TITLE: Cortical music selectivity does not require musical training

ABBREVIATED TITLE: Music selectivity in musicians and non-musicians

AUTHORS:

Dana Boebinger

Massachusetts Institute of Technology, Cambridge MA 02139 USA

Designed research, Performed research, Analyzed data, Wrote the paper

Sam Norman-Haignere

Zuckerman Institute for Brain Research, Columbia University, New York NY 10027 USA

Designed research, Contributed unpublished analytic tools, Edited the paper

Josh McDermott

Massachusetts Institute of Technology, Cambridge MA 02139 USA

Designed research, Edited the paper

Nancy Kanwisher

Massachusetts Institute of Technology, Cambridge MA 02139 USA

Designed research, Edited the paper

Corresponding author: Nancy Kanwisher (ngk@mit.edu)

Journal Section: Behavioral/Cognitive

Conflicts of Interest: The authors declare no competing financial interests.

Number of pages: 24

Number of figures: 10

Number of words:

Abstract: 248

Introduction: 889

Discussion: 1212

Cortical music selectivity does not require musical training

Dana Boebinger[1,2,3], Sam Norman-Haignere[5,6,7], Josh McDermott[1,2,3,4], Nancy Kanwisher[2,3,4]

[1] Speech and Hearing Bioscience and Technology, Harvard University, Cambridge MA 02138 USA

[2] Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA 02139 USA

[3] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge MA 02139 USA

[4] Center for Brains, Minds, and Machines, Massachusetts Institute of Technology, Cambridge MA 02139 USA

[5] HHMI Postdoctoral Fellow of the Life Sciences Research Foundation, Baltimore MD 21218 USA

[6] Laboratoire des Sytèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, CNRS, Paris France

[7] Zuckerman Institute for Brain Research, Columbia University, New York NY 10027 USA

ABSTRACT

Human auditory cortex contains neural populations that respond strongly to a wide variety of music sounds, but much less strongly to sounds with similar acoustic properties or to other real-world sounds. However, it is unknown whether this selectivity for music is driven by explicit training. To answer this question, we measured fMRI responses to 192 natural sounds in 10 people with extensive musical training and 10 with almost none. Using voxel decomposition (Norman-Haignere et al., 2015) to explain voxel responses across all 20 participants in terms of a small number of components, we replicated the existence of a music-selective response component similar in tuning and anatomical distribution to our earlier report. Critically, we also estimated components separately for musicians and non-musicians and found that a music-selective component was clearly present even in individuals with almost no musical training, which was very similar to the music component found in musicians. We also found that musical genres that were less familiar to our participants (e.g., Mongolian throat singing) produced strong responses within the music component, as did drum clips with rhythm but little melody. These data replicate the finding of music selectivity, broaden its scope to include unfamiliar musical genres and rhythms, and show that it is robustly present in people with almost no musical training. Our findings demonstrate that musical training is not necessary for music selectivity to emerge in non-primary auditory cortex, raising the possibility that music-selective brain responses could be a universal property of human auditory cortex.

SIGNIFICANCE STATEMENT

Recent research has revealed populations of neurons in the human brain that respond more to music than to other sounds. How do these music-selective responses arise, and what range of music do they respond to? We scanned 10 expert musicians and 10 non-musicians with fMRI while they listened to a variety of music and other sounds. We found that neural populations specifically responsive to music exist to a similar degree in non-musicians and musicians alike. We further showed that these neural populations respond strongly to unfamiliar musical genres (e.g., Mongolian throat singing) and to drum clips with rhythm but little melody. These results show that neural populations selective for a wide variety of music can arise without explicit musical training.

INTRODUCTION

Music is uniquely and universally human (Mehr et al., 2019) and arises early in development (Trehub, 2003). Further, recent evidence has revealed neural populations in nonprimary auditory cortex that respond selectively to music per se (Norman-Haignere et al., 2015; see also Leaver and Rauschecker, 2010; Rogalsky et al., 2011; Fedorenko et al., 2012; LaCroix et al., 2015; Norman-Haignere et al., 2019). How do these neural mechanisms for music arise, and what is the role of experience in their development? Most members of Western societies have received at least some explicit musical training in the form of lessons or classes. However, most Western individuals, including non-musicians, are believed to implicitly acquire knowledge of musical structure from a lifetime of exposure to music (Bigand, 1983; Bigand and Pineau, 1997; Koelsch et al., 2000; Tillmann et al., 2000; Tillmann, 2005; Bigand and Poulin-Charronnat, 2006), raising the possibility that the music-selective responses in the auditory cortex might also not depend strongly on explicit musical training. Here, we directly test whether music-selective neural responses result from or are modulated by musical training, by comparing the magnitude, anatomical location, and selectivity of these responses in individuals with extensive musical training, versus individuals with almost none.

Why might explicit musical training change the response properties of auditory cortex? In animals, exposure to specific sounds can elicit long-term changes in auditory cortex, such as sharper tuning of individual neurons (Recanzone et al., 1993; Fritz et al., 2003; Lee and Middlebrooks, 2011) and expansion of cortical maps (Recanzone et al., 1993; Polley et al., 2006; Bieszczad and Weinberger, 2010). These changes occur only for behaviorally relevant stimulus features (Ahissar et al., 1992, 1998; Fritz et al., 2005; Ohl and Scheich, 2005; Polley et al., 2006) related to the intrinsic reward value of the stimulus (Bakin and Weinberger, 1996; Fritz et al., 2005; David et al., 2012), and thus are closely linked to the neuromodulatory system (Bao et al., 2001; Kilgard et al., 2001; Blake et al., 2006). Additionally, the extent of cortical map expansion is correlated with the animal's subsequent improvement in behavioral performance (Recanzone et al., 1993; Rutkowski and Weinberger, 2005; Polley et al., 2006; Bieszczad and Weinberger, 2010, 2012; Reed et al., 2011). Most of this prior work on experience-driven plasticity in auditory cortex has been done in animals undergoing extensive training, and it has been unclear how this might generalize to humans in more natural settings. Musical training provides a unique way to investigate this question, as it meets virtually all of these criteria for eliciting functional plasticity: playing music requires focused attention, fine-grained sensory-motor coordination, and is known to engage the neuromodulatory system (Blood and Zatorre, 2001; Salimpoor et al., 2011, 2013). And expert musicians often begin training at a young age and hone their skills over many years.

Although many prior studies have measured fMRI responses in musicians and non-musicians (Ohnishi et al., 2001; Pantev et al., 2001; Shahin et al., 2003; Fujioka et al., 2004, 2005; Besson et al., 2007; Wong et al., 2007; Dick et al., 2011; Lee and Noppeney, 2011; Ellis et al., 2012, 2013; Angulo-

Perkins et al., 2014; Doelling and Poeppel, 2015; Lappe et al., 2016), it remains unclear whether extensive musical training changes the tuning properties of auditory cortex. Previous studies have found that fMRI responses to music are larger in musicians compared to non-musicians in posterior superior temporal gyrus (Ohnishi et al., 2001; Dick et al., 2011; Angulo-Perkins et al., 2014). However, these responses were not shown to be selective for music, and the increased responses in musicians could simply reflect increased attention to music, rather than increased neural selectivity for music *per se*.

The fact that prior studies have not observed group differences in music selectivity is perhaps unsurprising, as we have previously found that music selectivity is weak when measured in raw voxel responses using standard voxel-wise fMRI analyses, due to spatial overlap between music-selective neural populations and neural populations with other selectivities (e.g. pitch). To overcome these challenges, Norman-Haignere et al. (2015) introduced a voxel decomposition method that models each voxel in auditory cortex as a weighted sum of a small number of canonical response profiles ("components", each presumably reflecting a distinct neural population) across a large set of natural sounds. This method makes it possible to disentangle the response of neural populations that overlap within voxels, and has previously revealed a neural population with clear selectivity for music compared to both other real-world sounds (Norman-Haignere et al., 2015) and synthetic control stimuli matched in standard acoustic properties (Norman-Haignere and McDermott, 2018). These results have recently been confirmed by intracranial recordings, which show individual electrodes with clear selectivity for music (Norman-Haignere et al., 2019). Although the Norman-Haignere et al. (2015) study did not include actively practicing musicians (none had musical training in the five years preceding the study), many of the participants had substantial musical training earlier in their lives.

Our main goal in the current study was to ask whether the music selectivity reported by Norman-Haignere et al. (2015) is a widespread property of normal human brains, or whether it primarily arises due to explicit musical training. Using these same methods, we were also able to test whether the inferred music-selective neural population responds strongly to less familiar musical genres (e.g. Balinese *gamelan*), and to drum clips with rich rhythm but little melody.

MATERIALS & METHODS

Participants

Norman-Haignere et al. (2015) found a music-selective component in an analysis of ten participants. However, although these participants were described as "non-musicians," some of the participants had substantial musical training. We therefore used stricter inclusion criteria to recruit 10 musicians and 10 non-musicians for the current study, in order to have comparable statistical power within each group as in our previous study.

Twenty young adults (14 female, mean = 24.7 years) participated in the experiment: 10 musicians (8 female, mean = 23.5 years) and 10 non-musicians (6 female, mean = 25.8 years). Inclusion criteria for musicians were beginning formal training before the age of seven (Penhune, 2011), and continuing training until the current day. Our sample of ten musicians had an average of 16.30 years of training (ranging from 11-23 years, sd = 2.52). To be classified as a non-musician, participants were required to have less than two years of total music training, which could not have occurred either before the age of seven or within the last five years. Out of the ten non-musicians in our sample, eight had zero years of musical training, one had a single year of musical training (at the age of 20), and one had two years of training (starting at age 10). These training measures do not include any informal "music classes" included in participants' required elementary school curriculum.

There were no significant group differences in age, education, or socioeconomic status (t(18) = -0.845, p = 0.409). All participants were native English speakers and had normal hearing (audiometric thresholds <25 dB HL for octave frequencies 250Hz to 8kHz). The study was approved by MIT's human participants review committee (COUHES), and written informed consent was obtained from all participants.

Study design

Each participant underwent a 2-hour behavioral testing session as well as three 2-hour fMRI scanning sessions. During the behavioral session, participants completed an audiogram, filled out questionnaires about their musical experience, and completed a series of basic psychoacoustic tasks.

Behavioral data acquisition & analysis

To validate participants' self-reported musicianship, we measured participants' abilities on a variety of psychoacoustical tasks for which prior evidence suggested that musicians would outperform non-musicians. For all psychoacoustic tasks, stimuli were presented using Psychtoolbox for Matlab (Brainard, 1997). Sounds were presented to participants at 70dB SPL over circumaural Sennheiser HD280 headphones in a soundproof booth (Industrial Acoustics). After each trial, participants were given feedback about whether or not they had answered correctly. Group differences for each task were measured using 2-sample t-tests.

Pure tone frequency discrimination. Because musicians have superior frequency discrimination abilities when compared to non-musicians (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001; Micheyl et al., 2006), we first measured participants' pure tone frequency discrimination thresholds using an adaptive two-alternative forced choice (2AFC) task. In each trial, participants heard two pairs of tones. One of the tone pairs consisted of two identical 1 kHz tones, while the other tone pair contained a 1 kHz tone and a second tone of a different frequency. Participants determined which tone interval contained the frequency change. The magnitude of the frequency difference was varied adaptively using a 1-up 3-down procedure, which continued until 10 reversals had been measured. The frequency difference was changed initially by a factor of two, which was reduced to a factor of $\sqrt{2}$ after the fourth reversal.

Multiple threshold estimations were obtained per participant (3 threshold estimations for the first seven participants, and 5 for the remaining 13 participants), and then averaged.

Synchronized tapping to an isochronous beat. Sensorimotor abilities are crucial to musicianship, and finger tapping tasks show some of the most reliable effects of musicianship (Repp, 2005, 2010; Bailey and Penhune, 2010). Participants were asked to tap along with an isochronous click track. They heard ten 30-second click blocks, separated by 5 seconds of silence. The blocks varied widely in tempo, with interstimulus intervals ranging from 200ms to 1 second (tempos of 60 to 300 bpm). Each tempo was presented twice, and the order of tempi was permuted across participants. We recorded the timing of participants' responses using a tapping sensor we constructed and have used in previous studies (e.g. Jacoby and McDermott, 2017; Polak et al., 2018). We then calculated the difference between participants' response onsets and the actual stimulus onsets. As a measure of sensorimotor synchronization ability, we took the standard deviation of these asynchronies between corresponding stimulus and response onsets.

