american physiological society

JNP JOURNAL OF NEUROPHYSIOLOGY

J Neurophysiol 125: 2237–2263, 2021.
First published February 18, 2021; doi:10.1152/jn.00588.2020

RESEARCH ARTICLE

Sensory Processing

Music-selective neural populations arise without musical training

Dana Boebinger,[1,2,3] Sam V. Norman-Haignere,[4,5] Josh H. McDermott,[1,2,3,6] and Nancy Kanwisher[2,3,6]

[1]*Speech and Hearing Bioscience and Technology, Harvard University, Cambridge, Massachusetts;* [2]*Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts;* [3]*McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts;* [4]*Laboratoire des Sytèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, CNRS, Paris France;* [5]*Zuckerman Institute for Brain Research, Columbia University, New York, New York; and* [6]*Center for Brains, Minds, and Machines, Massachusetts Institute of Technology, Cambridge, Massachusetts*

Abstract

Recent work has shown that human auditory cortex contains neural populations anterior and posterior to primary auditory cortex that respond selectively to music. However, it is unknown how this selectivity for music arises. To test whether musical training is necessary, we measured fMRI responses to 192 natural sounds in 10 people with almost no musical training. When voxel responses were decomposed into underlying components, this group exhibited a music-selective component that was very similar in response profile and anatomical distribution to that previously seen in individuals with moderate musical training. We also found that musical genres that were less familiar to our participants (e.g., Balinese *gamelan*) produced strong responses within the music component, as did drum clips with rhythm but little melody, suggesting that these neural populations are broadly responsive to music as a whole. Our findings demonstrate that the signature properties of neural music selectivity do not require musical training to develop, showing that the music-selective neural populations are a fundamental and widespread property of the human brain.

NEW & NOTEWORTHY We show that music-selective neural populations are clearly present in people without musical training, demonstrating that they are a fundamental and widespread property of the human brain. Additionally, we show music-selective neural populations respond strongly to music from unfamiliar genres as well as music with rhythm but little pitch information, suggesting that they are broadly responsive to music as a whole.

auditory cortex; decomposition; expertise; fMRI; music

INTRODUCTION

Music is uniquely and universally human (1), and musical abilities arise early in development (2). Recent evidence has revealed neural populations in bilateral nonprimary auditory cortex that respond selectively to music and thus seem likely to figure importantly in musical perception and behavior (3–9). How does such selectivity arise? Most members of Western societies have received at least some explicit musical training in the form of lessons or classes, and it is possible that this training leads to the emergence of music-selective neural populations. However, most Western individuals, including nonmusicians, also implicitly acquire knowledge of musical structure from a lifetime of exposure to music (10–15). Thus, another possibility is that this type of passive experience with music is sufficient for the development of cortical music selectivity. The roles of these two forms of musical experience in the neural representation of music are not understood. Here, we directly test whether explicit musical training is necessary for the development of music-selective neural responses, by testing whether music-selective responses are robustly present—with similar response characteristics and anatomical distribution—in individuals with little or no explicit training.

Why might explicit musical training be necessary for neural tuning to music? The closest analogy in the visual domain is learning to read, where several studies have shown that selectivity to visual orthography (16) arises in high-level visual cortex only after children are taught to read (17, 18). In addition, exposure to specific sounds can elicit long-term

Correspondence: D. Boebinger (danaboebinger@g.harvard.edu).
Submitted 12 October 2020 / Revised 12 February 2021 / Accepted 12 February 2021

changes in auditory cortex, such as sharper tuning of individual neurons (19–21) and expansion of cortical maps (21–23). These changes occur primarily for behaviorally relevant stimulus features (23–27) related to the intrinsic reward value of the stimulus (26, 28, 29), and thus are closely linked to the neuromodulatory system (30–32). Additionally, the extent of cortical map expansion is correlated with the animal's subsequent improvement in behavioral performance (21–23, 33–35). Most of this prior work on experience-driven plasticity in auditory cortex has been done in animals undergoing extensive training on simple sensory sound dimensions, and it has remained unclear how the results from this work might generalize to humans in more natural settings with higher-level perceptual features. Musical training in humans meets virtually all of these criteria for eliciting functional plasticity: playing music requires focused attention, fine-grained sensory-motor coordination, it is known to engage the neuromodulatory system (36–38), and expert musicians often begin training at a young age and hone their skills over many years.

Although many prior studies have measured neural changes as a result of auditory experience (39, 40), including comparing responses in musicians and nonmusicians (41–54), it remains unclear whether any tuning properties of auditory cortex depend on musical training. Previous studies have found that fMRI responses to music are larger in musicians compared with nonmusicians in posterior superior temporal gyrus (41, 48, 52). However, these responses were not shown to be selective for music, leaving the relationship between musical training and cortical music selectivity unclear.

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. (7) used voxel decomposition to infer a small number of component response profiles that collectively explained voxel responses throughout auditory cortex to large set of natural sounds. This approach makes it possible to disentangle the responses of neural populations that overlap within voxels and has previously revealed a component with clear selectivity for music compared with both other real-world sounds (7) and synthetic control stimuli matched to music in many acoustic properties (55). These results have recently been confirmed by intracranial recordings, which show individual electrodes with clear selectivity for music (6). Although Norman-Haignere et al. (7) did not include actively practicing musicians, many of the participants had substantial musical training earlier in their lives.

Here, we test whether music selectivity arises only after explicit musical training. To this end, we probed for music selectivity in people with almost no musical training. On the one hand, if explicit musical training is necessary for the existence of music-selective neural populations, music selectivity should be weak or absent in these nonmusicians. If, however, music selectivity does not require explicit training but rather is either innate or arises as a consequence of passive exposure to music, then we would expect to see robust music selectivity even in the nonmusicians. A group of highly trained musicians was also included for comparison. 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.

Note that this is not a traditional group comparison study contrasting musicians and nonmusicians in an attempt to ascertain whether musical training has any detectable effect on music selective neural responses, as it would be unrealistic to collect the amount of data that would be necessary for a direct comparison between groups (see *Direct Group Comparisons of Music Selectivity* in the APPENDIX). Rather, our goal was to ask whether the key properties of music selectivity described in our earlier study are present in each group when analyzed separately, thus determining whether or not explicit training is necessary for the emergence of music selective responses in the human brain.

MATERIALS AND METHODS

Participants

Twenty young adults (14 female, mean = 24.7 yr, SD = 3.8 yr) participated in the experiment: 10 nonmusicians (6 female, mean = 25.8 yr, SD = 4.1 yr) and 10 musicians (8 female, mean = 23.5 yr, SD = 3.3 yr). This number of participants was chosen because our previous study (7) was able to infer a music-selective component from an analysis of 10 participants. Although these previous participants were described as "nonmusicians" (defined as no formal training in the 5 years preceding the study), many of the participants had substantial musical training earlier in life. We therefore used stricter inclusion criteria to recruit 10 musicians and 10 nonmusicians for the current study.

To be classified as a nonmusician, participants were required to have less than 2 years of total music training, which could not have occurred either before the age of seven or within the last 5 years. Out of the 10 nonmusicians 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 2 years of training (starting at age 10). These training measures do not include any informal "music classes" included in participants' required elementary school curriculum, because (at least in the United States) these classes are typically compulsory, are only for a small amount of time per week (e.g., 1 h), and primarily consist of simple sing-a-longs. Inclusion criteria for musicians were beginning formal training before the age of seven (56) and continuing training until the current day. Our sample of 10 musicians had an average of 16.30 years of training (ranging from 11–23 years, SD = 2.52). Details of participants' musical experience can be found in Table 1.

Nonparametric Wilcoxon rank sum tests indicated that there were no significant group differences in median age (musician median = 24.0 yr, SD = 3.3, nonmusician median = 25.0 yr, SD = 4.1, $Z = -1.03$, $P = 0.30$, effect size $r = -0.23$), postsecondary education (i.e., formal education after high school; musician median = 6.0 yr, SD = 8.2 yr, nonmusician median = 6.5 yr, SD = 7.5 yr, $Z = -0.08$, $P = 0.94$, effect size $r = -0.02$), or socioeconomic status as measured by the Barrett Simplified Measure of Social Status questionnaire (BSMSS; Ref. 57; musician median = 54.8, SD = 7.1, nonmusician median = 53.6, SD = 15.4, $Z = 0.30$, $P = 0.76$, effect size $r = 0.07$). Note that we report the group standard deviations because this measure is more robust than the interquartile

Table 1. *Details of participants' musical backgrounds and training, as measured by a self-report questionnaire*

	Subject No.	Instrument	Age of Onset (Yr)	Years of Lessons	Years of Regular Practice	Years of Training	Hours of Weekly Practice	Hours of Daily Music Listening
Nonmusicians	1			0	0	0	0	0
	2		20	1	0	1	0	2
	3			0	0	0	0	0.25
	4			0	0	0	0	5
	5			0	0	0	0	1
	6			0	0	0	0	1.5
	7		10	1.5	0	1.5	0	1
	8			0	0	0	0	0.1
	9			0	0	0	0	0.5
	10			0	0	0	0	5
Mean			24.4	0.3	0	0.3	0	1.64
Musicians	11	Violin	7	11	11	11	2	2
	12	Piano, French horn	5	7	18	12	6	4
	13	Piano, Bass	6	12	18	18	2	2
	14	Piano	3	15	15	15	2	2.5
	15	Piano, Violin, Viola	5	17	19	20	3	4
	16	Piano, Flute	5	13	13	13	12	3
	17	Flute	7	12	13	15	15	3
	18	Piano, Cello, Flute	4	19	15	18	25	1
	19	Violin	3	18	18	18	4	10
	20	Violin, Clarinet	4	13	8	23	4	5
Mean			5.2	13.7	14.8	16.3	7.1	3.65
Grand mean			14.8	7	7.4	8.3	3.55	2.64

range with our modest sample size of 10 participants per group. All participants were native English speakers and had normal hearing (audiometric thresholds <25 dB HL for octave frequencies 250 Hz to 8 kHz), as confirmed by an audiogram administered during the course of this study. The study was approved by Massachusetts Institute of Technology's (MIT) human participants review committee (Committee on the Use of Humans as Experimental Subjects), and written informed consent was obtained from all participants.

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 nonmusicians, including frequency discrimination, sensorimotor synchronization, melody discrimination, and "sour note" detection. As predicted, musician participants outperformed nonmusician participants on all behavioral psychoacoustic tasks. See Appendix for more details, and Fig. A1 for participants' performance on these behavioral tasks.

Study Design

Each participant underwent a 2-h behavioral testing session as well as three 2-h fMRI scanning sessions. During the behavioral session, participants completed an audiogram to rule out the possibility of hearing loss, filled out questionnaires about their musical experience, and completed a series of basic psychoacoustic tasks described in the APPENDIX.

Natural Sound Stimuli for fMRI Experiment

Stimuli consisted of 2-s clips of 192 natural sounds. These sounds included the 165-sound stimulus set used in Ref. 7, which broadly sampled frequently heard and recognizable sounds from everyday life. Examples can be seen in Fig. 1A. This set of 165 sounds was 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 compare responses to more versus less familiar musical genres. Stimuli were ramped on and off with a 25-ms 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 yr, SD = 11.9 yr) categorized each stimulus into one of 14 different categories. The categories were taken from Ref. 7, 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 ensure that American listeners were indeed less familiar with the non-Western music stimuli chosen for this experiment, but that they still perceived the stimuli as "music." In this experiment, 188 participants (75 females; mean age = 36.6 yr, SD = 10.5 yr) 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.

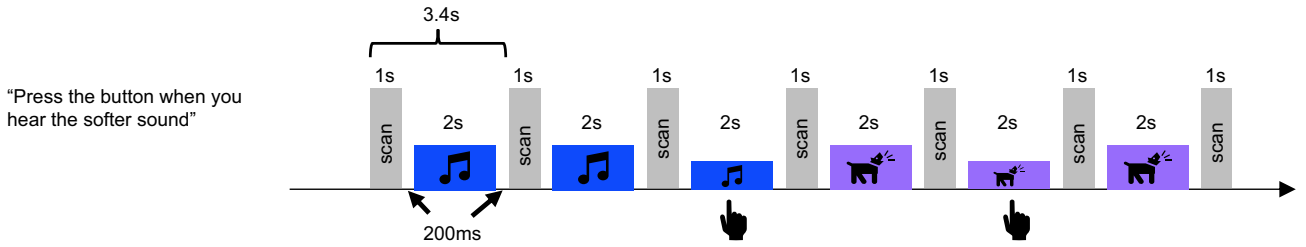
fMRI Data Acquisition and Preprocessing

Similar to the design of Ref. 7, sounds were presented during scanning in a "mini-block design," in which each 2-s natural sound was repeated three times in a row. Sounds were repeated because we have found this makes it easier to detect reliable hemodynamic signals. We used fewer repetitions than in our prior study (3 vs. 5), because we wanted to test a larger number of sounds and because we observed similarly reliable responses using fewer repetitions in pilot experiments. Each stimulus was presented in silence, with a single fMRI volume collected between each repetition [i.e., "sparse scanning" (58)]. To

A Stimulus Set of Commonly Heard Natural Sounds

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

B Scanning Procedure and Task Structure



C Voxel Decomposition Procedure

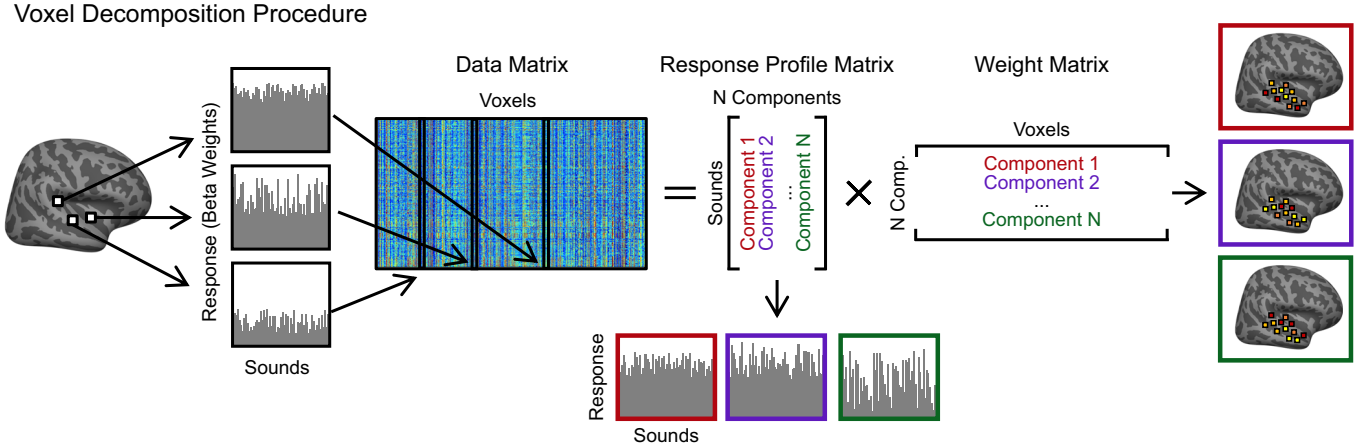


Figure 1. Experimental design and voxel decomposition method. *A*: fifty examples from the original set of 165 natural sounds used in Ref. 7 and in the current study, ordered by how often participants reported hearing them in daily life. An additional 27 music stimuli were added to this set of 165 for the current experiment. *B*: scanning paradigm and task structure. Each 2-s sound stimulus was repeated three times consecutively, 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. *C*: diagram depicting the voxel decomposition method, reproduced from Ref. 7. 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 × 26,792 voxels). This matrix is then factorized into a response profile matrix (192 sounds × N components) and a voxel weight matrix (N components × 26,792 voxels).

