20(5):1144–52.
65. Bookheimer S. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci*. 2002; 25:151–88. <https://doi.org/10.1146/annurev.neuro.25.1.12701.142946> PMID: 12052907
66. Griffiths TD, Warren JD. The planum temporale as a computational hub. *Trends Neurosci*. 2002 Jul; 25(7):348–53. [https://doi.org/10.1016/s0166-2236\(02\)02191-4](https://doi.org/10.1016/s0166-2236(02)02191-4) PMID: 12079762
67. Chen JL, Zatorre RJ, Penhune VB. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *NeuroImage*. 2006 Oct 1; 32(4):1771–81. <https://doi.org/10.1016/j.neuroimage.2006.04.207> PMID: 16777432
68. Chen JL, Penhune VB, Zatorre RJ. Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex N Y N 1991*. 2008 Dec; 18(12):2844–54.
69. Grahn JA, Brett M. Rhythm and beat perception in motor areas of the brain. *J Cogn Neurosci*. 2007 May; 19(5):893–906. <https://doi.org/10.1162/jocn.2007.19.5.893> PMID: 17488212
70. Merchant H, Honing H. Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front Neurosci [Internet]*. 2014 [cited 2018 Sep 7]; 7. Available from: <https://www.frontiersin.org/articles/10.3389/fnins.2013.00274/full>
71. Manning F, Schutz M. “Moving to the beat” improves timing perception. *Psychon Bull Rev*. 2013 Dec; 20(6):1133–9. <https://doi.org/10.3758/s13423-013-0439-7> PMID: 23670284
72. Vuust P, Witek MAG. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Front Psychol*. 2014; 5:1111. <https://doi.org/10.3389/fpsyg.2014.01111> PMID: 25324813
73. Lima CF, Krishnan S, Scott SK. Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. *Trends Neurosci*. 2016 Aug; 39(8):527–42. <https://doi.org/10.1016/j.tins.2016.06.003> PMID: 27381836
74. Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci*. 2010 Mar; 1191:62–88. <https://doi.org/10.1111/j.1749-6632.2010.05444.x> PMID: 20392276
75. Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*. 2012 Aug 15; 62(2):816–47. <https://doi.org/10.1016/j.neuroimage.2012.04.062> PMID: 22584224
76. Lartillot O, Toiviainen P. A Matlab Toolbox for Musical Feature Extraction from Audio. 2007; 8.
77. Aubé W, Angulo-Perkins A, Peretz I, Concha L, Armony JL. Fear across the senses: brain responses to music, vocalizations and facial expressions. *Soc Cogn Affect Neurosci*. 2014 May 1;
78. Beckmann CF, Jenkinson M, Smith SM. General multilevel linear modeling for group analysis in FMRI. *NeuroImage*. 2003 Oct; 20(2):1052–63. [https://doi.org/10.1016/S1053-8119\(03\)00435-X](https://doi.org/10.1016/S1053-8119(03)00435-X) PMID: 14568475
79. Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*. 2004 Apr; 21(4):1732–47. <https://doi.org/10.1016/j.neuroimage.2003.12.023> PMID: 15050594
80. Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC. Detecting changes in nonisotropic images. *Hum Brain Mapp*. 1999; 8(2–3):98–101. PMID: 10524599