Melody discrimination. Musicians have also been reported to outperform non-musicians on measures of melodic contour and interval discrimination (Fujioka et al., 2004; McDermott et al., 2010; McPherson and McDermott, 2018). In each trial, participants heard two five-note melodies, and were asked to judge whether the two melodies were the same or different. Melodies were composed of notes that were randomly drawn from a log uniform distribution of semitone steps from 150Hz to 270Hz. The second melody was transposed up by half an octave and was either identical to the first melody or contained a single note had that had been altered either up or down by 1 or 2 semitones. Half of the trials contained a second melody that was the same as the first melody, while 25% contained a pitch change that preserved the melodic contour and the remaining 25% contained a pitch change that violated the melodic contour. There were 20 trials per condition (same/different melody x same/different contour x 1/2 semitone change), for a total of 160 trials. This task was modified from McPherson and McDermott (2018).

<u>"Sour note" detection.</u> To measure participants' knowledge of Western music, we also measured participants' ability to determine whether a melody conforms to the rules of Western music theory. The melodies used in this experiment were randomly generated from a probabilistic generative model of Western tonal melodies that creates a melody on a note-by-note basis according to the principles that (1) melodies tend to be limited to a narrow pitch range, (2) note-to-note intervals tend to be small, and (3) the notes within the melody conform to a single key (Temperley, 2008). In each trial of this task, participants heard a 16-note melody and were asked to determine whether the melody contained an out-of-key ("sour") note. In half of the trials, one of the notes in the melody was modified so that it was rendered out of key. The modified notes were always scale degrees 1, 3, or 5 and they were modified by either 1 or 2 semitones accordingly so that they were out of key (i.e. scale degrees 1 and 5 were modified by 1 semitone, and scale degree 3 was modified by 2 semitones). Participants judged whether the melody contained a sour note (explained as a "mistake in the melody"). There were 20 trials per condition (modified or not x 3 scale degrees), for a total of 120 trials. This task was modified from McPherson and McDermott (2018).

Natural sound stimuli for fMRI Experiment

Stimuli consisted of 2-second clips of 192 familiar natural sounds. These sounds included the 165-sound stimulus set used in Norman-Haignere et al. (2015), which was designed to include the most frequently heard and recognizable sounds in everyday life. Examples can be seen in **Figure 1A**. This stimulus set was then supplemented with 27 additional music and drumming clips from a variety of musical cultures, so that we could examine responses to rhythmic features of music, as well as comparing responses to more versus less familiar musical genres. Stimuli were normalized (RMS =

0.1) and ramped on and off with a 25ms linear ramp. During scanning, auditory stimuli were presented over MR-compatible earphones (Sensimetrics S14) at 75 dB SPL.

An online experiment (via Amazon's Mechanical Turk) was used to assign a semantic category to each stimulus, in which 180 participants (95 females; mean age = 38.8 years, sd = 11.9 years) categorized each stimulus into one of fourteen different categories. The categories were taken from Norman-Haignere et al. (2015), with three additional categories ("non-Western instrumental music," "non-Western vocal music," "drums") added to accommodate the additional music stimuli used in this experiment.

A second Amazon Mechanical Turk experiment was run to compare the Western and non-Western music stimuli used in this experiment. In this experiment, 188 participants (75 females; mean age = 36.6 years, sd = 10.5 years) listened to each of the 62 music stimuli and rated them based on (1) how "musical" they sounded, (2) how "familiar" they sounded, (3) how much they "liked" the stimulus, and (4) how "foreign" they sounded.

Figure 1

A. Stimulus Set of Commonly Heard Natural Sounds

1. Man speaking
2. Flushing toilet
3. Pouring liquid
4. Tooth brushing
5. Woman speaking
6. Car accelerating
7. Biting and chewing
8. Laughing
9. Typing
10. Car engine running
11. Running water
12. Breathing
13. Keys jangling
14. Dishes clanking
15. Ringtone
16. Microwave
17. Dog barking
18. Walking (hard surface)
19. Road traffic
20. Zipper
21. Cellphone vibrating
22. Water dripping
23. Scratching
24. Car windows
25. Telephone ringing
26. Chopping food
27. Telephone dialing
28. Girl speaking
29. Car horn
30. Writing
31. Computer startup
32. Background speech
33. Songbird
34. Pouring water
35. Pop song
36. Water boiling
37. Guitar
38. Coughing
39. Crumpling paper
40. Siren
41. Splashing water
42. Computerized speech
43. Alarm clock
44. Walking with heels
45. Vacuum
46. Wind
47. Boy speaking
48. Chair rolling
49. Rock Song
50. Door knocking

…

B. Scanning Procedure and Task Structure

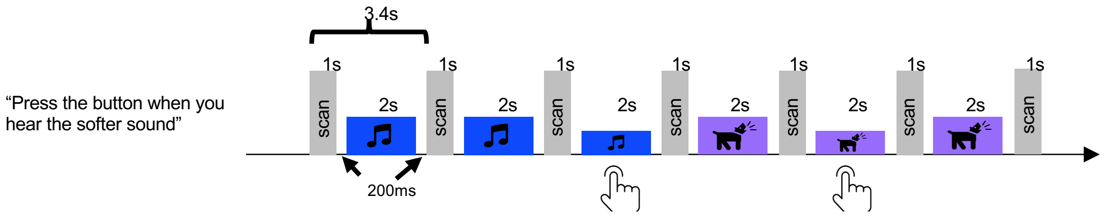

Figure 1. (**A**) Fifty examples of the original set of 165 natural sounds used in Norman-Haignere et al. (2015), ordered by how often participants' reported hearing them in daily life. An additional 27 music stimuli were added to this set for the current experiment. (**B**) Scanning paradigm and task structure. Each 2-second sound stimulus was repeated three times, with one repetition (the second or third) being 12 dB quieter. Subjects were instructed to press a button when they detected this quieter sound. A sparse scanning sequence was used, in which one fMRI volume was acquired in the silent period between stimuli.

Similar to the design of Norman-Haignere et al (2015), sounds were presented during scanning in a "mini-block design," in which each 2-second natural sound was repeated multiple times in a row. Each sound was repeated five times in a row in the previous experiment (Norman-Haignere et al., 2015), but due to the additional stimuli used in the current experiment, we repeated each sound three times in a row (which we have found yields similarly reliable responses in pilot experiments). Each stimulus was presented in silence, with a single fMRI volume collected between each repetition (i.e. "sparse scanning," Hall et al., 1999). To encourage participants to pay attention to the sounds, either the second or third repetition in each "mini-block" was 12dB quieter (presented at 67 dB SPL), and participants were instructed to press a button when they heard this quieter sound (**Figure 1B**). Overall, participants performed well on this task (musicians: mean = 92.06%, sd = 5.47%; non-musicians: mean = 91.47%, sd = 5.83%; no participant's average performance across runs fell below 80%). Each of the three scanning sessions consisted of sixteen 5.5-minute runs, for a total of 48 functional runs per participant. Each run consisted of 24 stimulus mini-blocks and five silent blocks during which no sounds were presented. These silent blocks were the same duration as the stimulus mini-blocks, and were distributed evenly throughout each run, providing a baseline. Each specific stimulus was presented in two mini-blocks per scanning session, for a total of six mini-block repetitions per stimulus over the three scanning sessions. Stimulus order was randomly permuted across runs and across participants.

MRI data were collected at the Athinoula A. Martinos Imaging Center of the McGovern Institute for Brain Research at MIT, on a 3T Siemens Prisma with a 32-channel head coil. Each volume acquisition lasted 1 second, and the 2-second stimuli were presented during periods of silence between each acquisition, with a 200ms buffer of silence before and after stimulus presentation. As a consequence, one brain volume was collected every 3.4 seconds (1 second + 2 seconds + 0.2*2 seconds) (TR = 3.4s, TA = 1.02s, TE = 33ms, 90 degree flip angle, 4 discarded initial acquisitions). Each functional acquisition consisted of 48 slices, each slice being 3mm thick and having an in-plane resolution of 2.1 x 2.1mm (96 x 96 matrix, 0.3mm slice gap). An SMS acceleration factor of 4 was used in order to minimize acquisition time (TA = 1.02s). To localize functional activity, a high-resolution anatomical T1-weighted image was obtained for every participant (TR = 2.53 seconds, voxel size: 1mm^3, 176 slices, 256 x 256 matrix).

Preprocessing and data analysis were performed using FSL software and custom Matlab scripts). Functional volumes were motion-corrected, slice-time-corrected, skull-stripped, linearly detrended, and aligned to each participant's anatomical image (using FLIRT and BBRegister; Jenkinson and Smith, 2001; Greve and Fischl, 2009). Preprocessed data were then resampled to the cortical surface reconstruction computed by FreeSurfer (Dale et al., 1999), and smoothed on the surface using a 3mm FWHM kernel to improve SNR. The data were then downsampled to a 2mm isotropic grid on the FreeSurfer-flattened cortical surface. Next, we estimated the response of this downsampled data to each of the 192 sound stimuli using a general linear model (GLM) in which each stimulus was modeled as a boxcar function convolved with the canonical hemodynamic response function (HRF). This differs from our prior paper (Norman-Haignere et al., 2015), in which signal averaging was used in place of a GLM. We made this change because responses were made more reliable using an HRF, potentially due to the use of shorter stimulus blocks causing more overlap between BOLD responses to different stimuli.

Voxel decomposition

The first step of this analysis method is to determine which voxels serve as input to the voxel decomposition algorithm. To select only the most informative voxels, all analyses were carried out on voxels within a large anatomical constraint region encompassing bilateral superior temporal and posterior parietal cortex, just as in Norman-Haignere et al. (2015). Within this region, we selected voxels that displayed a significant ($p < .001$) response to sound (pooling over all sounds compared to silence). This consisted of 51.45% of the total number of voxels within our large anatomical constraint region. We also selected only voxels that produced a reliable response pattern to the stimuli across scanning sessions. Note that rather than using a simple correlation to determine reliability, we used the equation from Norman-Haignere et al. (2015) to measure the reliability across split halves of our data. This reliability measure assigns high values to voxels that respond consistently to sounds even if their response does not vary much between sounds, which is the case for many voxels within primary auditory cortex:

$$r = 1 - \frac{\parallel \boldsymbol{v}_1 - proj_{\boldsymbol{v}_2}\boldsymbol{v}_1 \parallel^2}{\parallel \boldsymbol{v}_1 \parallel^2}$$

$$proj_{\boldsymbol{v}_2}\boldsymbol{v}_1 = \boldsymbol{v}_2\left(\frac{{\boldsymbol{v}_2}^T}{\parallel \boldsymbol{v}_2 \parallel^2}\,\boldsymbol{v}_1\right)$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ indicate the response vector of a single voxel to the 192 sounds measured in two different scans, and ‖ ‖ is the L2 norm. Note that these equations differ slightly from Equations 1 and 2 in Norman-Haignere et al. (2015). This is because the equations previously contained an error (the L2-norm terms were not squared), which has been corrected here. We used the same reliability cutoff as in our prior study ($r > 0.3$). Of the sound-responsive voxels, 54.47% of them also met the reliability criteria. Using these two selection criteria, a total of 26,792 voxels were selected for analysis (an average of 1,340 voxels per participant).

The voxel decomposition method approximates the response of each voxel as the weighted sum of a small number of component response profiles that are shared across voxels (**Figure 2A**). For example, the response of the ith voxel would be:

$$\boldsymbol{v}_i \approx \sum_{k=1}^{K} \boldsymbol{r}_k w_{k,i}$$

where $\boldsymbol{r}_k$ represents the kth component response profile that is shared across all voxels, $w_{k,i}$ represents the voxel-specific weight for that component, and K is the total number of components.