encourage participants to pay attention to the sounds, either the second or third repetition in each "mini-block" was 12 dB quieter (presented at 67 dB SPL), and participants were instructed to press a button when they heard this quieter sound (Fig. 1*B*). Overall, participants performed well on this task (musicians: mean = 92.06%, SD = 5.47%; nonmusicians: 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-min 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 s, and the 2-s stimuli were

presented during periods of silence between each acquisition, with a 200-ms buffer of silence before and after stimulus presentation. As a consequence, one brain volume was collected every 3.4 s (1 s + 2 s + 0.2*2 s; TR = 3.4 s, TA = 1.02 s, TE = 33 ms, 90 degree flip angle, 4 discarded initial acquisitions). Each functional acquisition consisted of 48 roughly axial slices (oriented parallel to the anterior-posterior commissure line) covering the whole brain, each slice being 3 mm thick and having an in-plane resolution of 2.1 × 2.1mm (96 × 96 matrix, 0.3-mm slice gap). A simultaneous multislice (SMS) acceleration factor of 4 was used to minimize acquisition time (TA = 1.02 s). To localize functional activity, a high-resolution anatomical T1-weighted image was obtained for every participant (TR = 2.53 s, voxel size: 1 mm^3, 176 slices, 256 × 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; Refs. 59, 60). Motion correction and function-to-anatomical registration was done separately for each run. Preprocessed data were then resampled to the cortical surface reconstruction computed by FreeSurfer (61) and smoothed on the surface using a 3-mm full-width half-maximum (FWHM) kernel to improve signal-to-noise ratio (SNR). The data were then downsampled to a 2-mm isotropic grid on the FreeSurfer-flattened cortical surface.

Next, we estimated the response to each of the 192 stimuli using a general linear model (GLM). Each stimulus mini-block was modeled as a boxcar function convolved with a canonical hemodynamic response function (HRF). The model also included six motion regressors and a first-order polynomial noise regressor to account for linear drift in the baseline signal. Note that this analysis differs from our prior paper (7), in which signal averaging was used in place of a GLM. We made this change because blood oxygen level-dependent (BOLD) responses were estimated more reliably using an HRF-based GLM, potentially due to the use of shorter stimulus blocks causing more overlap between BOLD responses to different stimuli.

Voxel Selection

The first step of this analysis method is to determine which voxels serve as input to the decomposition algorithm. All analyses were carried out on voxels within a large anatomical constraint region encompassing bilateral superior temporal and posterior parietal cortex (Fig. A2), as in Ref. 7. In practice, the vast majority of voxels with a robust and reliable response to sound fell within this region (Fig. A3), which explains why our results were very similar with and without this anatomical constraint (Fig. A4). Within this large anatomical region, we selected voxels that met two criteria. First, they displayed a significant ($P < 0.001$, uncorrected) response to sound (pooling over all sounds compared with silence). This consisted of 51.45% of the total number of voxels within our large constraint region. Second, they 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 Ref. 7 to measure the reliability across split halves of our data. This reliability measure differs from a Pearson correlation in that it assigns high values to voxels that respond consistently to sounds and does not penalize them even if their response does not vary much between sounds, which is the case for many voxels within primary auditory cortex:

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

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

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ indicate the vector of beta weights from a single voxel for the 192 sounds, estimated separately for the two halves of the data ($\boldsymbol{v}_1$ = first three repetitions from runs 1–24, $\boldsymbol{v}_2$ = last three repetitions from runs 25–48), and || || is the L2 norm. Note that these equations differ from *Eq. 1* and *2* in Ref. 7, because the equations as reported in that paper contained a typo: the L2-norm terms were not squared. 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, the median number of voxels per participant = 1,286, SD = 254 (Fig. A2*A*). The number of selected voxels did not differ significantly between musicians (median = 1,216, SD = 200) and nonmusicians (median = 1,341, SD = 284; $Z = -1.40$, $P = 0.16$, effect size $r = -0.31$, two-tailed Wilcoxon rank sum test), and the anatomical location of the selected voxels was largely similar across groups (Fig. A2*B*). When visualizing each group's data on the cortical surface (Fig. 3, *C* and *D*), we chose which voxels to include by first averaging voxel responses across participants within each group, and then applying the same selection criteria to the averaged data.

Unlike our prior study, we collected whole brain data in this experiment and thus were able to repeat our analyses without any anatomical constraint. Although a few additional voxels outside of the mask do meet our selection criteria (Fig. A3), the resulting components are very similar to those obtained using the anatomical mask, both in response profiles (Fig. A4*A*; correlations ranging from $r = 0.91$ to $r > 0.99$, SD = 0.03) and voxel weights (Fig. A4*B*).

Decomposition Algorithm

The decomposition algorithm approximates the response of each voxel ($\boldsymbol{v}_1$) as the weighted sum of a small number of component response profiles that are shared across voxels (Fig. 1*B*):

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

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

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

$$\boldsymbol{D} \approx \boldsymbol{RW}$$

The method used to infer components was described in detail in our prior paper (7) and 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 among 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 Fig. A5). The method differs from most standard ICA algorithms in that it maximizes non-Gaussianity by directly minimizing the entropy of the component weight distributions across voxels as measured by a histogram, which is feasible due to the large number of voxels. Entropy is a natural measure to minimize because the Gaussian distribution has maximal entropy. The algorithm achieves this goal in two 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 $\boldsymbol{U}^k$ contains the response profiles of the top K principal components (192 sounds $\times$ K components), $\boldsymbol{V}^k$ contains the whitened voxel weights for these components (K components $\times$ 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. Specifically, we chose a value of K that balanced these two measures such that the set of components explained a large fraction of the voxel response variance (which increases monotonically with additional components) but still maintained good prediction accuracy (which decreases once additional components begin to cause overfitting). In practice, the plateau in the amount of explained variance coincided with the peak of the prediction accuracy.

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

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

where $\boldsymbol{W}$ is the rotated weight matrix ($K \times$ 26,792), $\boldsymbol{T}$ is an orthonormal rotation matrix ($K \times K$), and $\boldsymbol{W}[c,:]$ is the cth row of $\boldsymbol{W}$. We estimated entropy using a histogram-based method (62) 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:

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

The optimization is performed by iteratively rotating pairs of components to maximize negentropy, which is a simple algorithm that does not require the computation of gradients and is feasible for small numbers of components [the number of component pairs grows as $\binom{K}{2}$].

This voxel decomposition analysis was carried out on three different data sets: *1*) on voxels from the 10 musicians only, *2*) on voxels from the 10 nonmusicians only, and *3*) on voxels from all 20 participants. We note that the derivation of a set of components using this method is somewhat akin to a fixed-effects analysis, in that it concatenates participants' data and infers a single set of components to explain the data from all participants at once. However, the majority of the analyses that we carried out using these components (as described in the following paragraphs) involve deriving participant-specific metrics and investigating the consistency of effects across participants.

Measuring Component Selectivity

To quantify the selectivity of the music component, we measured the difference in mean response profile magnitude between music and nonmusic sounds, divided by their pooled standard deviation (Cohen's d). This measure provides a measure of the separability of the two sound categories within the response profile. We measured Cohen's d for several different pairwise comparisons of sound categories. In each case, the significance of the separation of the two stimulus categories was determined using a permutation test (permuting stimulus labels between the two categories 10,000 times). This null distribution was then fit with a Gaussian, a P-value from which was assigned to the observed value of Cohen's d.

Anatomical Component Weight Maps

To visualize the anatomical distribution of component weights, individual participants' component weights were projected onto the cortical surface of the standard FsAverage template, and a random effects analysis (t test) was performed to determine whether component weights were significantly greater than zero across participants at each voxel location. To visualize the component weight maps separately for musicians and nonmusicians, a separate random effects analysis was run for the participants in each group. To correct for multiple comparisons, we adjusted the false discovery rate [FDR, $c(V)$ = 1, q = 0.05] using the method from Genovese et al. (63).

We note that the details of the analyses and plotting conventions used for visualizing component weight maps differ from those of our previous study (7). These differences include the process involved in aggregating weights across subjects (Norman-Haignere et al. smoothed individual participants' data and then averaged across participants, whereas there was no smoothing or averaging across participants in the current study), the use of different measures of statistical significance of the component weights (a random effects analysis across subjects in the current study, a permutation test across sounds in Norman-Haignere et al.), and different thresholding (an FDR threshold of q = 0.05 in the current study, whereas the component weight maps in Norman-Haignere et al. showed the entire weight distribution and were not thresholded). We made these changes in the current study so that we could ask about the consistency of effects across participants and better visualize which voxels' component weights passed standard significance thresholds. However, we note that these changes cause the maps we regenerated from the Ref. 7 data (Fig. 3, *C* and *D*) to look somewhat different from those shown in the original paper (replotted in Fig. 5*A*).

Component Voxel Weights within Anatomical ROIs

In addition to projecting the weight distributions on the cortical surface, we summarized their anatomical distribution by measuring the mean component voxel weight within a set of standardized anatomical ROIs. To create these ROIs, a set of 15 parcels were selected from an atlas (64) to fully encompass the superior temporal plane and superior temporal gyrus (STG). To identify a small set of ROIs suitable for evaluating the music component weights in our current study, we superimposed these anatomical parcels onto the weights of the music component from our previously published study (7, shown in Fig. 5*A*), and then defined ROIs by selecting sets of the anatomically defined parcels that best correspond to regions of high vs. low music component weights (Fig. 5*B*). The mean component weights within these ROIs were computed separately for each participant and then averaged across participants for visualization purposes (e.g., Fig. 5*C*). We then ran a 4 (ROI) × 2 (hemisphere) repeated-measures ANOVA on these weights. A separate ANOVA was run for musicians and nonmusicians, to evaluate each group separately.

To compare the magnitude of the main effect of ROI with the main effect of hemisphere, we bootstrapped across participants, resampling participants 1,000 times. We reran the repeated-measures ANOVA on each sample, each time measuring the difference in the effect size for the two main effects, i.e., $\eta_{pROI}^2 - \eta_{pHemi}^2$. We then calculated the 95% confidence interval (CI) of this distribution of effect size differences. The significance of the difference in main effects was evaluated by determining whether or not each group's 95% CI for the difference overlapped with zero.

In addition, we ran a Bayesian repeated-measures ANOVA on these same data, implemented in JASP v.0.13.1 (78), using the default prior (Cauchy distribution, $r = 0.5$). Effects are reported as the Bayes Factor for inclusion (BF_{inc}) of each main effect and/or interaction, which is the ratio between the likelihood of the data given the model including the effect in question vs. the likelihood of the next simpler model without the effect in question.

RESULTS

Our primary question was whether cortical music selectivity is present in people with almost no musical training. To that end, we scanned 10 people with almost no musical training (Table 1) and used voxel decomposition to infer a small set of response components that could explain the observed voxel responses. We also included another set of 10 participants with extensive musical training for comparison. First, we asked whether the response components across all of auditory cortex reported by Norman-Haignere et al. (7) replicate in both nonmusicians and highly trained musicians when analyzed separately. Second, we examined the detailed characteristics of music selectivity in particular, to test whether its previously documented key properties are present in both nonmusicians and highly trained musicians. Third, we took advantage of our expanded stimulus set to look at additional properties of music selectivity, such as the response to musical genres that are less familiar to our Western participants.

Replication of Functional Components of Auditory Cortex from Norman-Haignere et al. in Musicians and in Nonmusicians

Replication of previous voxel decomposition results. We first tested the extent to which we would replicate the overall functional organization of auditory cortex reported by Norman-Haignere et al. (7) in people with almost no musical training, using the voxel decomposition method introduced in that paper. We also performed the same analysis on a group of highly trained musicians. Specifically, in every participant, we measured the response of voxels within auditory cortex to 192 natural sounds (Fig. 1, *A* and *B*; the average response of each voxel to each sound was estimated using a standard hemodynamic response function). Then, separately for nonmusicians and musicians, we used voxel decomposition to model the response of these voxels as the weighted sum of a small number of canonical response components (Fig. 1*C*). This method factorizes the voxel responses into two matrices: one containing the components' response profiles across the sound set and the second containing voxel weights specifying the extent to which each component contributes to the response of each voxel.

Since the only free parameter in this analysis is the number of components recovered, the optimal number of components was determined by measuring the fraction of the reliable response variance the components explain. In the previous study (7), six components were sufficient to explain over 80% of the reliable variance in voxel responses. We found the same to be true in both participant groups of the current study: six components were needed to optimally model the data from the 10 participants in separate analyses of each group. The six components explained 88.56% and 88.09% of the reliable voxel response variance for nonmusicians and musicians, respectively, after which the amount of explained variance for each additional component plateaued (Fig. A6).

Next, we examined the similarity of the components inferred from nonmusicians to the components from our previous study, comparing their responses to the 165 sounds common to both. Because the order of the components inferred using ICA holds no significance, we first used the Hungarian algorithm (65) to optimally reorder the components, maximizing their correlation with the components from our previous study. For comparisons of the response profile matrices of two groups of subjects, we matched components using the weight matrices; conversely, for comparisons involving the voxel weights, we matched components using the response profile matrices (see MATERIALS AND METHODS). In practice, the component matches were identical regardless of which matrices were used for matching. We also conducted the same analysis for the musicians, comparing the components derived from their data with those in our previous study.

For both nonmusicians and musicians, corresponding pairs of components were highly correlated with those from our previous study, with r values for nonmusicians' components ranging from 0.63 to 0.99 (Fig. 2, *A* and *B*) and from 0.66 to 0.99 for musicians' components (Fig. 2, *C* and *D*).

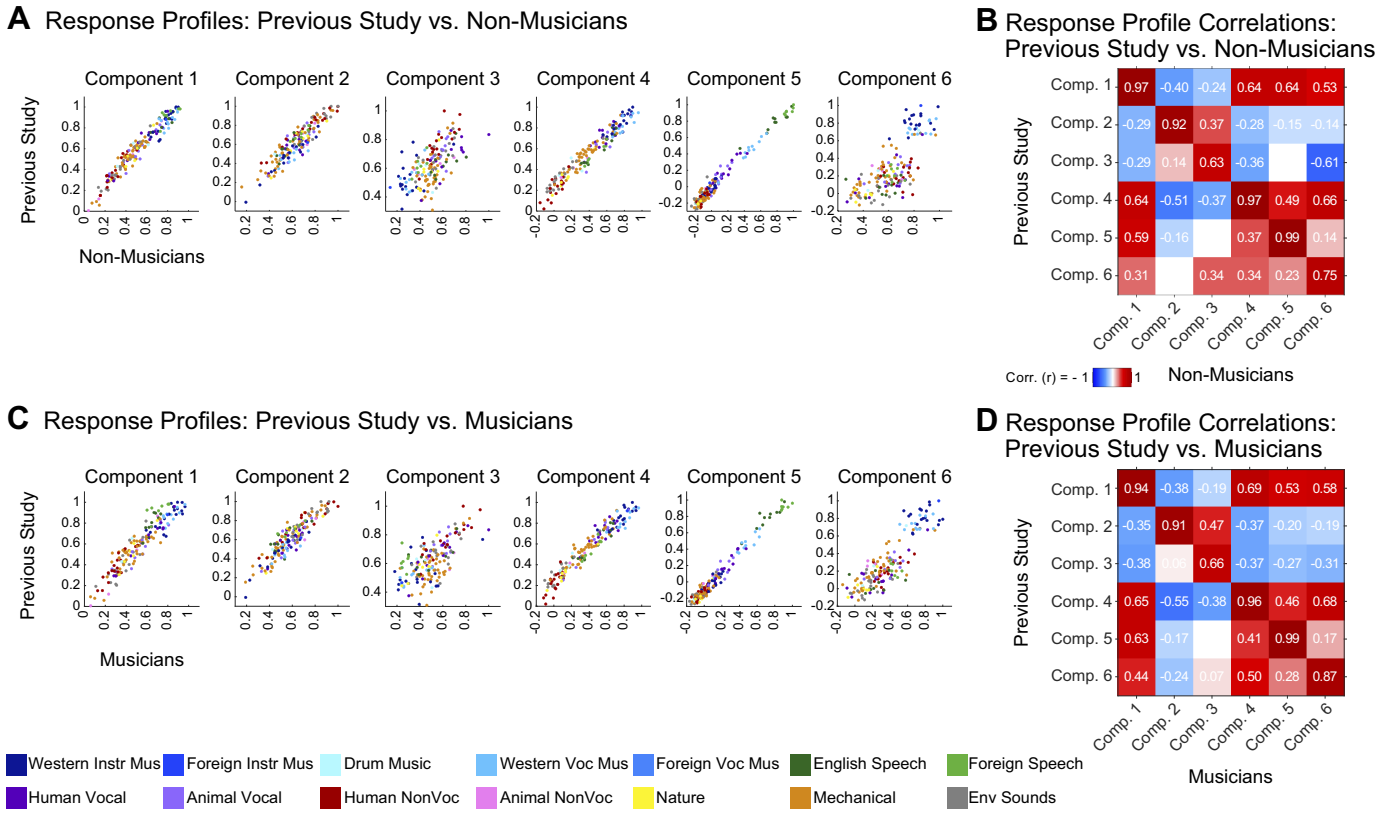


Figure 2. Replication of components from Ref. 7. *A*: scatterplots showing the correspondence between the component response profiles from the previous study ($n = 10$, *y*-axis) and those inferred from nonmusicians ($n = 10$, *x*-axis). The 165 sounds common to both studies are colored according to their semantic category, as determined by raters on Amazon Mechanical Turk. Note that the axes differ slightly between groups to make it possible to clearly compare the pattern of responses across sounds independent of the overall response magnitude. *B*: correlation matrix comparing component response profiles from the previous study (*y*-axis) and those inferred from nonmusicians ($n = 10$, *x*-axis). *C* and *D*: same as *A* and *B* but for musicians. Comp., component.

Component response profiles and selectivity for sound categories.

Four of the six components from our previous study captured expected acoustic properties of the sound set (e.g., frequency, spectrotemporal modulation; see Fig. A8*A* for analyses relating the responses of these components to audio frequency and spectrotemporal modulation) and were concentrated in and around primary auditory cortex (PAC), consistent with prior results (55, 66–72). The two remaining components responded selectively to speech (Fig. 3*A*, *left column*) and music (Fig. 3*B*, *left column*), respectively, and were not well accounted for using acoustic properties alone (Fig. A8*A*). The corresponding components inferred from nonmusicians (Fig. 3, *A* and *B*, *middle columns*) and musicians (Fig. 3, *A* and *B*, *right columns*) also show this category selectivity.

We note that the mean response profile magnitude for the music-selective component differed between groups, being lower in musicians than nonmusicians (Fig. 3*B*). This effect seems unlikely to be a consequence of musical training because the mean response magnitude was lower still in the previous study, whose participants had substantially less musical training than the musicians in the current study. Further, we have found that component response profile magnitude tends to vary depending on the method used to infer the components. For example, using a probabilistic parametric matrix factorization model instead of the simpler, nonparametric method presented throughout this paper resulted in components that had different mean responses despite otherwise being very similar to those obtained via ICA (see Appendix for details of the parametric model, and Fig. A10 for the component response profiles inferred using this method). Moreover, the information about music contained in the response as measured by the separability of music versus nonmusic sounds (Cohen's *d*) is independent of this overall mean response (see Fig. 4). For these reasons, we do not read much into this apparent difference between groups.

We also found similarities in the anatomical distribution of speech- and music-selective component weights between the previous study and both groups in the current study. The weights for the speech-selective component were concentrated in the middle portion of the superior temporal gyrus (midSTG, Fig. 3*C*), as expected based on previous reports (73–75). In contrast, the weights for the music-selective component were most prominent anterior to PAC in the planum polare, with a secondary cluster posterior to PAC in the planum temporale (Fig. 3*D*) (3, 7, 41, 48, 52, 76, 77).

Together, these findings show that we are able to twice replicate the overall component structure underlying auditory cortical responses described in our previous study (once for nonmusicians and once for musicians). Further, both nonmusicians and musicians show category-selective components that are largely similar to those in our previous study, including a single component that appears to be selective for music.

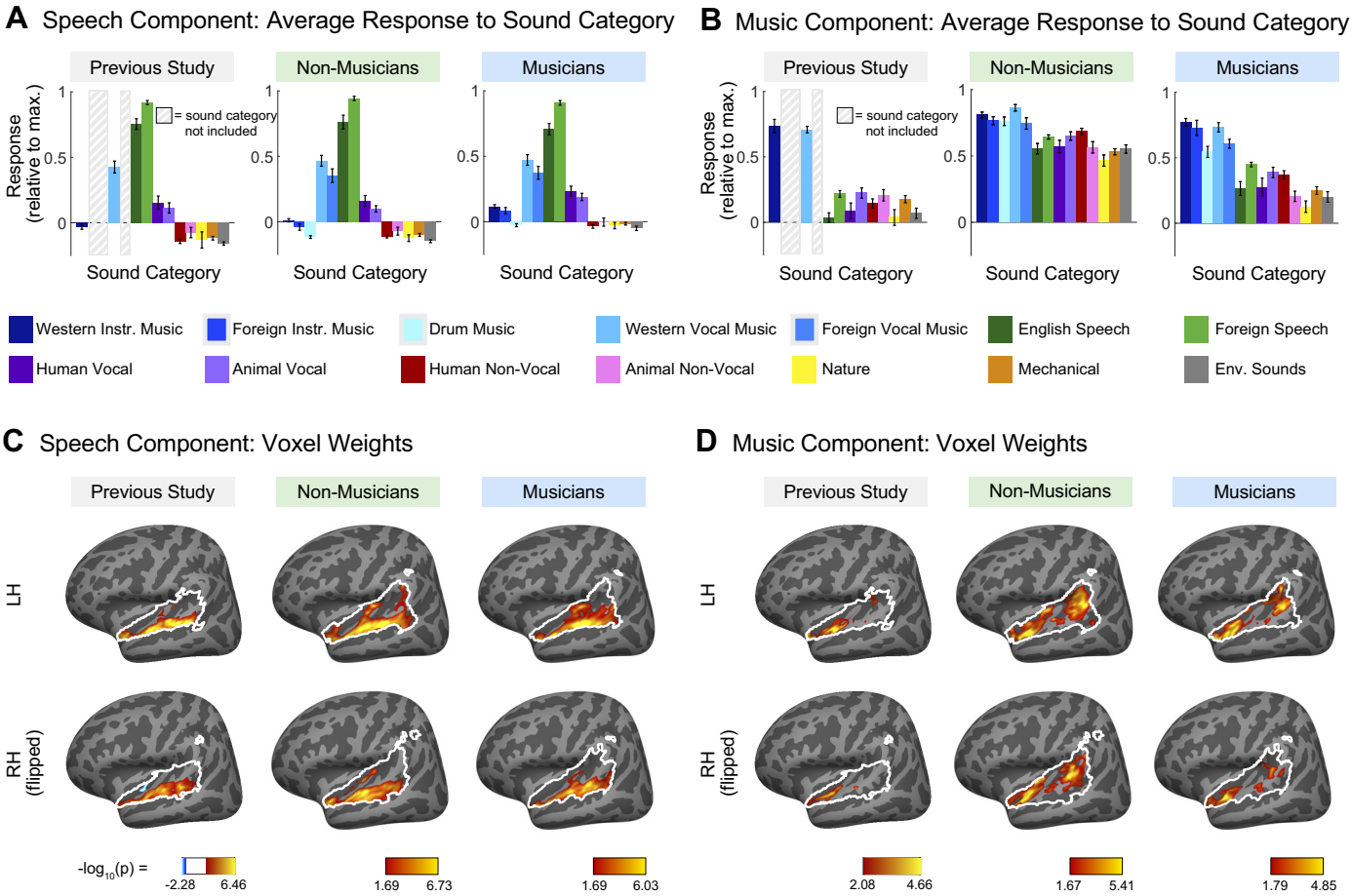


Figure 3. Comparison of speech-selective and music-selective components for participants from previous study (n = 10), nonmusicians (n = 10), and musicians (n = 10). *A* and *B*: component response profiles averaged by sound category (as determined by raters on Amazon Mechanical Turk). *A*: the speech-selective component responds highly to speech and music with vocals, and minimally to all other sound categories. Shown separately for the previous study (*left*), nonmusicians (*middle*), and musicians (*right*). Note that the previous study contained only a subset of the stimuli (165 sounds) used in the current study (192 sounds) so some conditions were not included and are thus replaced by a gray rectangle in the plots and surrounded by a gray rectangle in the legend. *B*: the music-selective component (*right*) responds highly to both instrumental and vocal music, and less strongly to other sound categories. Note that "Western Vocal Music" stimuli were sung in English. We note that the mean response profile magnitude differs between groups, but that selectivity as measured by separability of music and nonmusic is not affected by this difference (see text for explanation). For both *A* and *B*, error bars plot one standard error of the mean across sounds from a category, computed using bootstrapping (10,000 samples). *C*: spatial distribution of speech-selective component voxel weights in both hemispheres. *D*: spatial distribution of music-selective component voxel weights. Color denotes the statistical significance of the weights, computed using a random effects analysis across subjects comparing weights against 0; P values are logarithmically transformed ($-\log_{10}[P]$). The white outline indicates the voxels that were both sound-responsive (sound vs. silence, $P < 0.001$ uncorrected) and split-half reliable ($r > 0.3$) at the group level (see MATERIALS AND METHODS for details). The color scale represents voxels that are significant at FDR $q = 0.05$, with this threshold 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 right hemisphere (*bottom rows*) is flipped to make it easier to visually compare weight distributions across hemispheres. Note that the secondary posterior cluster of music component weights is not as prominent in this visualization of the data from Ref. 7 due to the thresholding procedure used here; we found in additional analyses that a posterior cluster emerged if a more lenient threshold is used. FDR, false discovery rate; LH, left hemisphere; RH, right hemisphere.

Characterizing Music Selectivity in Nonmusicians and Musicians Separately

We next asked whether nonmusicians exhibited the signature response characteristics of music selectivity documented in our prior paper: *1*) the music component response profile showed a high response to both instrumental and vocal music and a low response to all other categories, including speech; *2*) music component voxel weights were highest in anterior superior temporal gyrus (STG), with indications of a secondary concentration of weights in posterior STG, and low weights in both PAC and lateral STG; and *3*) music component voxel weights had a largely bilateral distribution. We examined whether each of these properties was present in the components separately inferred from nonmusicians and musicians.

Response profiles show selectivity for both instrumental and vocal music.

The defining feature of the music-selective component from Ref. 7 was that its response profile showed a very high response to stimuli that had been categorized by humans as "music," including both instrumental music and music with vocals, relative to nonmusic stimuli, including speech (Fig. 3*B*, *left column*). The category-averaged responses of the music-selective component showed similar effects in nonmusicians and highly trained musicians (Fig. 3*B*, *center* and *right columns*).

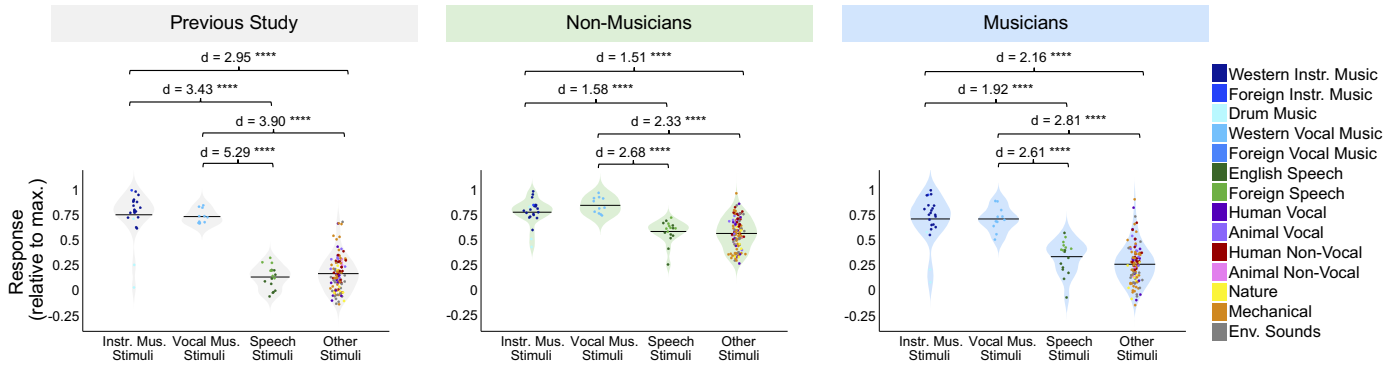


Figure 4. Separability of sound categories in music-selective components of nonmusicians and musicians. Distributions of *1*) instrumental music stimuli, *2*) vocal music stimuli, *3*) speech stimuli, and *4*) other stimuli within the music component response profiles from our previous study ($n = 10$; *left*, gray shading), as well as those inferred from nonmusicians ($n = 10$; *center*, green shading) and musicians ($n = 10$; *right*, blue shading). The mean for each stimulus category is indicated by the horizontal black line. The separability between pairs of stimulus categories (as measured using Cohen's *d*) is shown above each plot. The 165 individual sounds are colored according to their semantic category. Stimuli consisted of instrumental music ($n = 22$), vocal music ($n = 11$), speech ($n = 17$), and other ($n = 115$). See Table 2 for results of pairwise comparisons indicated by brackets; ****Significant at $P < 0.0001$, two-tailed.