We concatenated the responses from all participants' selected voxels into a data matrix $\boldsymbol{D}$ (192 sounds × 26,792 voxels). We then approximated the data matrix as the product of two smaller matrices: (1) a response matrix $\boldsymbol{R}$ (192 sounds x K components) containing the response profile of all inferred components to the sound set, and (2) a weight matrix $\boldsymbol{W}$ (K components × 26,792 voxels) containing the contribution of each component response profile to each voxel. Using matrix notation this yields:

$$\boldsymbol{D} \approx \boldsymbol{R}\boldsymbol{W}$$

The method used to infer components has been described in detail in our previous paper (Norman-Haignere et al., 2015) and the code is available online (https://github.com/snormanhaignere/nonparametric-ica). The method is similar to standard algorithms for independent components analysis (ICA) in that it searches amongst the many possible solutions to the factorization problem for components that have a maximally non-Gaussian distribution of weights across voxels (the non-Gaussianity of the components inferred in this study can be seen in **Figure 2B**). The method differs from most standard ICA algorithms in that it does so by directly minimizing the entropy of the component weight distributions across voxels (Gaussian variables have maximum entropy for a given variance), as measured by a histogram (feasible due to the large number of voxels). The algorithm achieves this goal in two main steps. First, PCA is used to whiten and reduce the dimensionality of the data matrix. This was implemented using the singular value decomposition:

$$\boldsymbol{D} \approx \boldsymbol{U}^k \boldsymbol{S}^k \boldsymbol{V}^k$$

where $\boldsymbol{U}^k$ are the response profiles of the top K principal components (192 sounds × K components), $\boldsymbol{V}^k$ is the whitened weight matrix for these components (K components × 26,792 voxels), and $\boldsymbol{S}^k$ is a diagonal matrix of singular values ($K \times K$). The number of components (K) was chosen by measuring the amount of voxel response variance explained by different numbers of components and the accuracy of the components in predicting voxel responses in left-out data.

The principal component weight matrix is then rotated to maximize the negentropy (*J*) summed across components:

$$\hat{\boldsymbol{T}} = \underset{\boldsymbol{T}}{\operatorname{argmax}} \sum_{c=1}^{N} J(\boldsymbol{W}[c,:]), \quad \text{where } \boldsymbol{W} = \boldsymbol{TV}$$

where $\boldsymbol{W}$ is the rotated weight matrix (K × 26,792), $\boldsymbol{T}$ is an orthonormal rotation matrix (K × K), and $\boldsymbol{W}[c,:]$ is the cth row of $\boldsymbol{W}$. We estimated entropy using a histogram-based method (Moddemeijer, 1989) applied to the voxel weight vector for each component ($\boldsymbol{W}[c,:]$), and defined negentropy as the difference in entropy between the empirical weight distribution and a Gaussian distribution of the same mean and variance:

$$J(\boldsymbol{y}) = H\left(\boldsymbol{y}_{gauss}\right) - H(\boldsymbol{y})$$

Because the order the components inferred using ICA holds no significance, we reordered the components to optimally match those from Norman-Haignere et al. (2015) using the Hungarian algorithm (Kuhn, 1955). We carried out this analysis on three different data sets: i) on the data matrix created using voxel responses from all twenty participants, ii) on the data matrix consisting of data from the 10 musicians only, and iii) on the data matrix consisting of data from the 10 non-musicians only.

Power analysis

We ran a power analysis using the data from Norman-Haignere et al. (2015), computing a measure of the strength of the music component in those 10 participants and then comparing them to a second population of 10 participants created by sampling participants with replacement and then shifting their component weights by various amounts, representing various models for how the music component weights might change in musicians. We quantified the strength of the music component by selecting the 10% of voxels with the largest music component weights, separately for each participant, and then using left-out data to measure the median music component weight in those voxels. Participants' median weights were averaged separately for each group, and the difference between these group averages was computed. The significance of this group difference was assessed by permuting participant groupings 1,000 times. For each shift amount, we repeated this entire resampling procedure 1,000 times, and found that we were able to detect a significant group difference 50% of the time if the two groups' median weights differed by 32%.

Figure 2

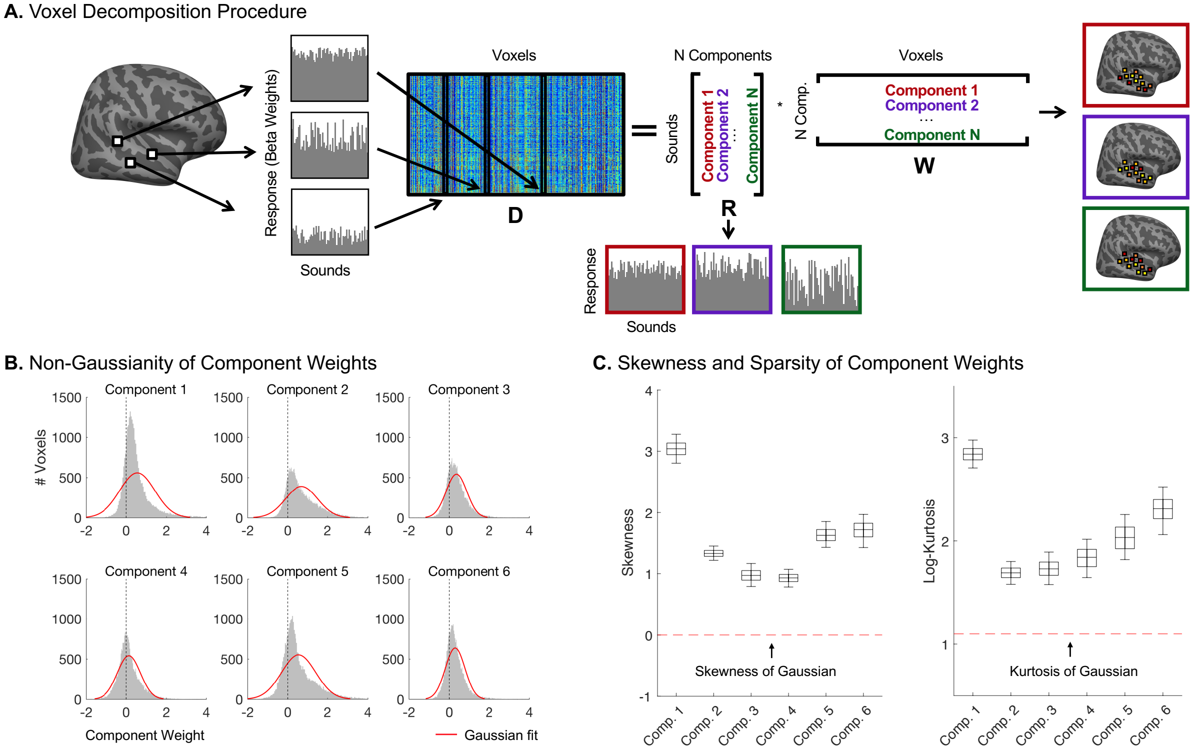

Figure 2. (**A**) Diagram depicting the voxel decomposition method, reproduced from Norman-Haignere et al. (2015). The average response of each voxel to the 192 sounds is represented as a vector, and the response vector for every voxel from all 20 subjects is concatenated into a matrix (192 sounds x 26,792 voxels). This matrix is then factorized into a response profile matrix (192 sounds x N components) and a voxel weight matrix (N components x 26,792 voxels). (**B**) Like ICA, this method searches amongst the many possible solutions to the factorization problem for components that have a maximally non-Gaussian distribution of weights across voxels. Histograms showing the weight distributions for each component can be seen here, along with their Gaussian fits (red). (**C**) Skewness and log-kurtosis (a measure of sparsity) for each component, illustrating that the inferred components are skewed and sparse compared to a Gaussian (red dotted lines). Box-and-whisker plots show central 50% (boxes) and central 95% (whiskers) of the distribution for each statistic (via bootstrapping across subjects). For both the weight distribution histograms and analyses of non-Gaussianity, we used independent data to infer components (runs 1-24) and to measure the statistical properties of the component weights (runs 25-48).

Data analyses

For all weight comparisons of music component weights between musicians and non-musicians, we used individual participants' component weights, which were calculated using their individual data (i.e. voxel responses to the 192 natural sound stimuli) and the pseudoinverse of the response profile matrix that was defined using the data from all 20 participants. To ensure that all participants had component weights defined for the same set of voxels, we selected only voxels that passed our selectivity criteria (sound-responsive, with a reliable response across scanning sessions) at the group level after averaging across individual participants' voxel response data. A total of 2,249 voxels met these criteria, and these are the voxels that are plotted in all figures (**Figures 4, 7, 9**).

The magnitudes of these individual participant component weights were used to determine (1) the amount of variance that was accounted for by the music component in each group, and (2) the strength of the music component. There are many ways to quantify the strength of the music component, but we reasoned that it seemed most sensible to evaluate the strength of the music component in the most music-selective neural populations. After all, music-selectivity is typically limited to a small fraction of voxels, and thus changes in music selectivity might only slightly change the overall median and variance across all voxels. Following this logic, the measure of component strength that we report throughout the paper is individual participants' median weight over the voxels with the top 10% of component weights, rather than over all voxels. To avoid statistical circularity from the selection process, we used cross-validation. Specifically, for a given participant, we inferred components using data from all other participants. Then, one half of the left-out participant's data (corresponding to runs 1-24, or the first three repetitions of each stimulus) was used to infer weights, and the voxels with the top 10% of weights were selected. Finally, we inferred the left-out participant's component weights in those selected voxels using the other half of their data, and calculated the median of those weights. This was done separately for each participant and each component. The significance of the group difference was assessed using a nonparametric test permuting participant groupings 10,000 times and recalculating the group difference to build up a null distribution against which to measure the observed group difference. We also repeated all analyses using individual participants' median component weight over all 2,249 selected voxels (not using cross-validation), and the results remain the same.

To visualize the anatomical distribution of component weights, individual participants' component weights were projected onto the cortical surface of the standard Freesurfer FsAverage template, and a random effects analysis was conducted separately for musicians and non-musicians. To correct for multiple comparisons, we adjusted the false discovery rate (FDR, c(V) = 1, q = 0.05) using the method from Genovese, Lazar, and Nichols (2002).

To get a measure of selectivity for the music components inferred from musicians and non-musicians separately, we used d-prime to calculate the distance between the distribution of component responses to music stimuli ("Western instrumental," "Non-Western instrumental," "Western vocal," "Non-Western vocal," and "drums") and the distribution of component responses to non-music stimuli. The significance of the observed group difference was evaluated using a nonparametric test permuting participant grouping, re-deriving new sets of components for each permuted participant grouping, and building up a null distribution of d-prime differences.

RESULTS

Our primary question was whether the organization of auditory cortex, and in particular its selectivity for music, is present in people with almost no musical training, and whether it differs in expert musicians compared to non-musicians. To that end, we scanned ten people with extensive musical training and ten with almost none, and used voxel decomposition (Norman-Haignere et al., 2015) to test whether the magnitude, anatomical location, or selectivity of music-selective neural populations is influenced by musical training.

Figure 3

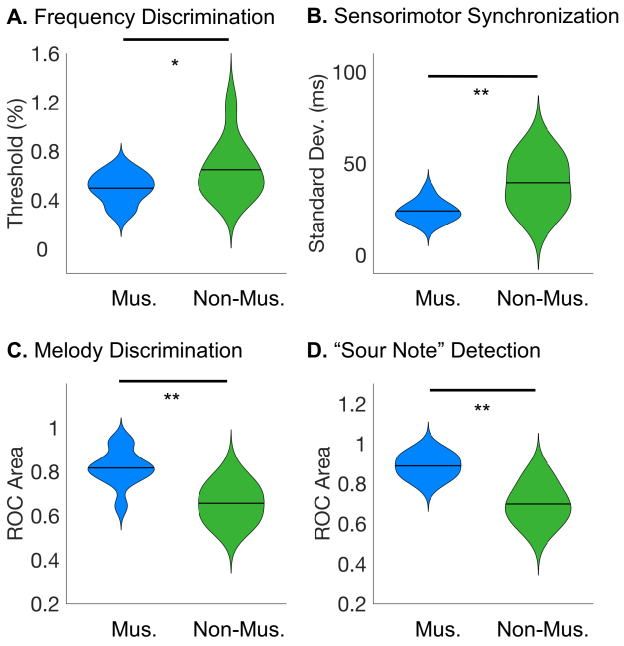

Figure 3. Musicians outperform non-musicians on psychoacoustic tasks. (**A**) Participants' pure tone frequency discrimination thresholds were measured using a 1-up 3-down adaptive two-alternative forced choice (2AFC) task, in which participants indicated which of two pairs of tones were different in frequency. Note that lower thresholds correspond to better performance. (**B**) Sensorimotor synchronization abilities were measured by instructing participants to tap along with an isochronous beat at various tempos, and comparing the standard deviation of the difference between participants' response onsets and the actual stimulus onsets. (**C**) Melody discrimination was measured using a 2AFC task, in which participants heard two five-note melodies (with the second one transposed up by a tritone) and were asked to judge whether the two melodies were the same or different. (**D**) We measured participants' ability to determine whether a melody conforms to the rules of Western music theory by creating 16-note melodies using a probabilistic generative model of Western tonal melodies, and instructing participants to determine whether or not the melody contained an out-of-key ("sour") note. Mus. = musicians, Non-Mus. = non-musicians. * = significant at $p < 0.05$, one-tailed; ** = significant at $p < 0.005$, one-tailed.