To quantify this music selectivity, we measured the difference in mean response profile magnitude between music and nonmusic sounds, divided by their pooled standard deviation (Cohen's *d*). So that we could compare across our previous and current experiments, this was done using the set of 165 sounds that were common to both studies. We measured Cohen's *d* separately for several different pairwise comparisons: *1*) instrumental music vs. speech stimuli, *2*) instrumental music vs. other nonmusic stimuli, *3*) vocal music vs. speech stimuli, and *4*) vocal music vs. other nonmusic stimuli (Fig. 4). In each case, the significance of the separation of the two stimulus categories was determined using a nonparametric test permuting stimulus labels 10,000 times. All four of these statistical comparisons were highly significant for nonmusicians when analyzed separately (all Ps $< 10^{-5}$, Table 2). This result shows that the music component is highly music-selective in nonmusicians, in that it responds highly to both instrumental and vocal music, and significantly less to both speech and other nonmusic sounds. Similar results were also found for musicians (all Ps $< 10^{-5}$, Table 2). We note that the selectivity of the music component inferred from nonmusicians seems to be slightly lower than that of the component inferred from musicians, but we are not sufficiently powered to directly test for differences in selectivity between groups (see *Direct Group Comparisons of Music Selectivity* in the Appendix). It's also true that the values of Cohen's *d* tend to be somewhat larger for the music component from Ref. 7 than for the components inferred from both groups of participants in the current study. It is not clear why this is the case, but it is more likely to be due to slight methodological differences between the experiments than an effect of musical training, because the participants in our previous study had intermediate levels of musical training.

Music component weights concentrated in anterior and posterior STG.

A second notable property of the music component from Ref. 7 was that the weights were concentrated in distinct regions of nonprimary auditory cortex, with the most prominent cluster

Table 2. *Results of pairwise comparisons between stimulus categories shown in Fig. 4*

Subject Group	Pairwise Comparison	Stimulus Category	Mean	SD	Cohen's *d*	*P* value
Nonmusicians	1	Instrumental music Speech	0.78 0.59	0.13 0.11	1.58	2.14E-06
	2	Instrumental music Other	0.78 0.57	0.13 0.15	1.51	1.88E-10
	3	Vocal music Speech	0.85 0.59	0.08 0.11	2.68	4.54E-11
	4	Vocal music Other	0.85 0.57	0.08 0.15	2.33	1.87E-12
Musicians	1	Instrumental music Speech	0.72 0.34	0.22 0.16	1.92	1.19E-08
	2	Instrumental music Other	0.72 0.27	0.22 0.19	2.16	7.40E-20
	3	Vocal music Speech	0.72 0.34	0.12 0.16	2.61	8.38E-11
	4	Vocal music Other	0.72 0.27	0.12 0.19	2.81	3.62E-18

The significance of the separation of each pair of stimulus categories was determined using a nonparametric test permuting stimulus labels 10,000 times. Stimuli consisted of instrumental music ($n = 22$), vocal music ($n = 11$), speech ($n = 17$), and other ($n = 115$).

located in anterior STG, and a secondary cluster located in posterior STG (at least in the left hemisphere). Conversely, music component weights were low in primary auditory cortex (PAC) and intermediate in nonprimary lateral STG (see Fig. 5*A*).

To assess whether these anatomical characteristics were evident for the music components inferred from our nonmusician and musician participants, we superimposed standardized anatomical parcels (64) on the data and defined anatomical regions of interest (ROIs) by selecting four sets of these anatomically defined parcels that best corresponded to anterior nonprimary, primary, and lateral nonprimary auditory cortex (Fig. 5*B*). We then calculated the average music component weight for each individual participant within each of these four anatomical ROIs, separately for each hemisphere (Fig. 5*C*). This was done separately for nonmusicians and musicians, using their respective music components.

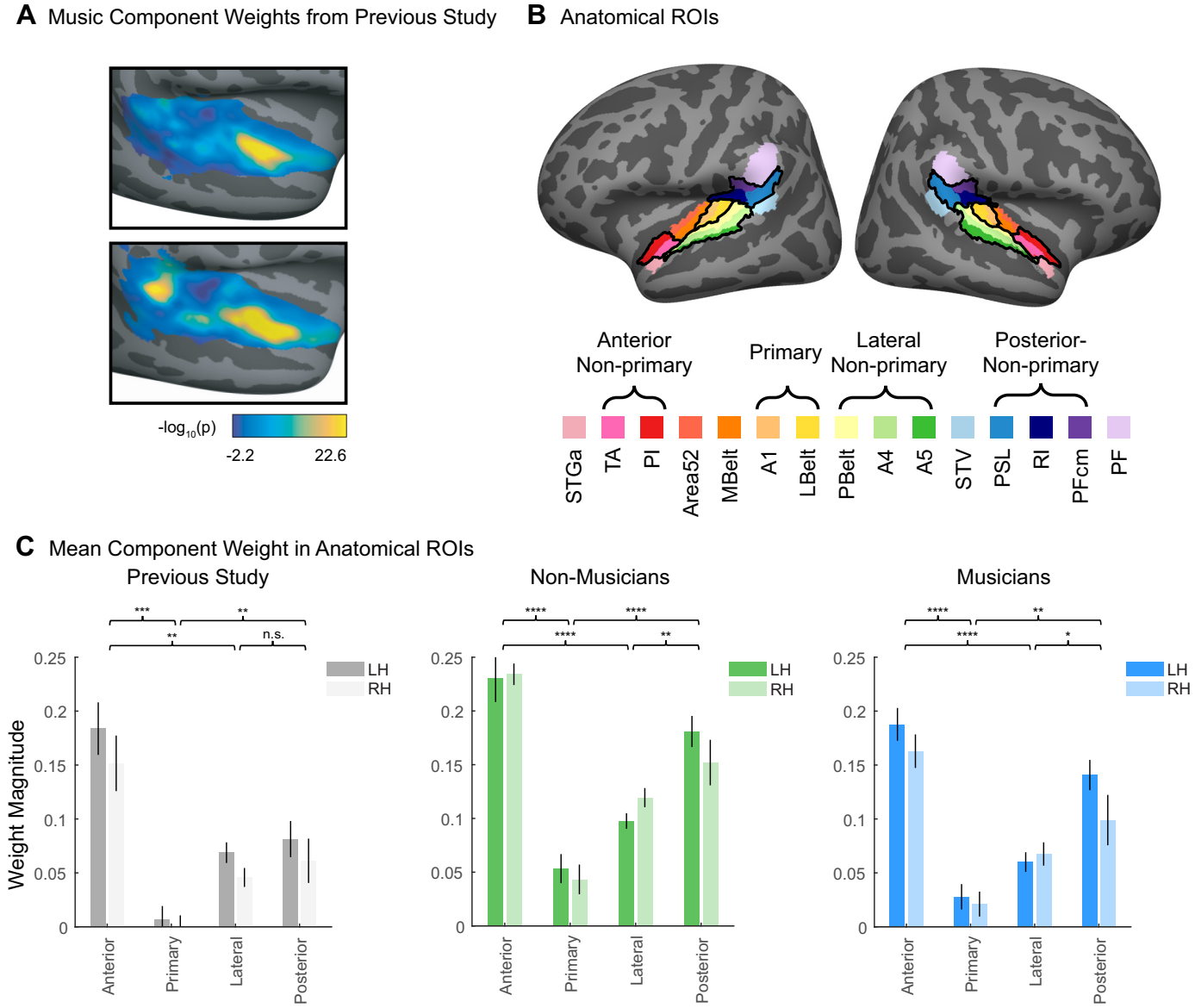


Figure 5. Quantification of bilateral anterior/posterior concentration of voxel weights for the music-selective components inferred in nonmusicians and musicians separately. *A*: music component voxel weights, reproduced from Ref. 7. See MATERIALS AND METHODS for details concerning the analysis and plotting conventions from our previous paper. *B*: fifteen standardized anatomical parcels were selected from Ref. 64, chosen to fully encompass the superior temporal plane and superior temporal gyrus (STG). To come up with a small set of ROIs to use to evaluate the music component weights in our current study, we superimposed these anatomical parcels onto the weights of the music component from our previously published study (7), and then defined ROIs by selecting sets of the anatomically defined parcels that correspond to regions of high (anterior nonprimary, posterior nonprimary) vs. low (primary, lateral nonprimary) music component weights. The anatomical parcels that comprise these four ROIs are indicated by the brackets and outlined in black on the cortical surface. *C*: mean music component weight across all voxels in each of the four anatomical ROIs, separately for each hemisphere, and separately for our previous study ($n = 10$; *left*, gray shading), nonmusicians ($n = 10$; *center*, green shading), and musicians ($n = 10$; *right*, blue shading). A repeated-measures ROI × hemisphere ANOVA was conducted for each group separately. Error bars plot one standard error of the mean across participants. Brackets represent pairwise comparisons that were conducted between ROIs with expected high vs. low component weights, averaged over hemisphere. See Table 3 for full results of pairwise comparisons, and Fig. A9 for component weights from all 15 anatomical parcels. *Significant at $P < 0.05$, two-tailed; **Significant at $P < 0.01$, two-tailed; ***Significant at $P < 0.001$, two-tailed; ****Significant at $P < 0.0001$, two-tailed. Note that because of our prior hypotheses and the significance of the omnibus F test, we did not correct for multiple comparisons. LH, left hemisphere; RH, right hemisphere; ROI, region of interest.

For each group, a 4 (ROI) × 2 (hemisphere) repeated measures ANOVA on these mean component weights showed a significant main effect of ROI for both nonmusicians [$F(3,27)$ = 50.12, P = 3.63e-11, η^2_p = 0.85] and musicians [$F(3,27)$ = 19.62, P = 5.90e-07, η^2_p = 0.69]. Pairwise comparisons showed that for each group, component weights were significantly higher in the anterior and posterior nonprimary ROIs than both the primary and lateral nonprimary ROIs when averaging over hemispheres (Table 3; nonmusicians: all Ps < 0.004, musicians: all Ps < 0.03).

These results show that in both nonmusicians and musicians, music selectivity is concentrated in anterior and posterior STG and present to a lesser degree in lateral STG, and only minimally in PAC.

Music component weights are bilaterally distributed. A third characteristic of the previously described music selectivity is that it was similarly distributed across hemispheres, with no obvious lateralization (7). The repeated-measures ANOVA described in the previous section showed no evidence of lateralization in either nonmusicians or musicians [Fig. 5C; nonmusicians: $F(1,9)$ = 0.15, P = 0.71, η^2_p = 0.02; musicians: $F(1,9)$ = 2.43, P = 0.15, η^2_p = 0.21]. Furthermore, for both groups, the effect size of ROI within a hemisphere was significantly larger than the effect size of hemisphere [measured by bootstrapping across participants to get 95% CIs around the difference in the effect size for the two main effects, i.e., $\eta_{pROI}^2 - \eta_{pHemi}^2$; the significance of the difference in main effects was evaluated by determining whether or not each group's 95% CI for the difference overlapped with zero: nonmusicians' CI: (0.37, 0.89), musicians' CI: (0.16, 0.82)].

Because the lack of a significant main effect of hemisphere could be due to insufficient statistical power, we ran a Bayesian version of the repeated-measures ANOVA, which allows us to quantify evidence both for and against the null hypothesis that there was not a main effect of hemisphere. We used JASP (78), with its default prior (Cauchy distribution, r = 0.5), and computed the Bayes Factor for inclusion of each main effect and/or interaction (the ratio between the likelihood of the data given the model including the effect in question vs. the likelihood of the next simpler model without the effect in question, with values further from 1 providing stronger evidence in favor of one model or the other). We found no evidence for a main effect of hemisphere (Bayes factor of inclusion, BF_{incl} = 0.77 for musicians, suggestive of weak evidence against inclusion; BF_{incl} = 0.24 for nonmusicians, suggestive of moderate evidence against inclusion, using the guidelines suggested by Ref. 79). By contrast, the main effect of ROI was well supported (BF_{incl} for ROI for nonmusicians = 1.02e17, and for musicians = 6.11e12, both suggestive of extreme evidence in favor of inclusion).

Neither group showed a significant ROI × hemisphere interaction [nonmusicians: $F(3,27)$ = 1.48, P = 0.24, η^2_p = 0.14; musicians: $F(3,27)$ = 2.24, P = 0.11, η^2_p = 0.20]. This was also the case for the Bayesian repeated-measures ANOVA, in which the Bayes Factors for the interaction between ROI and hemisphere provided weak evidence against including the interaction term in the models (BF_{incl} = 0.38 for nonmusicians, BF_{incl} = 0.36 for musicians).

The fact that the music component inferred from nonmusicians exhibits all of the previously described features of music selectivity (7) suggests that explicit musical training is not necessary for a music-selective neural population to arise in the human brain. In addition, the results from musicians suggest that that the signature properties of music selectivity are not drastically altered by extensive musical experience. Both groups exhibited a single response component selective for music. And in both groups, this selectivity was present for both instrumental and vocal music, was localized to anterior and posterior nonprimary auditory cortex, and was present bilaterally.

New Insights into Music Selectivity: Music-Selective Regions of Auditory Cortex Show High Responses to Drum Rhythms and Unfamiliar Musical Genres

Because our experiment utilized a broader stimulus set than the original study (7), we were able to use the inferred components to ask additional questions about the effect of experience on music selectivity, as well as gain new insights into the nature of cortical music selectivity. The set of natural sounds used in this study included a total of 60 music

Table 3. *Results of pairwise comparisons between mean weights in ROIs shown in Fig. 5C*

Subject Group	Pairwise Comparison	ROI	*t* Value	Cohen's *d*	*P* value
Nonmusicians	1	Anterior nonprimary Primary	15.91	5.03	6.75E-08
	2	Anterior nonprimary Lateral nonprimary	8.99	2.84	8.61E-06
	3	Posterior nonprimary Primary	7.25	2.29	4.81E-05
	4	Posterior nonprimary Lateral nonprimary	3.84	1.21	0.0040
Musicians	1	Anterior nonprimary Primary	11.40	3.61	1.19E-06
	2	Anterior nonprimary Lateral nonprimary	7.66	2.42	3.14E-05
	3	Posterior nonprimary Primary	4.49	1.42	0.0015
	4	Posterior nonprimary Lateral nonprimary	2.72	0.86	0.0237

Component weights were first averaged over hemispheres, and significance between ROI pairs was evaluated using paired t tests (n = 10). Note that because of our prior hypotheses and the significance of the omnibus F test, we did not correct for multiple comparisons. ROI, region of interest.