Musicians outperform non-musicians on psychoacoustical tasks

We first examined participants' psychoacoustic abilities to determine whether our musicians showed the commonly observed perceptual markers of highly trained musicians. As expected, musicians outperformed non-musicians on all behavioral psychoacoustic tasks, replicating prior findings (**Figure 3**). Consistent with previous reports (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001; Micheyl et al., 2006), musicians performed better on the frequency discrimination task (mean discrimination threshold = 0.50%) than non-musicians (mean discrimination threshold = 0.65%, t = 1.82, p = 0.04, one-tailed t-test, **Figure 3A**). Musicians were also better able to synchronize their finger tapping with an isochronous beat, showing significantly less variability than non-musicians in their response (sd = 24.0ms) than non-musicians (sd = 39.4ms) (t(18) = -3.48, p = 0.0013, one-tailed t-test, **Figure 3B**). When presented with musical melodies, musicians were better able to discriminate between two similar melodies (musician mean ROC area = 0.82 correct, non-musician mean ROC area = 0.66, t(18)= 4.22, p = 0.0003, one-tailed t-test, **Figure 3C**), and to detect scale violations within melodies (musician mean ROC area = 0.89, non-musicians mean ROC area = 0.70, t(18) = 5.27, p = 2.60 x 10-5, one-tailed t-test, **Figure 3D**). These behavioral effects validate our participants' self-reported status as trained musicians or non-musicians.

Replication of music-selective component using voxel decomposition method

Our first question was whether we would replicate the component structure of auditory cortex reported by Norman-Haignere et al. (2015), especially the music-selective component. We measured the response of voxels within auditory cortex to 192 natural sounds (responses were averaged across time because the sounds were short relative to the hemodynamic response). We then modeled the response of these voxels as the weighted sum of a small set of components, using the voxel decomposition method depicted in **Figure 2A**. This method factorizes the voxel responses (**D**) into two matrices: one containing the components' response profiles across the sound set (**R**), and the second containing voxel weights specifying the extent to which each component contributes to the response of each voxel (**W**). The method searches for components that have maximally non-Gaussian weight distributions, using a variant of independent components analysis (ICA) that seeks to minimize the entropy of the component weights. The logic of this approach is that independent non-Gaussian variables become more Gaussian when they are linearly mixed, and thus non-Gaussianity provides a statistical signature that can be used to detect the unmixed components. Since the only free parameter in this analysis is the number of components recovered, two metrics were used to determine the optimal number of components: (1) the variance of voxel responses the components can explain, and

(2) the accuracy of the components' prediction of left-out data. In the previous study (Norman-Haignere et al., 2015), six components were sufficient to explain over 80% of the reliable variance in voxel responses. Four of these components captured expected acoustic properties of the sound set (e.g. frequency, spectrotemporal modulation). The other two components were highly selective for speech sounds and music sounds, respectively.

In our previous study (Norman-Haignere et al., 2015), prior to applying the voxel decomposition algorithm, each participant's responses were de-meaned across voxels, such that each participant had the same mean response (across voxels) for a given sound. This normalization was included to prevent the voxel decomposition algorithm from discovering additional components that were driven by a single participant (e.g. due to non-replicable sources of noise, such as motion during a scan). However, this analysis step would also remove any group difference in the average response to certain sounds (e.g. music stimuli). To prevent this effect from influencing the comparison of the two groups, we ran the voxel decomposition algorithm without demeaning by individual participants. As expected, this resulted in a larger number of components, with 8 components being needed to optimally model the data from all 20 participants, explaining 89.83% of the voxel response variance, after which the amount of explained variance for each additional component plateaued.

Of the 8 components derived from the non-demeaned data, six of them were each very similar to one of the 6 components from Norman-Haignere et al. (2015) and accounted for 87.54% of voxel response variance. Because the order the components inferred using ICA holds no significance, we first optimally reordered the components to match those of our previous study using the Hungarian algorithm (Kuhn, 1955). After re-ordering, corresponding pairs of components were found to be highly correlated, with r-values ranging from 0.7567 to 0.9847 (**Figure 4A**; see **Figure 4B** for the response profiles of the components, and **Figure 4C** for the profiles averaged within sound categories). The additional two components were much less correlated with any of the six original components, with the strongest correlation being r = 0.2752. As expected, the weights of these additional two components were concentrated in a small number of participants. For this reason, we omitted these two components from further analyses and focused on the set of six components that closely match those discussed previously (**Figure 4B-F**). The non-Gaussianity of these 6 components can be seen in **Figure 2B** (skewness ranging from 1.06 to 2.96, log-kurtosis ranging from 1.70 to 2.79).

As in Norman-Haignere et al. (2015), four of the components were selective for different acoustic properties of sound (**Figure 4D&E**), while two components were selective for speech (component 5) and music respectively (component 6) (**Figure 4B&C)**. The components replicated all of the functional and anatomical properties from our prior study, which we briefly describe here.

Components 1 and 2 exhibited high correlations between their response profiles and measures of stimulus energy in either low (component 1) or high frequency bands (component 2) (**Figure 4D**). Although we did not measure tonotopy in these subjects, the anatomical weights for components 1 and 2 concentrated in what are typically found to be low and high-frequency regions of primary auditory cortex (**Figure 4F**) (Rauschecker et al., 1995; Humphries et al., 2010; Da Costa et al., 2011; Baumann et al., 2013). Components also showed tuning to spectrotemporal modulations (**Figure 4E**), with a tradeoff between selectivity for fine spectral and slow temporal modulation (components 1 and 4) verses coarse spectral and fast temporal modulation (components 2 and 3) (Singh and Theunissen, 2003; Rodríguez et al., 2010). Component 4, which exhibited selectivity for fine spectral modulation was concentrated anterior to Heschl's gyrus (component 4, **Figure 4F**), similar to prior work that has identified tone-selective regions in anterolateral auditory cortex in humans (Patterson et al., 2002; Penagos et al., 2004; Norman-Haignere et al., 2013). Conversely, selectivity for coarse spectral modulation and fast temporal modulation was concentrated in posterior regions of auditory cortex (component 3, **Figure 4F**) (Santoro et al., 2014), consistent with previous studies reporting selectivity for sound onsets in caudal areas of human auditory cortex (Hamilton et al., 2018).

Figure 4

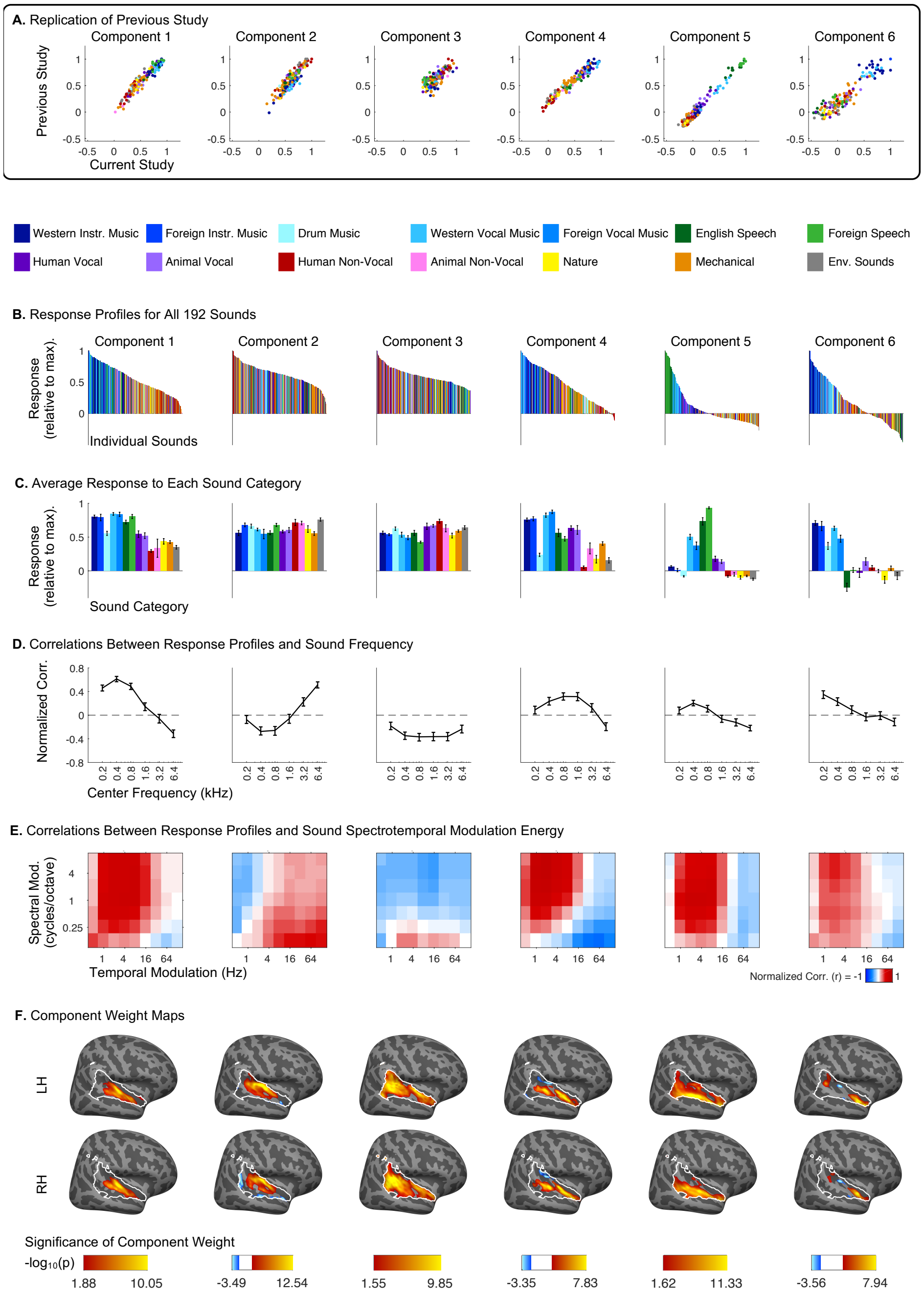

Figure 4. Independent components inferred from voxel decomposition of auditory cortex, replicating the results of Norman-Haignere et al. (2015). (**A**) Scatterplots showing the correspondence between the components from our previous study (y-axis) and those from the current study (x-axis). For the response profiles from the current study, only the 165 sounds that were common between the two studies are plotted. Sounds are colored according to their semantic category, as determined by raters on Amazon Mechanical Turk. (**B**) Response profiles of components inferred from all participants (n = 20), showing the full distribution of all 192 sounds. Sounds are colored according to their semantic category. Note that "Western Vocal Music" stimuli were sung in English. (**C**) The same response profiles as above, but showing the average response to each sound category. (**D**) Correlation of component response profiles with stimulus energy in different frequency bands. (**E**) Correlation of component response profiles with spectrotemporal modulation energy in the cochleograms for each sound. (**F**) Spatial distribution of component voxel weights, computed using a random effects analysis of participants' individual component weights. Each map plots the contrast of component weight > 0; p values are logarithmically transformed ($-\log_{10}[p]$). The white outline indicates the 2,249 voxels that were included in the analysis. The color scale represents voxels that are significant at FDR q = 0.05, with this threshold being computed for each component separately. Voxels that do not survive FDR correction are not colored, and these values appear as white on the color bar. The left hemisphere (top row) is flipped to make it easier to visually compare weight distributions across hemispheres.