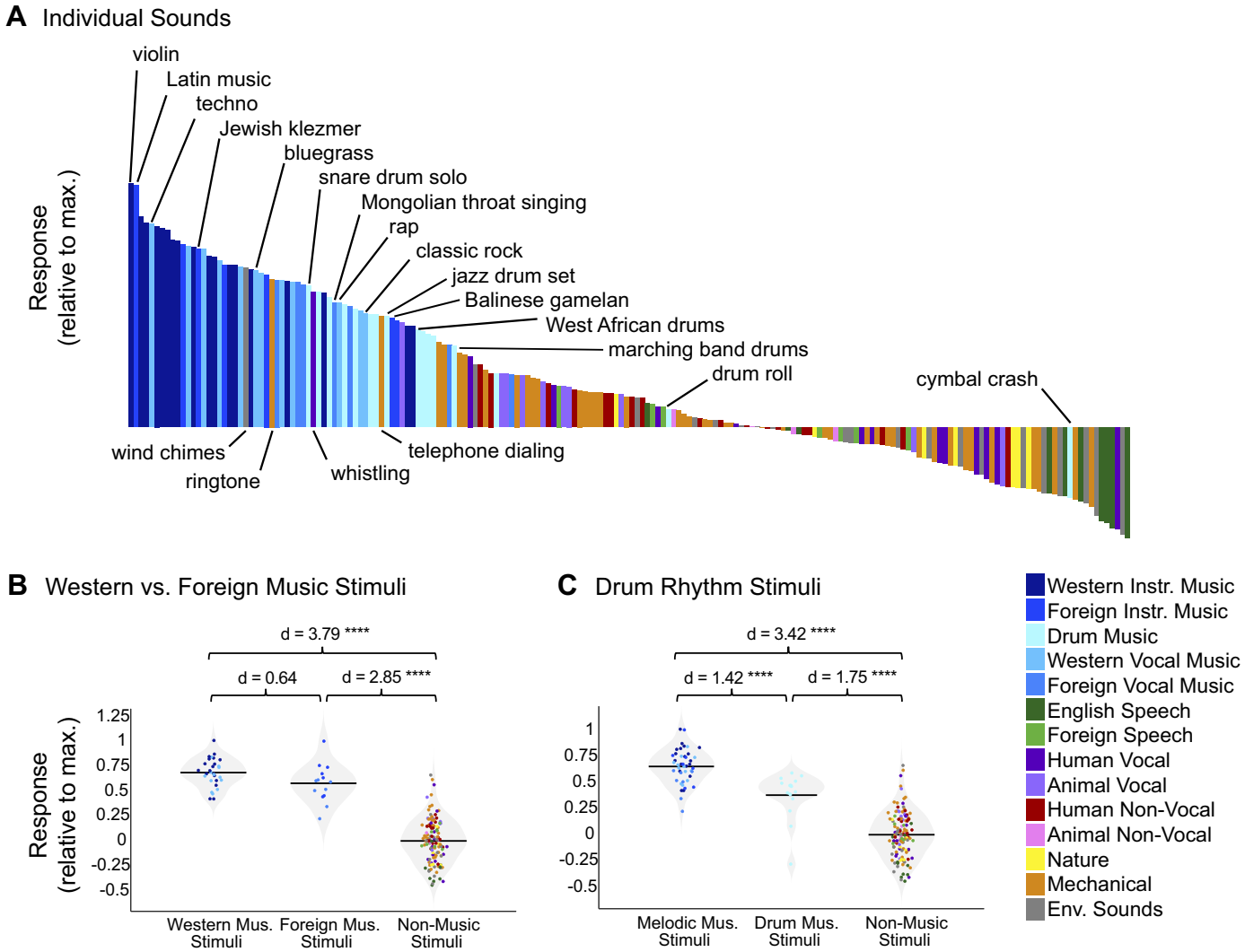


Figure 6. *A*: close-up of the response profile (192 sounds) for the music component inferred from all participants (n = 20), with example stimuli labeled. Note that there are a few "nonmusic" stimuli (categorized as such by Amazon Mechanical Turk raters) with high component rankings, but that these are all arguably musical in nature (e.g., wind chimes, ringtone). Conversely, "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*: distributions of Western music stimuli (n = 30), non-Western music stimuli (n = 14), and nonmusic stimuli (n = 132) within the music component response profile inferred from all 20 participants, with the mean for each stimulus group indicated by the horizontal black line. The separability between categories of stimuli (as measured using Cohen's d) is shown above the plot. Note that drum stimuli were left out of this analysis. *C*: distributions of melodic music stimuli (n = 44), drum rhythm stimuli (n = 16), and nonmusic stimuli (n = 132) within the music component response profile inferred from all 20 participants, with the mean for each stimulus group indicated by the horizontal black line. The separability between categories of stimuli (as measured using Cohen's d) is shown above the plot, and significance was evaluated using a nonparametric test permuting stimulus labels 10,000 times. ****Significant at $P < 0.0001$, two-tailed. Sounds are colored according to their semantic category.

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. Here, we analyze the music component inferred from all 20 participants since similar music components were inferred separately from musicians and nonmusicians (see Appendix and Fig. A7 for details of the components inferred from all 20 participants).

To expand beyond the original stimulus set from Ref. 7, 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; Fig. 6*A*). We expected that our American participants would have less exposure to these musical genres, allowing us to see whether the music component makes a distinction between familiar and less familiar music. The non-Western music stimuli were rated by American Mechanical Turk participants as being similarly musical (mean rating on 1–100 scale for Western music = 86.28, SD = 7.06; non-Western music mean = 79.63, SD = 9.01; P = 0.37, 10,000 permutations) but less familiar (mean rating on 1–100 scale for Western music = 66.50, SD = 8.23; non-Western music mean = 45.50, SD = 15.83; $P < 1.0e-5$, 10,000 permutations) 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 (Cohen's d = 0.64), a difference that was only marginally significant (Fig. 6*B*; P = 0.052, nonparametric test permuting music stimulus labels 10,000 times). Moreover, the magnitudes of both Western and non-Western music stimuli were both much higher than nonmusic stimuli (Western music stimuli vs. nonmusic stimuli: Cohen's d = 3.79, $P < 0.0001$, 10,000 permutations; non-Western music vs. nonmusic: Cohen's d = 2.85; $P <$ 0.0001, 10,000 permutations). Taken together, these results suggest that music-selective responses in auditory cortex occur even for relatively unfamiliar musical systems and genres.

Which stimulus features drive music selectivity? One of the most obvious distinctions is between melody and rhythm. Although 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 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. (7) study only included two drum stimuli, one of which was just a stationary snare drum roll that produced a low response in the music component, likely because it lacks both musical rhythm and pitch structure. The drum stimuli in our study ranked below the other instrumental and vocal music category responses (Cohen's d = 1.42, $P <$ 8.76e-07), but higher than the other nonmusic stimulus categories (Cohen's d = 1.75, $P <$ 9.60e-11; Fig. 6*C*). This finding suggests that the music component is not simply tuned to melodic information but is also sensitive to rhythm.

DISCUSSION

Our results show that cortical music selectivity is present in nonmusicians and hence does not require explicit musical training to develop. Indeed, the same six response components that characterized human auditory cortical responses to natural sounds in our previous study were replicated twice here, once in nonmusicians, and once in musicians. Our goal in this study was not to make statistical comparisons between nonmusicians and musicians (which would have required a prohibitive amount of data, see *Direct Group Comparisons of Music Selectivity* in the Appendix) but rather to assess whether the key properties of music selectivity were present in each group. Thus, although we cannot rule out the possibility that there are some differences between music-selective neural responses in musicians and nonmusicians, we have shown that in both groups, voxel decomposition produced a single music-selective component, which was selective for both instrumental and vocal music, and which was concentrated bilaterally in anterior and posterior superior temporal gyrus (STG). We also observed that the music-selective component responds strongly to both drums and less familiar non-Western music. Together, these results suggest that passive exposure to music is sufficient for the development of music selectivity in nonprimary auditory cortex, and that music-selective responses extend to rhythms with little melody, and to relatively unfamiliar musical genres.

Origins of Music Selectivity

Our finding of music-selective responses in nonmusicians is inconsistent with the hypothesis that explicit training is necessary for the emergence of music selectivity in auditory cortex and suggests rather that music selectivity is either present from birth or results from passive exposure to music. If present from birth, music selectivity could in principle represent an evolutionary adaptation for music, definitive evidence for which has long been elusive (80). But it is also plausible that music-specific representations emerge over development due to the behavioral importance of music in everyday life. For example, optimizing a neural network model to solve ecological speech and music tasks yields separate processing streams for the two tasks (81), suggesting that musical tasks sometimes require music-specific features. Another possibility is that music-specific features might emerge in humans or machines without tasks per se, due to the fact that music is acoustically distinct from other natural sounds. One way of testing this hypothesis might be to use generic unsupervised learning, for instance for producing efficient representations of sound (82–84), which might produce a set of features that are activated primarily by musical sounds.

Nearly all of our participants reported listening to music on a daily basis, and in other contexts, this everyday musical experience has clear effects (10–15), providing an example of how unsupervised learning from music might alter representations in the brain. Additionally, behavioral studies of nonindustrialized societies who lack much contact with Western culture show pronounced differences from Westerners in many aspects of music perception (85–88) and might plausibly also exhibit differences in the degree or nature of cortical music selectivity. Thus, our data do not show that music selectivity in the brain is independent of experience but rather that typical exposure to music in Western culture is sufficient for cortical music selectivity to emerge. It remains possible that the brains of people who grow up with less extensive musical exposure than our participants would not display such pronounced music selectivity.

What Does Cortical Music Selectivity Represent?

The music-selective component responds strongly to a wide range of music and weakly to virtually all other sounds, demonstrating that it is driven by a set of features that are relatively specific to music. One possibility is that there are simple acoustic features that differentiate music from other types of stimuli. Speech and music are known to differ in their temporal modulation spectra, peaking at 5 Hz and 2 Hz, respectively (89), and some theories suggest that these acoustic differences lead to neural specialization for speech vs. music in different cortical regions (90). However, standard auditory models based on spectrotemporal modulation do not capture the perception of speech and music (91) or neural responses selective for speech and music (55, 74, 81, 92). In particular, the music-selective component responds substantially less to sounds that have been synthesized to have the same spectrotemporal modulation statistics as natural music, suggesting that the music component does not simply represent the audio or modulation frequencies that are prevalent in music (55).

Our finding that the music-selective component shows high responses to less familiar musical genres places some constraints on what these properties might be, as does the short duration of the stimuli used to characterize music selectivity. For instance, the music-specific features that drive the response are unlikely to be specific to Western music and must unfold over relatively short timescales. Features that are common to nearly all music, but not other types of sounds, include stable and sustained pitch organized into discrete note-like elements, and temporal patterning with regular time intervals. Because the music component anatomically overlaps with more general responses to pitch (7), it is natural to wonder if it represents higher-order aspects of pitch, such as the previously mentioned stability, or discrete jumps from one note to another. However, the high response to drum rhythms in the music component that we observed here indicates that the component is not only sensitive to pitch structure. Instead, this result suggests that melody and rhythm might be jointly analyzed, rather than dissociated, at least at the level of auditory cortex. One possibility is that the underlying neural circuits extract temporally local representations of melody and rhythm motifs that are assembled elsewhere into the representations of contour, key, meter, groove etc. that are the basis of music cognition (3, 93–96).

Limitations

Our paradigm used relatively brief stimuli since music-selective regions are present just outside of primary auditory cortex, where integration periods appear to be short (74, 97). And we intentionally used a simple task (intensity discrimination) to encourage subjects to attend to all stimuli. But because the responses of auditory cortical neurons have been known to change based on task demands (e.g., 98), it is possible that more complex stimuli or tasks would reveal additional aspects of music-selective responses, which might not be present to a similar degree in nonmusicians. One relevant point of comparison is the finding that amusic participants with striking pitch perception deficits show pitch-selective auditory cortical responses that are indistinguishable from those of control participants with univariate analyses (99). Recent evidence suggests it is nonetheless possible to discriminate amusic participants from controls and to predict participants' behavioral performance, using fMRI data collected in the context of a pitch task (100). Utilizing a music-related task might produce larger differences between musicians and nonmusicians, as might longer music stimuli (compared with the 2-s clips used in this experiment), which could be argued to contain richer melodic, harmonic, and/or rhythmic information.

Finally, our study is limited by the resolution of fMRI. Voxel decomposition is intended to help overcome the spatial limitations of fMRI, and indeed appears to reveal responses that are not evident in raw voxel responses but can be seen with finer-grained measurement substrates such as electrocorticography (6). But the spatial and temporal resolution of the BOLD signal inevitably constrain what is detectable and place limits on the precision with which we can observe the activity of music-selective neural populations. Music-selective brain responses might well exhibit additional characteristics that would only be evident in fine-grained spatial and temporal response patterns that cannot be resolved with fMRI. Thus, we cannot rule out the possibility that there are additional aspects of music-selective neural responses that might be detectable with other neuroimaging methods (e.g., M/EEG, ECoG) and which are absent or altered in nonmusicians.

Future Directions

One of the most interesting open questions raised by our findings is whether cortical music selectivity reflects implicit knowledge gained through typical exposure to music, or whether it is present from birth. These hypotheses could be addressed by testing people with very different musical experiences from non-Western cultures or other populations whose lifetime perceptual experience with music is limited in some way (e.g., people with musical anhedonia, children of deaf adults). It would also be informative to test whether music selectivity is present in infants or young children. Finally, much remains to be learned about the nature of cortical music selectivity, such as what acoustic or musical features might be driving it. The voxel decomposition approach provides one way of answering these questions and exploring the quintessentially human ability for music.

APPENDIX

Psychoacoustic Data Acquisition and 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 nonmusicians. For all psychoacoustic tasks, stimuli were presented using Psychtoolbox for Matlab (101). Sounds were presented to participants at 70dB SPL over circumaural Sennheiser HD280 headphones in a soundproof booth (Industrial Acoustics; SPL level was computed without any weighting across time or frequency). After each trial, participants were given feedback about whether or not they had answered correctly. Group differences for each task were measured using nonparametric Wilcoxon rank sum tests.

Pure tone frequency discrimination.

Because musicians have superior frequency discrimination abilities when compared with nonmusicians (102–104), 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, whereas 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 (105), which targets participants' 79.4% threshold. The frequency difference was changed initially by a factor of two, which was reduced to a factor of $\sqrt{2}$ after the fourth reversal. Once 10 reversals had been measured, participants' thresholds were estimated as the average of these 10 values.

Multiple threshold estimations were obtained per participant (three threshold estimations for the first seven participants, and five 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 (106–108). Participants were asked to tap along with an isochronous click track. They heard ten 30-s click blocks, separated by 5 s of silence. The blocks varied widely in tempo, with interstimulus intervals ranging from 200 ms to 1 s (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 used in previous studies (85, 109). We then calculated the difference between participants' response onsets and the actual stimulus onsets. The standard deviation of these asynchronies between corresponding stimulus and response onsets was used as a measure of sensorimotor synchronization ability (109).

Melody discrimination.

Musicians have also been reported to outperform nonmusicians on measures of melodic contour and interval discrimination (44, 110, 111). 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 150 Hz to 270 Hz. 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, whereas 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 × same/different contour × 1/2 semitone change), for a total of 160 trials. This task was modified from McPherson and McDermott (111).

"Sour note" detection.

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 (112). 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 increased 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 × 3 scale degrees), for a total of 120 trials. This task was modified from McPherson and McDermott (111).

Psychoacoustic Results

As predicted, musicians outperformed nonmusicians on all behavioral psychoacoustic tasks, replicating prior findings (Fig. A1). Consistent with previous reports (102–104), musicians performed slightly better on the frequency discrimination task (median discrimination threshold = 0.51%, SD = 0.12%) than nonmusicians (median discrimination threshold = 0.57%, SD = 0.23%); this difference was marginally

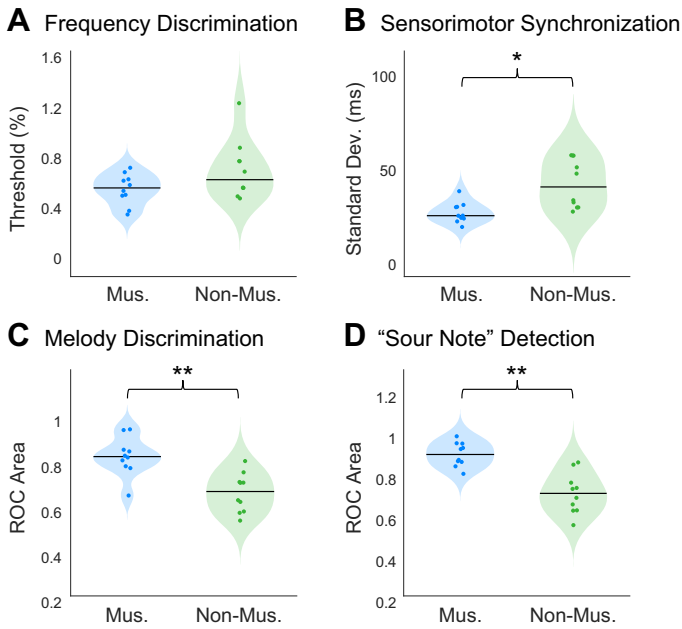


Figure A1. Musicians (n = 10) outperform nonmusicians (n = 10) on psychoacoustic tasks. *A*: participants' pure tone frequency discrimination thresholds were measured using a 1-up 3-down adaptive two-alternative forced choice (2AFC) task, in which participants indicated which of two pairs of tones were 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 (112) and instructing participants to determine whether or not the melody contained an out-of-key ("sour") note. Colored dots represent individual participants, and the median for each participant group is indicated by the horizontal black line. Mus., musicians; Non-Mus., nonmusicians. *Significant at $P < 0.01$ one-tailed, **Significant at $P < 0.001$ one-tailed.

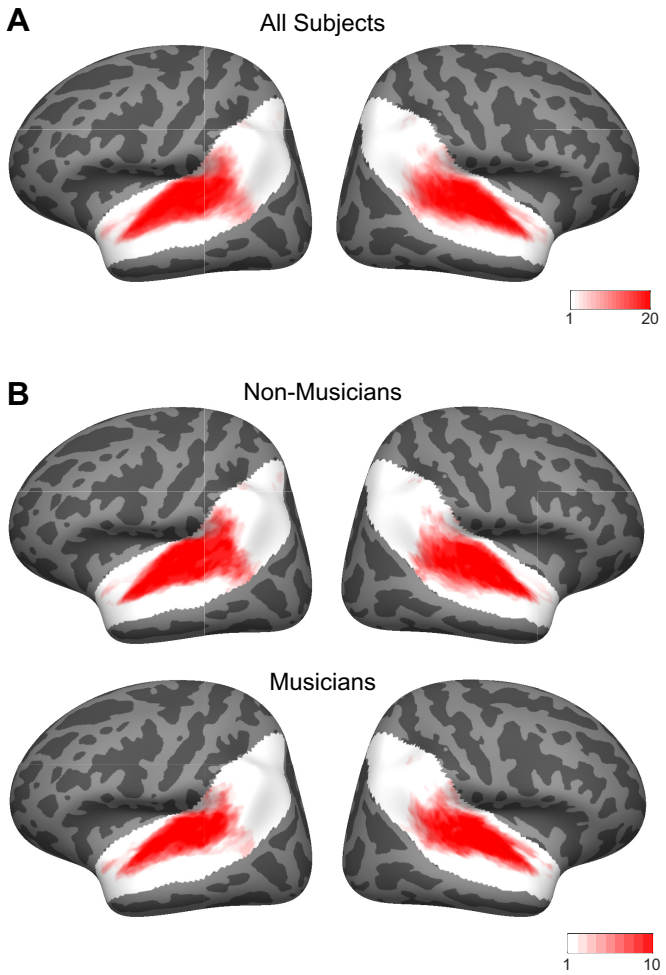


Figure A2. Subject overlap maps showing which voxels were selected in individual subjects to serve as input to the voxel decomposition algorithm. The white area shows the anatomical constraint regions from which voxels were selected. *A*: overlap map for all 20 subjects. *B*: overlap maps for nonmusicians (n = 10) and musicians (n = 10) separately, illustrating that the anatomical location of the selected voxels was largely similar across groups.

significant (Z = −1.32, P = 0.09, effect size r = −0.30, one-tailed Wilcoxon rank sum test, Fig. A1*A*). Musicians were also better able to synchronize their finger tapping with an isochronous beat, showing significantly less variability in their response (SD = 22.4 ms) than nonmusicians (SD = 37.7 ms, Z = −2.68, P < 0.01, effect size r = −0.60, one-tailed Wilcoxon rank sum test, Fig. A1*B*). When presented with musical melodies, musicians were better able to discriminate between two similar melodies (musician median ROC area = 0.82, SD = 0.08, nonmusician mean ROC area = 0.66, SD = 0.09, Z = 3.21, P < 0.001, effect size r = 0.72, one-tailed Wilcoxon rank sum test, Fig. A1*C*) and to detect scale violations within melodies (musician median ROC area = 0.89, SD = 0.06, nonmusicians median ROC area = 0.70, SD = 0.10, Z = 3.44, P < 0.001, effect size r = 0.77, one-tailed Wilcoxon rank sum test, Fig. A1*D*). These behavioral effects validate our participants' self-reported status as trained musicians or nonmusicians.

Details of Voxel Selection

To be included as input to the decomposition algorithm, a voxel must display a significant (P < 0.001, uncorrected) response to sound (pooling over all sounds compared with silence) and produce a reliable response pattern to the stimuli across scanning sessions (see equations in MATERIALS AND METHODS section). For the main analyses, voxels were selected from within a large anatomical constraint region. Voxels selected in individual participants according to these criteria can be seen in Fig. A2. To see whether the anatomical constraint was missing a substantial number of reliably sound-responsive voxels, we also selected voxels without the anatomical constraint (Fig. A3), and the resulting components were very similar (Fig. A4).

Details of Voxel Decomposition

Like ICA, the voxel decomposition method (7) searches among the many possible solutions to the factorization problem for components that have a maximally non-Gaussian distribution of weights across voxels. The voxel weights of

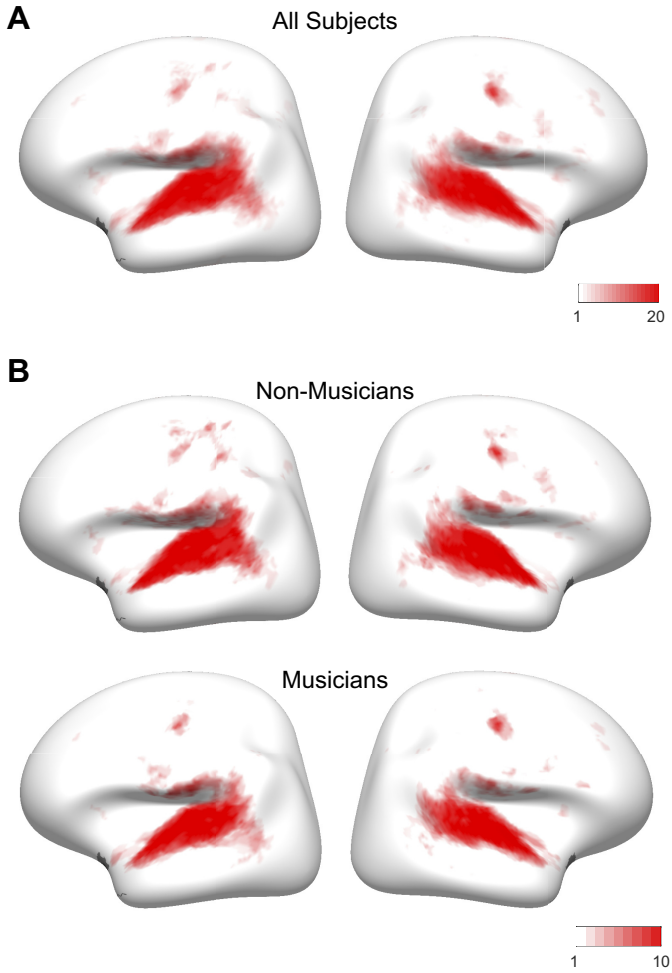


Figure A3. Subject overlap maps showing which voxels pass the selection criteria as described in Fig. A2, but without any anatomical mask applied before selecting voxels.

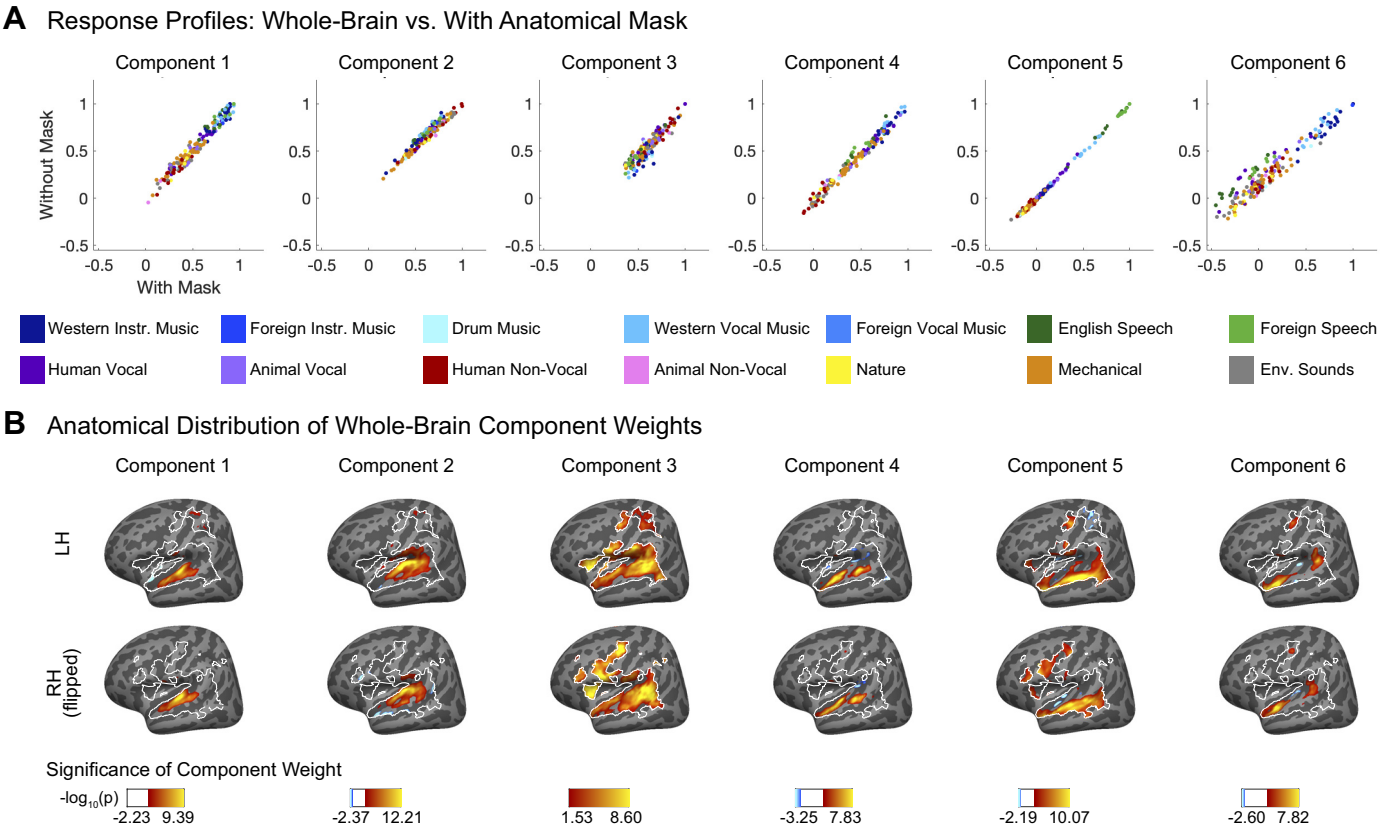


Figure A4. Similarity between components with anatomical mask vs. whole-brain. *A*: scatter plots showing the components inferred from all 20 participants, using the voxel decomposition algorithm both with and without the anatomical mask shown in Fig. A2. Individual sounds are colored according to their semantic category. *B*: spatial distribution of whole brain component voxel weights, computed using a random effects analysis of participants' individual component weights. Weights are compared against 0; *P* values are logarithmically transformed ($-\log_{10}[P]$). The white outline indicates the voxels that were both sound-responsive (sound vs. silence, $P < 0.001$ uncorrected) and split-half reliable ($r > 0.3$) at the group level. 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 right hemisphere (*bottom row*) is flipped to make it easier to visually compare weight distributions across hemispheres. FDR, false discovery rate.

the inferred components were indeed more skewed and kurtotic than would be expected from a Gaussian distribution (Fig. A5).

As explained in the MATERIALS AND METHODS, the only free parameter in the voxel decomposition analysis is the number of components recovered. To determine the optimal number of components, we measured the fraction of the reliable response variance explained by a given number of components and chose the number after which the explained variance plateaued (Fig. A6).