The weights for the speech-selective component (component 5) were concentrated in the middle portion of the superior temporal gyrus (midSTG, **Figure 4F**), as expected (e.g., Scott et al., 2000; Hickok and Poeppel, 2007; Overath et al., 2015). In contrast, the weights for the music-selective (component 6) were most prominent anterior to PAC in the planum polare, with a secondary cluster posterior to PAC in the planum temporale (**Figure 4F**) (Ohnishi et al., 2001; Fedorenko et al., 2012; Angulo-Perkins et al., 2014; Armony et al., 2015; Norman-Haignere et al., 2015).

These results closely replicate the functional organization of human auditory cortex reported by Norman-Haignere et al. (2015), including the existence and anatomical location of inferred music-selective neural populations.

Figure 5

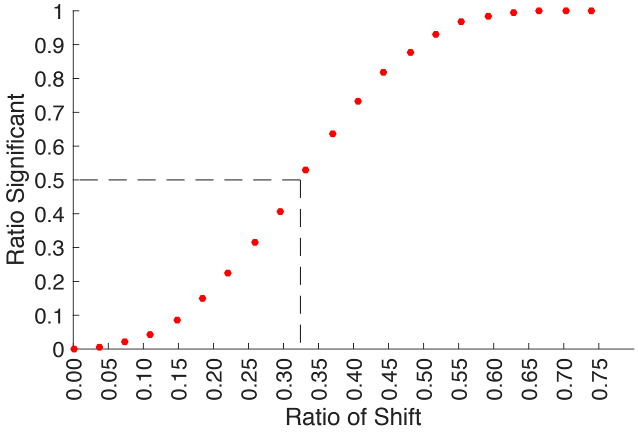

Figure 5. Power analysis results using the data from Norman-Haignere et al. (2015), in which we computed a measure of the strength of the music component in those 10 participants and then compared them to second population of 10 participants created by sampling participants with replacement and shifting their component weights by various amounts, representing various models for how the music component weights might change in musicians. We quantified the strength of the music component by selecting the 10% of voxels with the largest music component weights, separately for each participant, and then using left-out data to measure the median music component weight in those voxels. Participants' median weights were averaged separately for each group, and the difference between these group averages was computed. The significance of this group difference was assessed by permuting participant groupings 1,000 times. For each shift amount (x-axis), we repeated this entire resampling procedure 1,000 times, and plotted the ratio of resamplings in which we detected a significant difference (y-axis). The dotted line represents the shift amount in which we were able to detect a significant difference in the two groups' median weights 50% of the time.

Music component weights differ only slightly between musicians and non-musicians

To determine whether musical training affects stimulus selectivity in auditory cortex, we compared both the magnitude and spatial distribution of the weights of the music-selective component (inferred from the data of all 20 participants) between expert musicians and non-musicians. To get a sense for how big of a group difference we would be able to detect, we conducted a power analysis using the data from Norman-Haignere et al. (2015). We compared the music component weights for his participants (n = 10) with a second population of 10 participants created by sampling participants with replacement and then shifting their component weights by various amounts, representing various models for how the music component weights might change in musicians. The results of this analysis suggested that we could expect to detect a significant group difference the majority of the time if the groups' music component weights differed by just over 30% (**Figure 5**), which is a small difference compared to the variability between individual participants.

To compare the magnitude of the music component in musicians versus non-musicians we used two different measures: (1) the amount of variance that was accounted for by the music component in each group, and (2) the strength of the music component, quantified as participants' median component weight over the top 10% of voxels (see Methods). Group differences were assessed using a nonparametric test permuting participant grouping (10,000 permutations, two-tailed tests, not corrected for the six components tested). We found that the music component did not account for significantly more variance in the voxel responses of musicians (7.56%) than of non-musicians (5.45%) ($p = 0.2066$), and that the median weight over the top 10% of voxels with the highest music component weights did not differ significantly by group (component 6, **Figure 6**, $p = 0.1019$). There was also no significant group difference for the other five components (components 1-5, **Figure 6**, all p's > 0.12). Results were similar if we instead used participants' median weight over all 2,249 voxels as a measure of the strength of the music component, rather than the median over the top 10% of voxels.

Figure 6

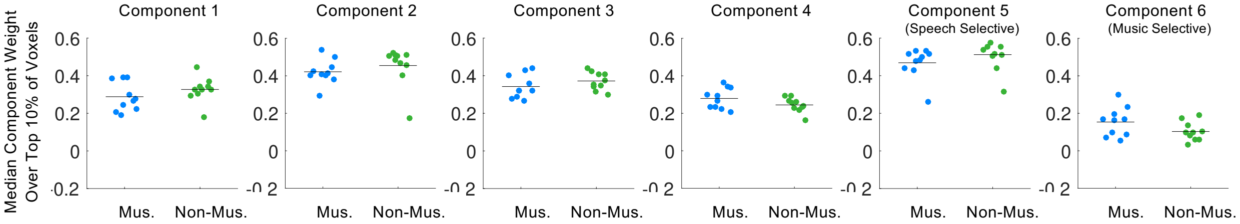

Figure 6. Comparison of musicians' (blue) and non-musicians' (green) component weights. We quantified the strength of the music component in each participant by selecting the 10% of voxels with the largest music component weights, separately for each participant, and then using left-out data to measure the participant's median music component weight in those voxels. Black horizontal line indicates the mean across participants within each group. Mus. = musicians, Non-Mus. = non-musicians.

The analysis described above finds no significant group difference in music component weights when pooled across voxels from all over auditory cortex, but does not rule out the possibility that musicians and non-musicians might show different anatomical distributions of the music component weights. To examine this possibility, we conducted a random effects analysis on individual participants' music component weight matrices that had been projected to the cortical surface. First, we performed a random effects analysis separately for musicians and non-musicians in order to visualize which voxels showed significant music component weights in each group separately (**Figure 7**). The resulting weight maps show a general anatomical correspondence between the weight maps for the two groups. A random effects analysis directly comparing the two groups supported this impression, showing very little difference between the weight magnitudes of musicians and non-musicians, with not a single voxel surviving FDR correction (q = 0.05). Similar comparisons between mismatching components (e.g. non-musicians' music component and musicians' other components) resulted in significant differences over large swaths of auditory cortex (not shown), suggesting that the lack of an observed group difference for the music component weight maps is not due to a lack of power.

Rather than grouping participants as either "musicians" or "non-musicians," another way to examine the effect of musical training is to look at the relationship between various aspects of individual participants' musical training and the strength of their individual component. As part of their participation in the experiment, participants had completed a detailed questionnaire about their musical experience

Figure 7

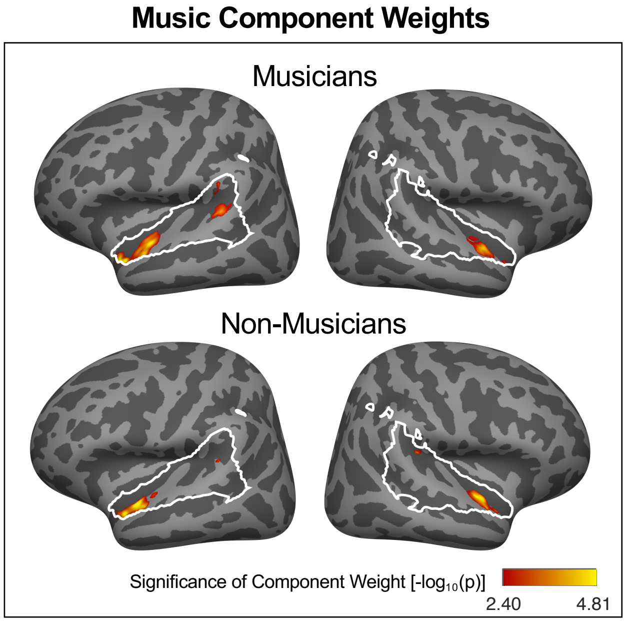

Figure 7: Voxel weights for the music component inferred from the full set of 20 participants, shown separately for musicians and non-musicians. First, individual participant component weights were inferred using the component response profiles and individual participant voxel responses to the 192-sound stimulus set. Then, to visualize group data, a random effects analysis tested whether component weights were significantly greater than zero across participants at each voxel. This analysis was performed separately for the 10 musicians (top) and 10 non-musicians (bottom). Each map plots the contrast of component weight > 0; p values are logarithmically transformed ($-\log_{10}[p]$). The white outline indicates the 2,249 voxels that were included in the analysis. The color scale represents voxels that are significant at FDR q = 0.05.

and training, as well as their music listening behaviors (e.g. average number of hours of music listening per day). We correlated these musical behavior metrics with participants' median component weights over the top 10% of voxels, and while a few questionnaire items were found to correlate with individual components, none of the correlations surpassed the Bonferroni-corrected p-level for six tests ($p = 0.05/6 = 0.0083$). Note that one of the questionnaire measures was participants' self-reported amount of music listening per day, which we did not find to be correlated with participants' music component weights ($r = -0.0412$, $p = 0.8629$). However, because many these questionnaire metrics were highly correlated (e.g. number of years of formal training, number of years of private lessons, number of years of daily practice), we used principal component analysis (PCA) to extract the first PC as a single composite measure of each participants' musical training (accounting for 84.1507% of the total variance). The correlation between this measure of musical training and the median weight of the music component did not reach significance (r = -0.4160, p = 0.0681), and was not close to significance given Bonferroni correction. The same was true for all other components (all p's > 0.09). These results hold when considering only musician participants, for both the raw questionnaire responses (all p's > 0.13) and with the first PC of musicians' questionnaire responses (r = -0.1477, p = 0.6840).

Separate voxel decomposition analyses on musicians and non-musicians reveal similar component structures

Because the components inferred in the previous analysis were defined using the data from all 20 participants, and because the algorithm seeks to infer components that explain responses across all participants, meaningful group differences in the response components themselves, rather than the component weights, might not be detected. We therefore next ran the voxel decomposition analysis separately on the musicians and non-musicians. This analysis allows us to determine whether different canonical response profiles underlie the functional organization of auditory cortex of musicians and non-musicians.

For these separate group analyses, each participant's responses were de-meaned across voxels (as in Norman-Haignere et al., 2015), such that each participant had the same mean response across voxels to each sound. Measures of explained variance indicated that the optimal number of components from the voxel decomposition analysis was six for both musicians and non-musicians, explaining 88.09% and 88.56% of participants' voxel response variance, respectively.

The set of six components inferred from the separate groups were very similar to each other (**Figure 8A**). After optimally reordering the components using the Hungarian algorithm, pairs of corresponding components (between those derived from the musicians, and those derived from the non-musicians) were highly correlated (r values between 0.7735 to 0.9904; r = 0.9018 for the music-selective component) (**Figure 8B**). These results indicate that similar component structures underlie auditory cortical responses in musicians and non-musicians. Additionally, these two sets of components were very similar to those derived from the whole-group analysis of all 20 participants, with the corresponding components from the two sets having correlations ranging from 0.89 to 0.99 for musicians, and from 0.67 to 0.99 for non-musicians. Moreover, the music component was similarly selective in both groups (**Figure 8C&D**), as measured by the difference in means between music and non-music sounds, divided by their standard deviation (**Figure 8D**; d-prime; musicians d' = 1.59, non-musicians' d' = 1.21; not significantly different, p = 0.19; 1,000 permutations). We note that the mean response magnitude of the music-selective component was higher in non-musicians (note different axes in **Figure 8A** & **Figure 8C**), but we have found such additive response profile "offsets" to occur occasionally as an unstable feature of the voxel decomposition algorithm. For example, discarding a scan of data from the analysis caused the mean of the music component to fluctuate, but not the response pattern. The critical result thus appears to be the separability of music and non-music in the component, which is similar for musicians and non-musicians.