Replication of Norman-Haignere et al. Using Data from All 20 Participants

In addition to conducting the voxel decomposition analysis separately for musicians and nonmusicians, we were able replicate the full results from Ref. 7 using data from all 20 participants from both groups. Here, we describe that analysis in more detail and explain more about the four components that are selective for acoustic stimulus features.

In our previous study (7), as in the analyses described for the current study, prior to applying the voxel decomposition algorithm, each participant's responses were demeaned across voxels (see MATERIALS AND METHODS), 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 nonreplicable 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 removing differences between musicians and nonmusicians that might be of interest, we ran the voxel decomposition algorithm without demeaning by individual participants. When we varied the number of components as we normally do to determine the number of components to use for ICA, we found that the best results were obtained with eight components, which included the expected set of six plus two "extra" components that emerged as a result of the omitted normalization step. If we include the normalization step, we found the expected set of six components.

Of the eight components derived from the nondemeaned data, six of them were each very similar to one of the six components from Ref. 7 and accounted for 87.54% of voxel response variance. Because the components inferred using

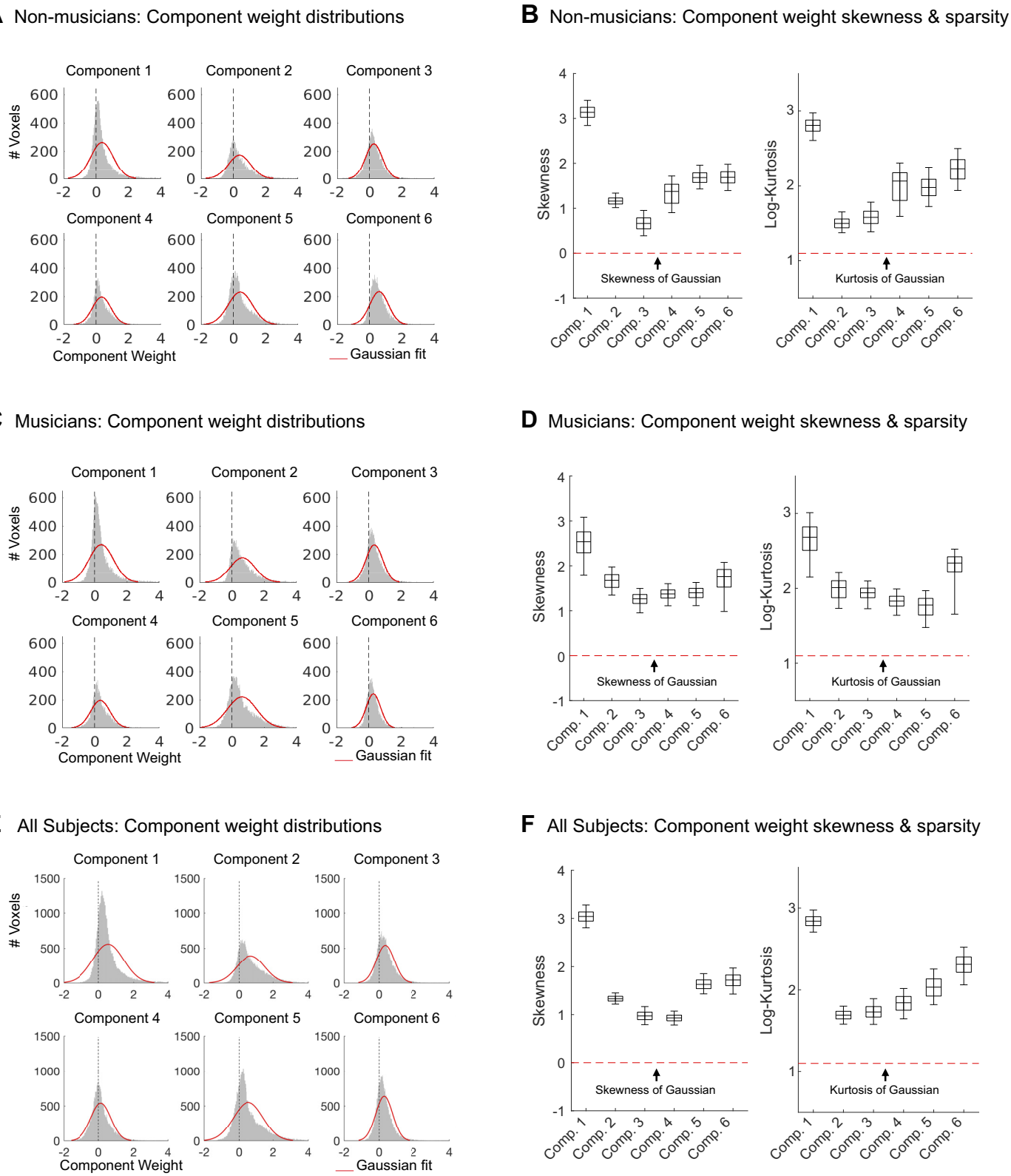


Figure A5. *A*: histograms showing the weight distributions for each component inferred from nonmusicians (n = 10), along with their Gaussian fits (red). *B*: skewness and log-kurtosis (a measure of sparsity) for each component inferred from nonmusicians (n = 10), illustrating that the inferred components are skewed and sparse compared with 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*). *C* and *D*: same as *A* and *B*, but for the components inferred from musicians (n = 10). *E* and *F*: same as *A* and *B*, but for the components inferred from all 20 participants. Comp, component.

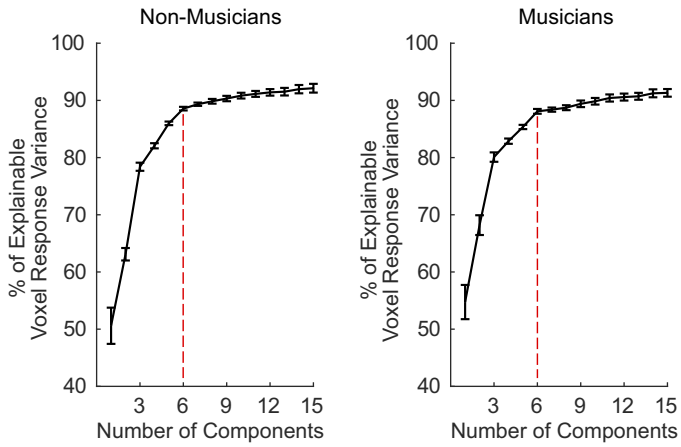


Figure A6. The proportion of voxel response variance explained by different numbers of components, for both nonmusicians (n = 10, *left*) and musicians (n = 10, *right*). The figure plots the median variance explained across voxels (noise corrected by split-half reliability using the Spearman correction for attenuation; 121), calculated separately for each subject and then averaged across the 10 subjects in each group. Error bars plot one standard error of the mean across subjects. For both groups, six components were sufficient to explain over 88% of the noise-corrected variance.

ICA have no order, we first used the Hungarian algorithm (65) to optimally reorder the components, maximizing their correlation with the components from our previous study. As expected, the reordered components were highly correlated, with r values ranging from 0.76 to 0.98 (Fig. A7*A*; see Fig. A7*B* for the response profiles of the components, and Fig. A7*C* for the profiles averaged within sound categories). To confirm that these strong correlations are not simply an artifact of the Hungarian algorithm matching procedure, we ran a permutation test in which we reordered the sounds within each component 1,000 times, each time using the Hungarian algorithm to match these permuted components with those from our previous study. The resulting correlations between the original components their corresponding permuted components were very low (mean r values ranging from 0.086 to 0.098), with the maximum correlation over all 1,000 permutations not exceeding r = 0.3 for any component.

The additional two components were much less correlated with any of the six original components, with the strongest correlation being r = 0.28. As expected, the weights of these additional two components were concentrated in a small number of participants (one almost entirely loading onto a single nonmusician participant, and the other onto a small group composed of both musicians and nonmusicians). 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 (Fig. A7, *B*–*H*). The non-Gaussianity of these six components can be seen in Fig. A5*A* (skewness ranging from 1.06 to 2.96, log-kurtosis ranging from 1.70 to 2.79).

As in Ref. 7, four of the components were selective for different acoustic properties of sound (Fig. A7, *D* and *E*, Fig. A8), whereas two components were selective for speech (*component 5*) and music respectively (*component 6*; Fig. A7, *B* and *C*, Fig. A8). 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*; Fig. A7*D*). The group anatomical weights for *components 1* and *2* concentrated in the low- and high-frequency regions of primary auditory cortex (Fig. A7, *F* and *H*) (67, 70, 113, 114). We did not measure tonotopy in the individual participants from this study, but our previous study did so and found a close correspondence between individual participant tonotopic maps and the weights for these two components. Components also showed tuning to spectrotemporal modulations (Fig. A7*E*), with a tradeoff between selectivity for fine spectral and slow temporal modulation (*components 1* and *4*) versus coarse spectral and fast temporal modulation (*components 2* and *3*) (115, 116). *Component 4*, which exhibited selectivity for fine spectral modulation, was concentrated anterior to Heschl's gyrus (*component 4*, Fig. A7, *F* and *H*), similar to prior work that has identified tone-selective regions in anterolateral auditory cortex in humans (117–119). Conversely, selectivity for coarse spectral modulation and fast temporal modulation was concentrated in posterior regions of auditory cortex (*component 3*, Fig. A7, *F* and *H*) (71), consistent with previous studies reporting selectivity for sound onsets in caudal areas of human auditory cortex (120).

The two remaining components responded selectively to speech and music, respectively (*component 5* and *6*, Fig. A7*C*) and were not well accounted for using acoustic properties alone (Fig. A8). The weights for the speech-selective component (*component 5*) were concentrated in the middle portion of the superior temporal gyrus (midSTG, Fig. A7, *F* and *H*), as expected (73–75). In contrast, the weights for the music-selective component (*component 6*) were most prominent anterior to PAC in the planum polare, with a secondary cluster posterior to PAC in the planum temporale (Fig. A7, *F* and *H*) (3, 7, 41, 48, 52, 76, 77).

These results closely replicate the functional organization of human auditory cortex reported by Norman-Haignere et al. (7), including the existence and anatomical location of inferred music-selective neural populations.

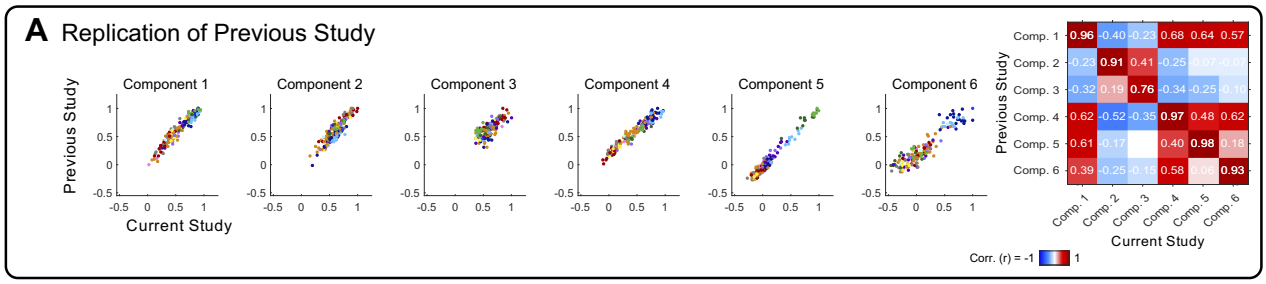
Component Voxel Weights within 15 Anatomical ROIs

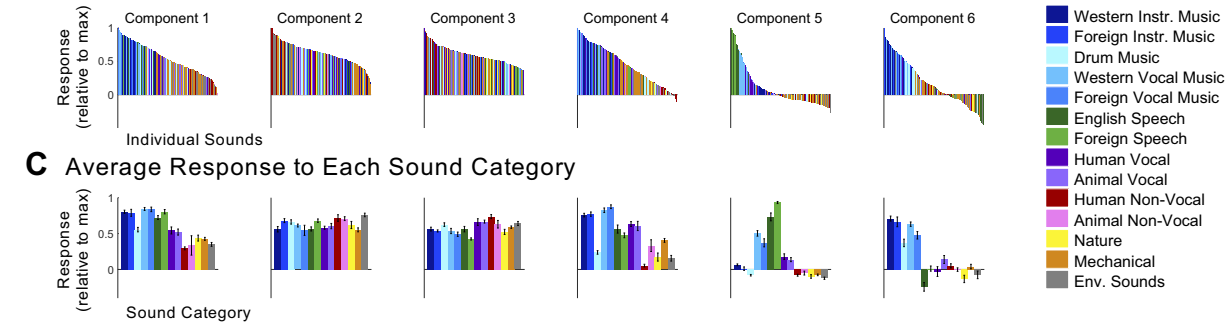
In addition to visualizing component voxel weights as cortical maps depicting the results of a group random effects analysis (e.g., Fig. 3), we measured individual participants' mean voxel weights within a set of 15 standardized anatomical parcels from (84), chosen to fully encompass the superior temporal plane and superior temporal gyrus (STG; Fig. A9). Combinations of subsets of these parcels were used in the analysis described in Fig. 5.

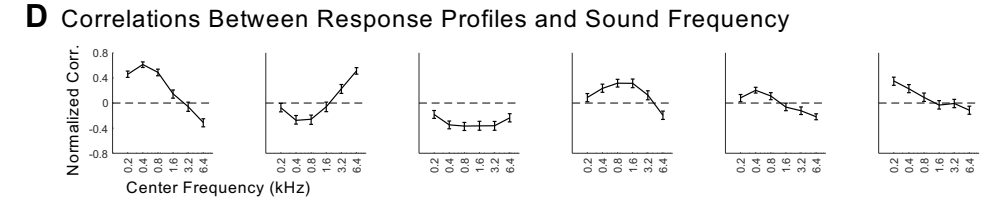
Direct Group Comparisons of Music Selectivity

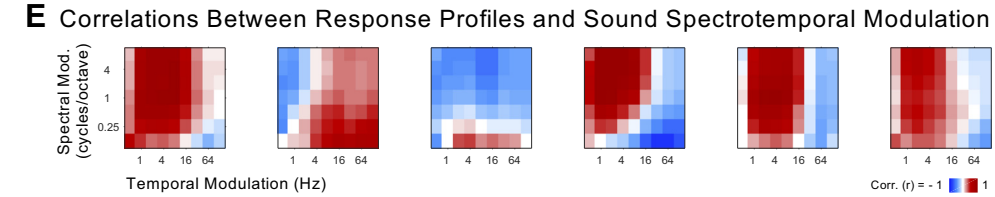
Music component selectivity.