Comparing musicians and non-musicians using standard methods

All of the analyses described above are based on the voxel decomposition method from Norman-Haignere et al. (2015). As a complementary analysis, we conducted a standard group random effects analysis, contrasting responses to music vs. non-music sound stimuli. Many brain regions responded significantly more to music sounds than non-music sounds, presumably because these sound sets differ both in low-level acoustic properties and higher-order category-specific properties (Norman-Haignere and McDermott, 2018) (**Figure 9A**). In particular, the music-preferring voxels overlapped both the music-selective component and the "pitch-selective" component (component 4), consistent with the fact that both of these components responded preferentially to music. In agreement with our findings comparing music component weights between groups, the distribution of music-preferring voxels was similar in musicians and non-musicians (**Figure 9B**). A direct comparison between groups (**Figure 9C**) revealed a few voxels that survived FDR correction (q = 0.05) with the

Figure 8

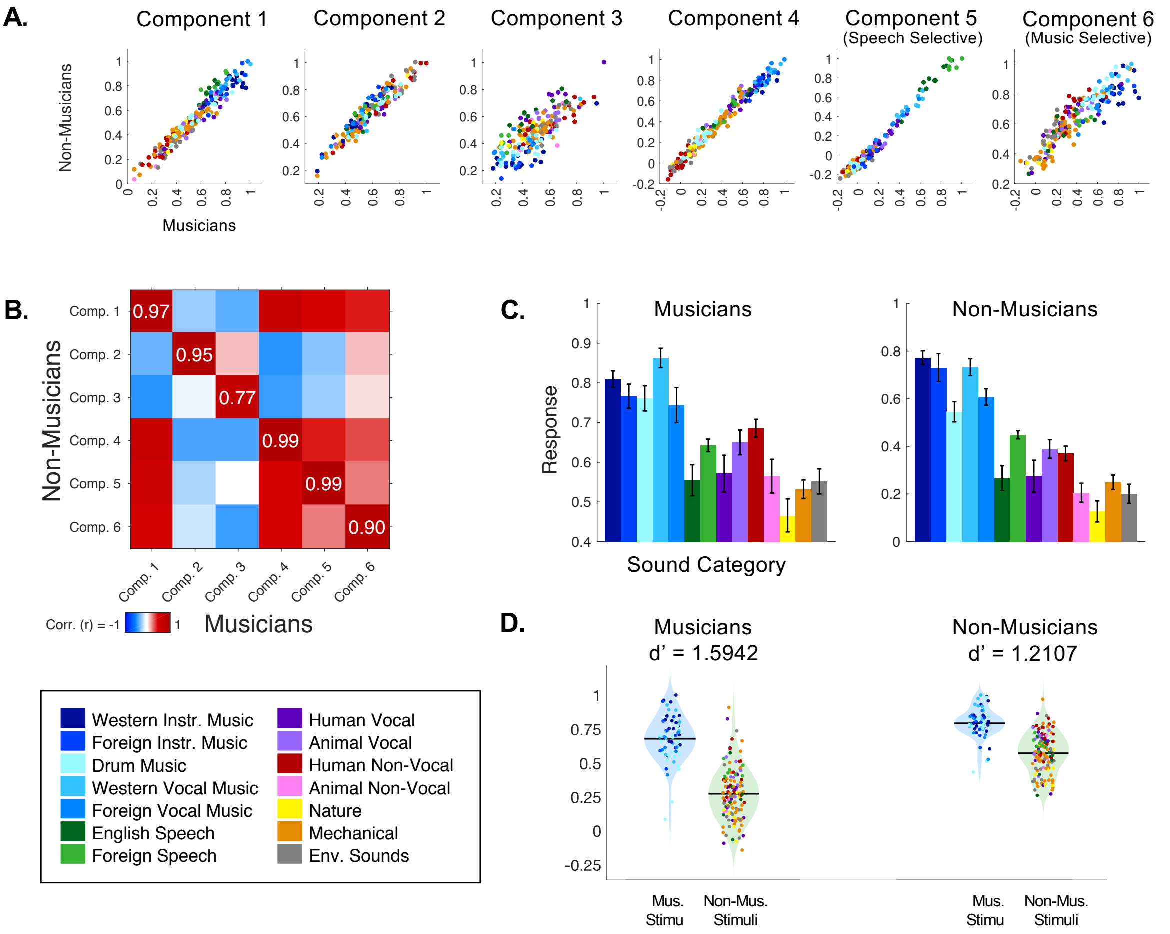

Figure 8. (**A**) Scatterplots showing the correspondence between the components inferred from musicians (x-axis) and non-musicians (y-axis). Sounds are colored according to their semantic category, as determined by raters on Amazon Mechanical Turk. Note that to allow the comparison of component response profiles regardless of any positive offset (which is an unstable artifact of the voxel decomposition algorithm), the axes differ between groups. (**B**) Correlation matrix depicting relationships between component response profiles inferred from musicians (x-axis) and non-musicians (y-axis) separately. The Pearson correlation coefficient is included for values on the diagonal. (**C**) Response profiles of music components inferred from musicians (n = 10, left) and non-musicians (n = 10, right), averaged over sound categories. Note the different axis limits for the two groups. (**D**) Distributions of music stimuli (blue shading) and non-music stimuli (green shading) within the music component response profiles inferred from musicians (n = 10, left) and non-musicians (n = 10, right), with the mean for each stimulus group indicated by the horizontal black line. The d-prime reflecting the distance between music and non-music stimuli for each group is shown above each plot. Sounds are colored according to their semantic category. We note that the positive offset in the response profile inferred from non-musicians (right) had only a small, non-significant effect on the d-prime between music and non-music stimuli.

primary cluster being located in the very anterior tip of left planum polare. A further ROI analysis of this region using independent data indicates that it is not music-selective (even though the difference between music and non-music stimuli is slightly greater in musicians), and in fact seems to respond to speech and song more than any other stimulus category. This finding provides additional support for our conclusion that musicians and non-musicians do not differ substantially in their cortical music selectivity.

Figure 9

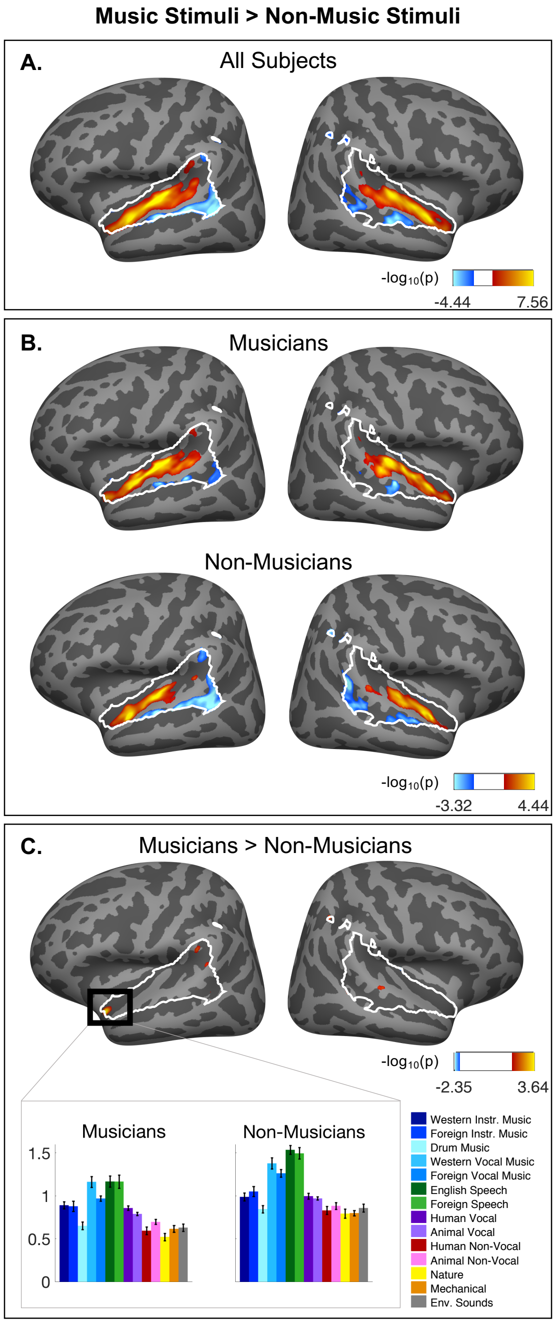

Figure 9. Significance maps of group random effects analysis on raw fMRI responses (not components). (**A**) Main effect of music stimuli > non-music stimuli, computed as a random effects analysis over all participants. (**B**) Same as above, but computed separately for musicians (top) and non-musicians (bottom). (**C**) Group difference, which is equivalent to the stimulus (music vs. non-music) by group (musician vs. non-musician) interaction. Each map plots logarithmically transformed p-values ($-\log_{10}[p]$), signed such that positive values indicate positive weights, and is thresholded at $-\log_{10}[p] > 3$ ($p < 0.001$), uncorrected. The color scale represents voxels that are significant at FDR q = 0.05. The white outline indicates the 2,249 voxels that were included in the analysis.

New insights into music selectivity: Music-selective regions of auditory cortex show high responses to drum stimuli and unfamiliar musical genres

Because our experiment utilized a broader stimulus set that the original study (Norman-Haignere et al., 2015), we were able to use voxel decomposition to gain new insights into the nature of cortical music selectivity in addition to examining the effect of explicit training on these populations. The set of natural sounds used in this study included a total of 62 music stimuli, spanning a variety of instruments, genres, and cultures. Using this diverse set of music stimuli, we can begin to address the questions of (1) whether music selectivity is specific to the music of one's own culture, and (2) whether music selectivity is driven solely by features related to pitch, like the presence of a melody.

To expand beyond the original stimulus set from Norman-Haignere et al. (2015), which contained music exclusively from traditionally Western genres and artists, we selected additional music clips from several non-Western musical cultures that varied in tonality and rhythmic complexity (e.g. Indian raga, Balinese gamelan, Chinese opera, Mongolian throat singing, Jewish klezmer, Ugandan lamellophone music) (**Figure 10A**). These non-Western music stimuli were rated by American participants as being similarly musical ($p = 0.37$) but less familiar ($p < 1.0e\text{-}5$) than typical Western music. Despite this difference in familiarity, the magnitude of non-Western music stimuli within the music component was only slightly smaller than the magnitude of Western music stimuli (**Figure 10B**). A nonparametric test permuting music stimulus labels shows that the observed separation between the responses to Western and non-Western music stimuli was small (d-prime = 0.45), though it approached significance ($p = 0.056$, 10,000 permutations). However, the magnitudes of both Western and non-Western music stimuli were both much higher than non-music stimuli. Indeed, both the separability between Western music stimuli and non-music stimuli (d-prime = 2.68) as well as that between non-Western music stimuli and non-music stimuli (d-prime = 2.01) were large and highly significant ($p < 0.0001$; 10,000 permutations). Taken together, these results suggest that music selectivity does not depend upon detailed familiarity with structure of a particular culture's music.

Which stimulus features drive music selectivity? One of the most obvious distinctions is melody and rhythm. While music typically involves both melody and rhythm, when assembling our music stimuli we made an attempt to pick clips that varied in the prominence and complexity of their melodic and rhythmic content. In particular, we included 13 stimuli consisting of drumming from a variety of genres and cultures, because drum music mostly isolates the rhythmic features of music while minimizing (though not completely eliminating) melodic features. Whether music-selective auditory cortex would respond highly to these drum stimuli was largely unknown, partially because the Norman-Haignere et al. (2015) study only included two drum stimuli, one of which was just a stationary snare drum roll. However, the drum stimuli in our study ranked relatively high in the music component response profile, averaging only slightly below the other instrumental and vocal music category responses (d-prime = 1.08), and considerably higher than the other non-music stimulus categories (d-prime = 1.90) (**Figure 10B**). This finding suggests that the music component is not simply tuned to melodic information, but is also responsive to rhythm.

DISCUSSION

In this study, we tested whether cortical music selectivity depends upon explicit musical training. Our results show a clear music component in people with almost no musical training. Indeed, all of the key response patterns that characterize the functional organization of human auditory cortex are robustly present in both musicians and non-musicians, suggesting that explicit training does not substantially alter the functional organization of auditory cortex. The small group difference in the music component weights did not reach statistical significance. These results demonstrate that passive exposure to music is sufficient for the existence of music selectivity in auditory cortex, which is not dependent on or strongly modified by extensive explicit musical training.

Figure 10

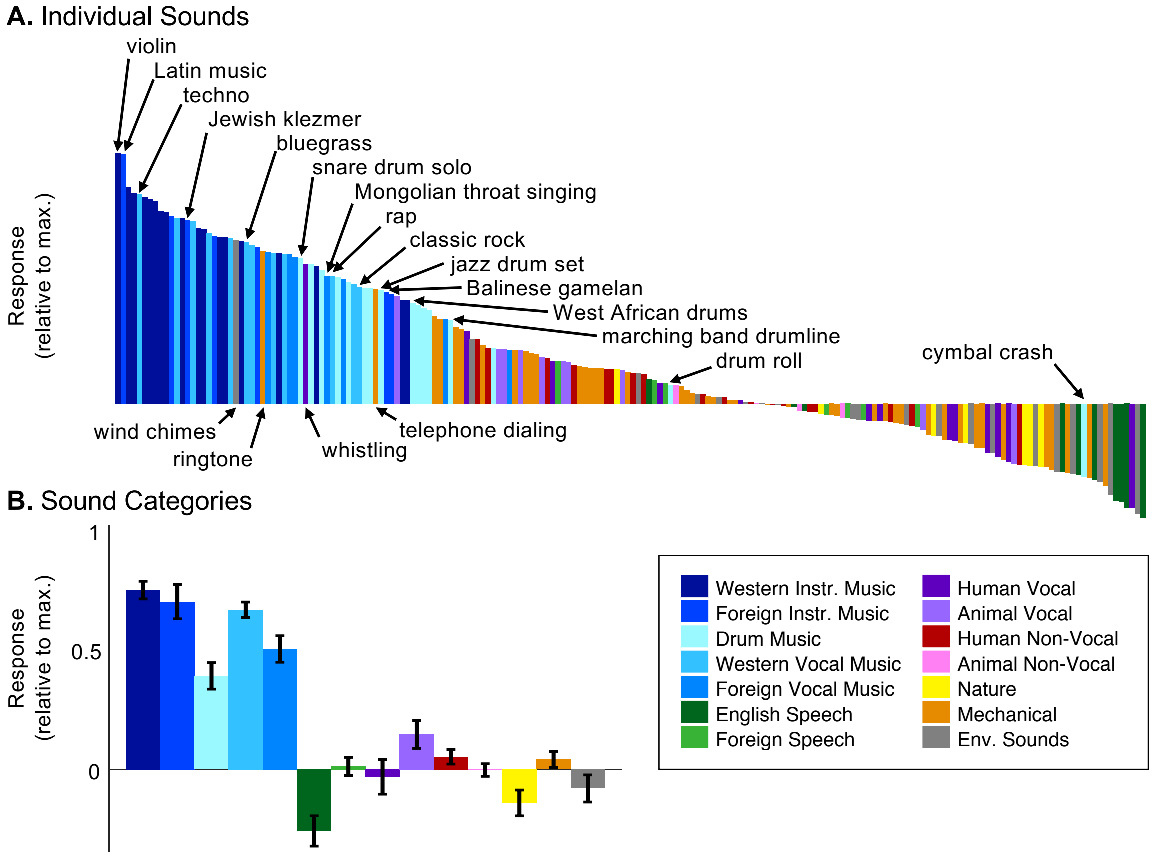

Figure 10. (**A**) Close-up of the response profile for the music component inferred from all participants (n = 20), with example stimuli labeled. Note that there are a few "non-music" stimuli (categorized as such by Amazon Mechanical Turk raters) with high component rankings, but that these are all somewhat melodic (e.g. wind chimes, ringtone). Similarly, "music" stimuli with low component rankings (e.g. "drumroll" and "cymbal crash") do not contain salient melody or rhythm, despite being classified as "music" by human listeners. (**B**) Response profiles of components inferred from all participants (n = 20), averaged over sound categories, reproduced from Figure 4C.

Despite the lack of an effect of explicit training observed here, our study leaves open the possibility that music selectivity might instead reflect implicit knowledge of musical structure (Bigand, 1983; Bigand and Pineau, 1997; Koelsch et al., 2000; Tillmann et al., 2000; Tillmann, 2005; Bigand and Poulin-Charronnat, 2006) gleaned from a lifetime of passive exposure to music. We did not find a significant correlation between participants' self-reported amount of music listening per day, but it may be that everyone in our sample had enough (or nearly enough) passive exposure to music to develop neural selectivity to music, and that additional exposure does not strengthen that selectivity. Indeed, behavioral studies of non-industrialized societies who lack electricity and much contact with western culture show pronounced differences in many aspects of music perception (McDermott et al., 2016;

Jacoby and McDermott, 2017; Jacoby et al., 2019). It is therefore possible that the brains of people from these cultures with less extensive musical exposure would not display such pronounced music selectivity. Thus, our data do not show that music selectivity in the brain is independent of experience. Rather, these results show that normal exposure to music (by the standards of modern Western culture) is sufficient for music selectivity, and that this selectivity is not greatly modified by extensive and explicit training.

We note that our musician participants had substantial explicit musical training, and our non-musicians virtually none, thus maximizing our chance of detecting a difference if one was present. And as expected, our musician participants showed better behavioral performance across several psychoacoustic tests that are associated with expert musicians. Of course, it remains possible that a meaningful difference between music selectivity in musicians and non-musicians exists, but was too small to be detected in our study. Our experiment was well powered to detect moderate differences between groups (~30% increases in weights), but scanning a much larger number of participants could enable detection of a smaller group difference in music selectivity. We note that our study required collecting 60 scan sessions across 20 participants, each 2 hours in duration, so collecting substantially more data would be a non-trivial enterprise. Regardless, our results indicate that any stable group difference, if present, is small.

The fact that we observed no clear group differences in music-selective neural responses within auditory cortex raises the question of what constitutes the neural basis of music expertise. Perhaps musical expertise alters neural responses at finer spatial or temporal scales than can be resolved with fMRI. Alternatively, musical training could modify neural responses exclusively outside of auditory cortex, such as within frontal or parietal regions involved in decision making or attention (Strait and Kraus, 2011; Harris and De Jong, 2015; Alluri et al., 2017; Puschmann et al., 2018), or in motor or limbic regions (Janata and Grafton, 2003; Baumann et al., 2007; Chen et al., 2008; Grahn and Rowe, 2009; Luo et al., 2012; Alluri et al., 2015; Saari et al., 2018; de Aquino et al., 2019). There is also evidence that musical training can alter the anatomical structure of auditory (Schneider et al., 2002, 2005; Bermudez et al., 2009; Foster and Zatorre, 2010), motor (Elbert et al., 1995; Gaser and Schlaug, 2003; Bermudez and Zatorre, 2005; Bermudez et al., 2009; Bashwiner et al., 2016), or limbic regions like the hippocampus (Teki et al., 2012). Another possibility is that musical expertise may influence the strength of anatomical (Imfeld et al., 2009; Halwani et al., 2011; de Manzano and Ullén, 2017) or functional (Chen et al., 2008; Grahn and Rowe, 2009; Luo et al., 2012; Alluri et al., 2015; Palomar-García et al., 2016) connections between auditory and other cortical or subcortical regions. Finally, it could still be the case that the commencement of musical training as a child is associated with auditory cortical plasticity, but that these changes are transient and fade over time. Some precedent for this possibility is found in animal research: auditory cortical map expansion is associated with the degree of perceptual learning, but the cortical map can revert back to its default organization without a corresponding decrement in behavioral performance (e.g. Reed et al., 2011).

Another possibility is that differences between musicians and non-musicians would emerge during more challenging musical tasks. While the participants in this study were instructed to listen carefully to the sound clips while they were in the scanner, they were performing a very simple task intentionally designed to be easy for all participants (detecting a change in sound intensity) in order to gauge whether participants were alert and paying attention. Perhaps differences in music-selective responses would emerge if participants were given a more attentionally-demanding task or asked to make judgements about certain aspects of musical stimuli. Indeed, neurophysiological experiments in animals have found that top-down task-dependent influences can powerfully modulate the response properties of auditory cortical neurons (e.g. Polley et al., 2006), and that actively engaging in a task increases the sharpness of tuning compared to passive listening (Lee and Middlebrooks, 2011).

Many open questions remain about cortical music selectivity. A more thorough understanding of what the observed tuning for music is actually selective for could help explain why we did not observe a difference between musicians and non-musicians in this study. For example, musicians and non-musicians could differ in their responses to certain features of music that are not captured by the music selective component that is isolated using our methodology, such as metrical structure that unfolds over time. Further experiments using longer musical clips and/or neuroimaging methods with better temporal resolution, such as EEG and/or MEG, could help address this question.

The lack of a significant difference in cortical music selectivity between non-musicians and highly trained musicians suggests that music selectivity in the auditory cortex does not rely on the formal knowledge of musical structure that is acquired through years of explicit musical training, and instead may reflect the implicit musical knowledge that listeners gain through casual exposure to music. It is also possible that cortical music selectivity does not reflect experience at all, and instead could be present from birth. These hypotheses could be further tested by scanning a wider range of people from different cultures. We could also potentially learn more about how music selectivity arises in development by scanning infants and children, or by testing populations of people whose lifetime perceptual experience with music is limited in some way (e.g., people with musical anhedonia, children of deaf adults). Because the voxel decomposition technique (Norman-Haignere et al., 2015) enables us to isolate music-selective neural populations using fMRI, it provides a new avenue for exploring the origins of the quintessentially human ability for music.

ACKNOWLEDGEMENTS

This work was supported by NSF grant BCS-1634050 to J.M. and NIH grant DP1HD091947 to N.K.

REFERENCES

Ahissar E, Abeles M, Ahissar M, Haidarliu S, Vaadia E (1998) Hebbian-like functional plasticity in the auditory cortex of the behaving monkey. Neuropharmacology 37:633–655.

Ahissar E, Vaadia E, Ahissar M, Bergman H, Arieli A, Abeles M (1992) Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. Science 257:1412–1415.

Alluri V, Brattico E, Toiviainen P, Burunat I, Bogert B, Numminen J, Kliuchko M (2015) Musical expertise modulates functional connectivity of limbic regions during continuous music listening. Psychomusicology Music Mind, Brain 25:443–454.

Alluri V, Toiviainen P, Burunat I, Kliuchko M, Vuust P, Brattico E (2017) Connectivity patterns during music listening: Evidence for action-based processing in musicians. Hum Brain Mapp 38:2955–2970.

Angulo-Perkins A, Aubé W, Peretz I, Barrios FA, Armony JL, Concha L (2014) Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. Cortex 59:126–137.

Armony JL, Aubé W, Angulo-Perkins A, Peretz I, Concha L (2015) The specificity of neural responses to music and their relation to voice processing: An fMRI-adaptation study. Neurosci Lett 593:35–39.

Bailey JA, Penhune VB (2010) Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. Exp Brain Res 204:91–101.

Bakin JS, Weinberger NM (1996) Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. Proc Natl Acad Sci U S A 93:11219–11224.

Bao S, Chan VT, Merzenich MM (2001) Cortical remodelling induced by activity of ventral tegmental dopamine neurons. Nature 412:92–95.

Bashwiner DM, Wertz CJ, Flores RA, Jung RE (2016) Musical Creativity "Revealed" in Brain Structure: Interplay between Motor, Default Mode, and Limbic Networks. Sci Rep 6:1–8.

Baumann S, Koeneke S, Schmidt CF, Meyer M, Lutz K, Jancke L (2007) A network for audio-motor coordination in skilled pianists and non-musicians. Brain Res 1161:65–78.

Baumann S, Petkov CI, Griffiths TD (2013) A unified framework for the organisation of the primate auditory cortex. Front Syst Neurosci 7:1–19.

Bermudez P, Lerch JP, Evans AC, Zatorre RJ (2009) Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. Cereb Cortex 19:1583–1596.

Bermudez P, Zatorre RJ (2005) Differences in gray matter between musicians and nonmusicians. Ann N Y Acad Sci 1060:395–399.

Besson M, Schön D, Moreno S, Santos A, Magne C (2007) Influence of musical expertise and musical training on pitch processing in music and language. Restor Neurol Neurosci 25:399–410.

Bieszczad KM, Weinberger NM (2010) Learning strategy trumps motivational level in determining learning-induced auditory cortical plasticity. Neurobiol Learn Mem 93:229–239.

Bieszczad KM, Weinberger NM (2012) Extinction reveals that primary sensory cortex predicts reinforcement outcome. Eur J Neurosci 35:598–613.

Bigand E (1983) Perceiving Musical Stability: The Effect of Tonal Structure, Rhythm, and Musical Expertise. J Exp Psychol Hum Percept Perform 23:808–822.

Bigand E, Pineau M (1997) Global context effects on musical expectancy. Percept Psychophys 59:1098–1107.

Bigand E, Poulin-Charronnat B (2006) Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. Cognition 100:100–130.

Blake DT, Heiser MA, Caywood M, Merzenich MM (2006) Experience-Dependent Adult Cortical Plasticity Requires Cognitive Association between Sensation and Reward. Neuron 52:371–381.

Blood AJ, Zatorre RJ (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci U S A 98:11818–11823.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Chen JL, Penhune VB, Zatorre RJ (2008) Moving on Time: Brain Network for Auditory–Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. J Cogn Neurosci 20:226–239.

Da Costa S, Van Der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. J Neurosci 31:14067–14075.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

David S V., Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc Natl Acad Sci U S A 109:2144–2149.

de Aquino MPB, Verdejo-Román J, Pérez-García M, Pérez-García P (2019) Different role of the supplementary motor area and the insula between musicians and non-musicians in a controlled musical creativity task. Sci Rep 9:1–13.

de Manzano Ö, Ullén F (2017) Same Genes, Different Brains: Neuroanatomical Differences Between Monozygotic Twins Discordant for Musical Training. Cereb Cortex:1–8.

Dick F, Ling Lee H, Nusbaum H, Price CJ (2011) Auditory-Motor Expertise Alters "'Speech Selectivity'" in Professional Musicians and Actors. Cereb Cortex 21:938–948.

Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. Proc Natl Acad Sci 112:E6233-42.

Elbert T, Pantev C, Wienbruch C, Rockstroh B, Taub E (1995) Increased cortical representation of the fingers of the left hand in string players. Science 270:305–307.

Ellis RJ, Bruijn B, Norton AC, Winner E, Schlaug G (2013) Training-mediated leftward asymmetries during music processing: A cross-sectional and longitudinal fMRI analysis. Neuroimage 75:97–

107.

Ellis RJ, Norton AC, Overy K, Winner E, Alsop DC, Schlaug G (2012) Differentiating maturational and training influences on fMRI activation during music processing. Neuroimage 60:1902–1912.

Fedorenko E, McDermott JH, Norman-Haignere S V., Kanwisher NG (2012) Sensitivity to musical structure in the human brain. J Neurophysiol 108:3289–3300.

Foster NE V, Zatorre RJ (2010) Cortical structure predicts success in performing musical transformation judgments. Neuroimage 53:26–36.

Fritz J, Elhilali M, Shamma S (2005) Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hear Res 206:159–176.

Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci 6:1216–1223.

Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C (2004) Musical training enhances automatic encoding of melodic contour and interval structure. J Cogn Neurosci 16:1010–1021.

Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C (2005) Automatic encoding of polyphonic melodies in musicians and nonmusicians. J Cogn Neurosci 17:1578–1592.

Gaser C, Schlaug G (2003) Grey Matter Differences between Musicians and Nonmusicians. Ann N Y Acad Sci 999:514–517.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870–878.

Grahn JA, Rowe JB (2009) Feeling the Beat: Premotor and Striatal Interactions in Musicians and Nonmusicians during Beat Perception. J Neurosci 29:7540–7548.

Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48:63–72.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) “Sparse” temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223.

Halwani GF, Loui P, Rüber T, Schlaug G (2011) Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. Front Psychol 2:156.

Hamilton LS, Edwards E, Chang EF (2018) A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. Curr Biol 28:1860–1871.

Harris R, De Jong BM (2015) Differential parietal and temporal contributions to music perception in improvising and score-dependent musicians, an fMRI study. Brain Res 1624:253–264.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. Neuroimage 50:1202–1211.

Imfeld A, Oechslin MS, Meyer M, Loenneker T, Jancke L (2009) White matter plasticity in the corticospinal tract of musicians : A diffusion tensor imaging study. Neuroimage 46:600–607.

Jacoby N, McDermott JH (2017) Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. Curr Biol 27:359–370.

Jacoby N, Undurraga EA, McPherson MJ, Valdés J, Ossandón T, McDermott JH (2019) Universal and Non-universal Features of Musical Pitch Perception Revealed by Singing. Curr Biol 29:3229–3243.

Janata P, Grafton ST (2003) Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. Nat Neurosci 6:682–687.

Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5:143–156.

Kilgard MP, Pandya PK, Vazquez J, Gehi A, Schreiner CE, Merzenich MM (2001) Sensory input directs spatial and temporal plasticity in primary auditory cortex. J Neurophysiol 86:326–338.

Kishon-Rabin L, Amir O, Vexler Y, Zaltz Y (2001) Pitch discrimination: Are professional musicians better than non-musicians? J Basic Clin Physiol Pharmacol 12:125–144.

Koelsch S, Gunter T, Friederici AD, Èger ES (2000) Brain Indices of Music Processing:

"Nonmusicians'" are Musical." J Cogn Neurosci 12:520–541.

Kuhn HW (1955) The Hungarian Method for the Assignment Problem. Nav Res Logist Q 2:83–97.

LaCroix AN, Diaz AF, Rogalsky C (2015) The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. Front Psychol 6:1–19.

Lappe C, Lappe M, Pantev C (2016) Differential processing of melodic, rhythmic and simple tone deviations in musicians -an MEG study. Neuroimage 124:898–905.

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30:7604–7612.

Lee CC, Middlebrooks JC (2011) Auditory cortex spatial sensitivity sharpens during task performance. Nat Neurosci 14:108–116.

Lee H, Noppeney U (2011) Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc Natl Acad Sci 108.

Luo C, Guo Z wei, Lai Y xiu, Liao W, Liu Q, Kendrick KM, Yao D zhong, Li H (2012) Musical training induces functional plasticity in perceptual and motor networks: Insights from resting-state fMRI. PLoS One 7:1–10.

Margulis EH, Mlsna LM, Uppunda AK, Parrish TB, Wong PCM (2009) Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Hum Brain Mapp 30:267–275.

McDermott JH, Keebler M V., Micheyl C, Oxenham AJ (2010) Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. J Acoust Soc Am 128:1943–1951.

McDermott JH, Schultz AF, Undurraga EA, Godoy RA (2016) Indifference to dissonance in native Amazonians reveals cultural variation in music perception. Nature 535:547–550.

McPherson MJ, McDermott JH (2018) Diversity in pitch perception revealed by task dependence. Nat Hum Behav 2:52–66.

Mehr SA, Singh M, Knox D, Ketter DM, Pickens-jones D, Atwood S, Lucas C, Egner A, Jacoby N, Hopkins EJ, Howard M, Donnell TJO, Pinker S, Krasnow MM, Glowacki L (2019) Universality and diversity in human song. Science 366.

Micheyl C, Delhommeau K, Perrot X, Oxenham AJ (2006) Influence of musical and psychoacoustical training on pitch discrimination. Hear Res 219:36–47.

Moddemeijer R (1989) On estimation of entropy and mutual information of continuous distributions. Signal Processing 16:233–248.

Norman-Haignere S, Feather J, Brunner P, Ritaccio A, McDermott JH, Schalk G, Kanwisher N (2019) Intracranial recordings from human auditory cortex reveal a neural population selective for musical song. bioRxiv.

Norman-Haignere S V., Kanwisher NG, McDermott JH (2013) Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. J Neurosci 33:19451–19469.

Norman-Haignere S V., Kanwisher NG, McDermott JH (2015) Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. Neuron 88:1281–1296.

Norman-Haignere S V, McDermott JH (2018) Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. PLoS Biol 16.

Ohl FW, Scheich H (2005) Learning-induced plasticity in animal and human auditory cortex. Curr Opin Neurobiol 15:470–477.

Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M, Katoh A, Imabayashi E (2001) Functional Anatomy of Musical Perception in Musicians. Cereb Cortex 11:754–760.

Overath T, Mcdermott JH, Zarate JM, Poeppel D (2015) The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. Nat Neurosci 18:903–911.

Palomar-García M-Á, Zatorre RJ, Ventura-Campos N, Bueichekú E, Ávila C (2016) Modulation of Functional Connectivity in Auditory–Motor Networks in Musicians Compared with Nonmusicians.

Cereb Cortex:1–11.

Pantev C, Larry Roberts CE, Schulz M, Engelien A, Ross B (2001) Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12:959–4965.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci 24:6810–6815.

Penhune VB (2011) Sensitive periods in human development: evidence from musical training. Cortex 47:1126–1137.

Polak R, Jacoby N, Fischinger T, Goldberg D, Holzapfel A, London J (2018) Rhythmic prototypes across cultures: A comparative study of tapping synchronization. Music Percept 36:1–23.

Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual Learning Directs Auditory Cortical Map Reorganization through Top-Down Influences. J Neurosci 26:4970–4982.

Puschmann S, Baillet S, Zatorre RJ (2018) Musicians at the Cocktail Party: Neural Substrates of Musical Training During Selective Listening in Multispeaker Situations. Cereb Cortex:1–13.

Rauschecker JP, Tian B, Hauser M (1995) Processing of Complex Sounds in the Macaque Nonprimary Auditory Cortex. Science 268:111–114.

Recanzone GH, Schreiner CE, Merzenich MM (1993) Plasticity in the Frequency Representation of Primary Auditory Cortex following Discrimination Training in Adult Owl Monkeys. J Neurosci 13:87–103.

Reed A, Riley J, Carraway R, Carrasco A, Perez C, Jakkamsetti V, Kilgard MP (2011) Cortical Map Plasticity Improves Learning but Is Not Necessary for Improved Performance. Neuron 70:121–131.

Repp BH (2005) Sensorimotor synchronization: A review of the tapping literature. Psychon Bull Rev 12:969–992.

Repp BH (2010) Sensorimotor synchronization and perception of timing: Effects of music training and task experience. Hum Mov Sci 29:200–213.

Rodríguez FA, Read HL, Escabí MA (2010) Spectral and Temporal Modulation Tradeoff in the Inferior Colliculus. J Neurophysiol 103:887–903.

Rogalsky C, Rong F, Saberi K, Hickok G (2011) Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging. J Neurosci 31:3843–3852.

Rutkowski RG, Weinberger NM (2005) Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. Proc Natl Acad Sci U S A 102:13664–13669.

Saari P, Burunat I, Brattico E, Toiviainen P (2018) Decoding Musical Training from Dynamic Processing of Musical Features in the Brain. Sci Rep 8.

Salimpoor VN, Benovoy M, Larcher K, Dagher A, Zatorre RJ (2011) Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nat Neurosci 14:257–264.

Salimpoor VN, Van Den Bosch I, Kovacevic N, McIntosh AR, Dagher A, Zatorre RJ (2013) Interactions between the nucleus accumbens and auditory cortices predict music reward value. Science 340:216–219.

Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (2014) Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. PLoS Comput Biol 10.

Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A (2002) Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nat Neurosci 5:688–694.

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A (2005) Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat Neurosci 8:1241–1247.

Scott SK, Blank CC, Rosen S, Wise RJS (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 23:5545–5552.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. Cit J Acoust Soc Am 114:3394.

Spiegel MF, Watson CS (1984) Performance on frequency-discrimination tasks by musicians and nonmusicians. J Acoust Soc Am 76:1690–1695.

Strait DL, Kraus N (2011) Can You Hear Me Now? Musical Training Shapes Functional Brain Networks for Selective Auditory Attention and Hearing Speech in Noise. Front Psychol 2:10.

Teki S, Kumar S, von Kriegstein K, Stewart L, Rebecca Lyness C, Moore BCJ, Capleton B, Griffiths TD (2012) Navigating the auditory scene: An expert role for the hippocampus. J Neurosci 32:12251–12257.

Temperley D (2008) A probabilistic model of melody perception. Cogn Sci 32:418–444.

Tierney A, Dick F, Deutsch D, Sereno M (2013) Speech versus Song: Multiple Pitch-Sensitive Areas Revealed by a Naturally Occurring Musical Illusion. Cereb Cortex 23:249–254.

Tillmann B (2005) Implicit investigations of tonal knowledge in nonmusician listeners. Ann N Y Acad Sci 1060:100–110.

Tillmann B, Bigand E, Bharucha JJ (2000) Implicit Learning of Tonality: A Self-Organizing Approach. Psychol Rev 107:885–913.

Trehub SE (2003) The developmental origins of musicality. Nat Neurosci 6:669–673.

Wong PCM, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 10:420–422.