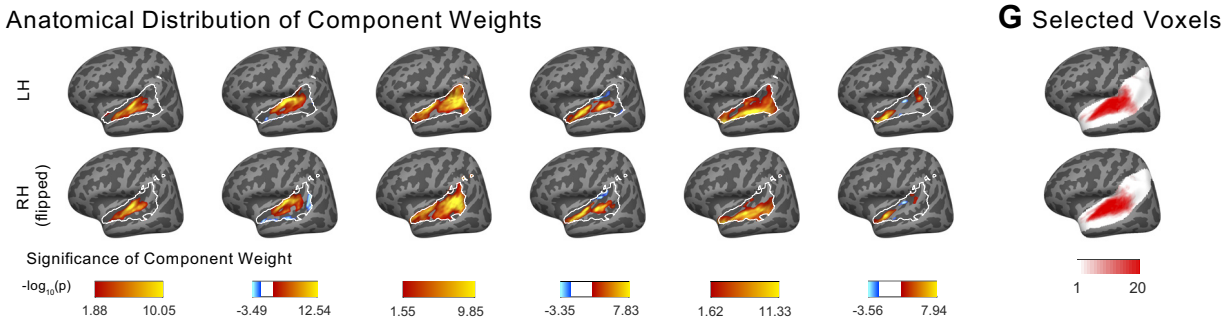
One natural extension of the analyses presented in this paper is to directly compare music selectivity in our group of nonmusicians to that observed in our group of musicians. To compare the selectivity of the music components

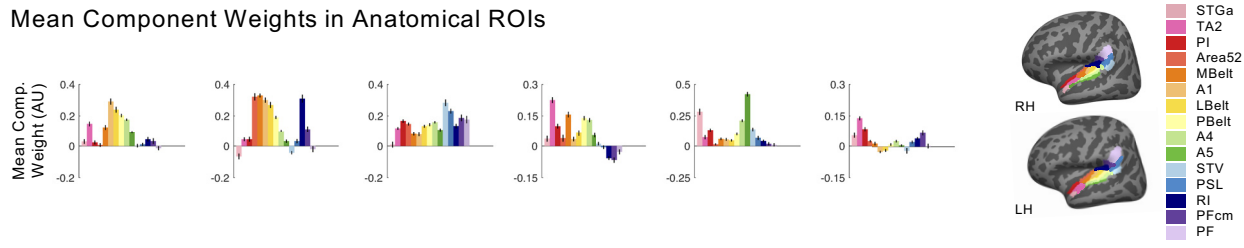












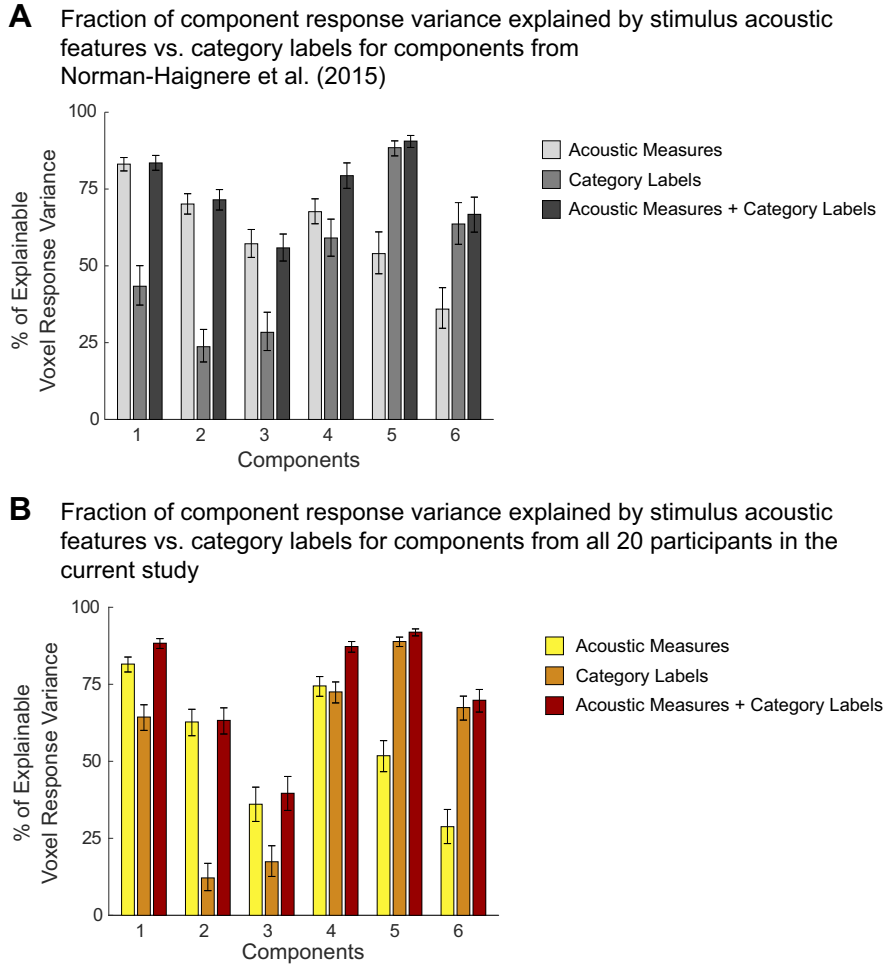


Figure A8. Total amount of component response variation explained by *1*) all acoustic measures (frequency content and spectrotemporal modulation energy), *2*) all category labels (as assigned by Amazon Mechanical Turk workers), and *3*) the combination of acoustic measures and category labels. *A*: results for components from our previous study (n = 10; 7). For *components 1–4*, category labels expla-ined little additional variance beyond that explained by acoustic features. For *components 5* (speech-selective) and *6* (music-selective), category labels explained most of the response variance, and acoustic features accounted for little additional variance. *B*: same as *A* but for the components inferred from all 20 participants in the current study.

inferred from each group, we computed Cohen's d 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 nonmusic stimuli (all other sound categories).

The significance of the observed group difference was evaluated using a nonparametric test in which we permuted participant groupings (i.e., randomly assigning 10 participants to *group A*, and the remaining 10 to *group B*) and inferred a set of components for each permuted group. We then calculated Cohen's d for each group's music component and computed the absolute value of the difference between these two values. We did this 1,000 times to build up a null distribution of Cohen's d differences and then compared this with the observed difference in Cohen's d for nonmusicians vs. musicians and found the observed difference to be nonsignificant (P = 0.21, one-tailed).

Figure A7. Independent components inferred from voxel decomposition of auditory cortex of all 20 participants (as compared with the components in Figs. 2–5, which were inferred from musicians and nonmusicians separately). Additional plots are included here to show the extent of the replication of the results of Ref. 7. *A*: scatterplots showing the correspondence between the components from our previous study (n = 10; y-axis) and those from the current study (n = 20; x-axis). 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 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. Error bars plot one standard error of the mean across sounds from a category, computed using bootstrapping (10,000 samples). *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. Weights are compared against 0; P values are logarithmically transformed ($-\log_{10}[P]$). The white outline indicates the 2,249 voxels that were both sound-responsive (sound vs. silence, $P < 0.001$ uncorrected) and split-half reliable ($r > 0.3$) at the group level. 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 right hemisphere (*bottom row*) is flipped to make it easier to visually compare weight distributions across hemispheres. *G*: subject overlap maps showing which voxels were selected in individual subjects to serve as input to the voxel decomposition algorithm (same as Fig. A2*A*). To be selected, a voxel must display a significant ($P < 0.001$, uncorrected) response to sound (pooling over all sounds compared to silence) and produce a reliable response pattern to the stimuli across scanning sessions (see equations in MATERIALS AND METHODS section). The white area shows the anatomical constraint regions from which voxels were selected. *H*: mean component voxel weights within standardized anatomical parcels from Ref. 84, chosen to fully encompass the superior temporal plane and superior temporal gyrus (STG). Error bars plot one standard error of the mean across participants. LH, left hemisphere; RH, right hemisphere; ROI, region of interest.

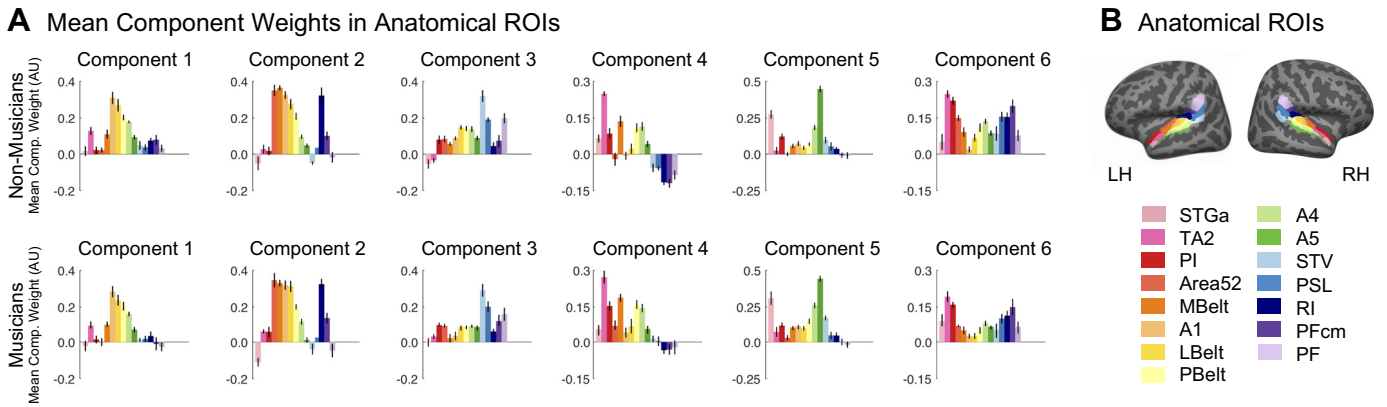


Figure A9. *A*: mean component weight in a set of 15 anatomical parcels from Glasser et al. (64), plotted separately for nonmusicians (n = 10; *top*) and musicians (n = 10; *bottom*). Error bars plot one standard error of the mean across participants. *B*: selected anatomical parcels from Glasser et al. (64), chosen to fully encompass the superior temporal plane and superior temporal gyrus (STG). LH, left hemisphere; RH, right hemisphere; ROI, region of interest.

It was not possible to conduct a power analysis using the data from Ref. 7 to determine how large of a difference we are powered to detect with our sample size, because the component analysis is sometimes unstable when the number of unique participants drops below 10, which made it impossible to bootstrap and infer a large number of sets of components. For that reason, we cannot rule out the possibility that a small difference does exist, but that this test is underpowered.

Music component weight magnitude and power analysis.

We thought it would be interesting to directly compare the magnitude of the weights of the music-selective component between expert musicians and nonmusicians. So that individual participants' component weight magnitudes could be compared in a meaningful way, we planned to run the voxel decomposition analysis on the data from all 20 participants (as reported in the APPENDIX section titled *Replication of Norman-Haignere et al. Using Data from All 20 Participants*) and directly compare the weights of musicians vs. nonmusicians for the resulting music component. Although there are many ways to summarize a participant's component magnitude, we used the median weight over their voxels with the top 10% of component weights (using independent data to select the voxels vs. quantify selectivity), though results were robust to the fraction of voxels selected (i.e., measuring participants' median weight over the top 5%, 7.5%, 10%, 15%, and 20% of voxels led to similar

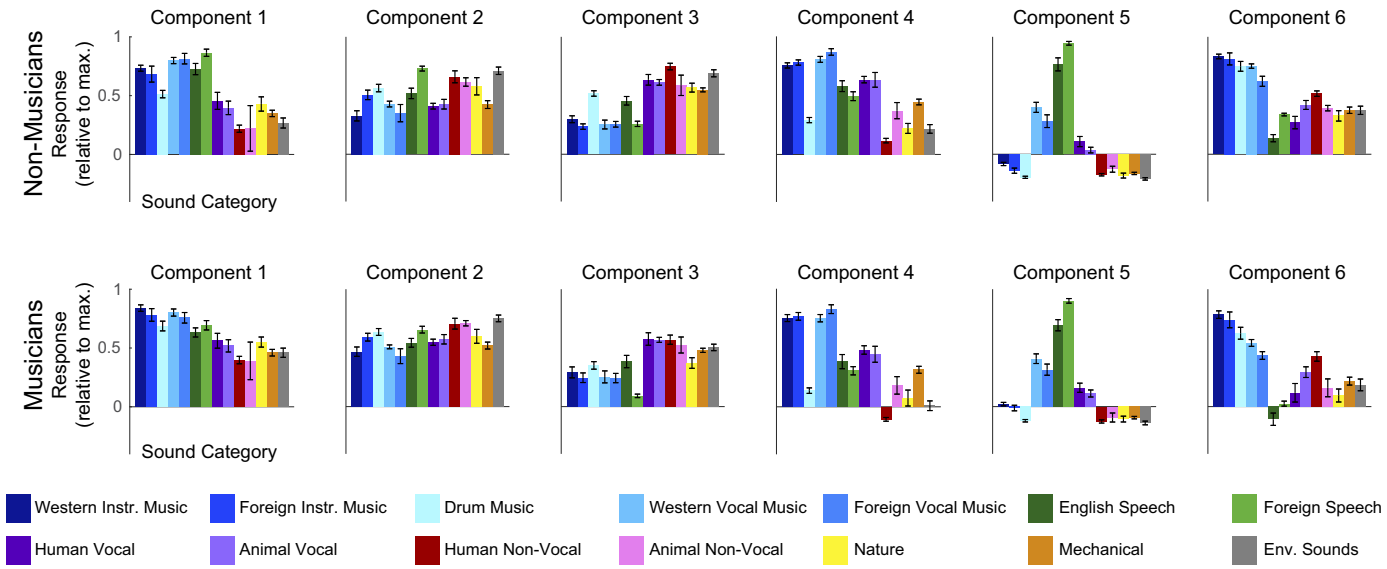


Figure A10. Response profiles discovered using the probabilistic parametric method, separately for nonmusicians (n = 10; *top*) and musicians (n = 10; *bottom*). These components were very highly correlated with those inferred using the nonparametric ICA-based voxel decomposition method presented in the main text, with the main difference between the two methods being the mean response profile magnitude (i.e., the "offset" from baseline). Because this mean response varies depending on the details of the analysis used to infer the components, while the components themselves remain highly similar, we chose to quantify selectivity using a measure (Cohen's d) that does not take the baseline into account but rather quantifies the separation between stimulus categories within the response profile. ICA, independent components analysis.

results). This decision to select a subset of voxels was made because music selectivity is typically sparse and limited to a small fraction of voxels, and we thought it reasonable to expect the largest group difference in the regions of auditory cortex with the highest music component weights.

To get a sense for how large a group difference we would be able to reliably detect given our sample size, we conducted a power analysis using the data from Ref. 7. We compared the music component weights for the participants in that study (n = 10) with a second population of 10 participants created by sampling participants with replacement and then shifting their component weights by various amounts (ranging from 0% to 100% in increments of 5%), representing various models for how the music component weights might change in musicians. The difference between the groups' median weights was computed, and the significance of this group difference was assessed by permuting participant groupings 1,000 times. For each shift amount, we repeated this entire procedure 1,000 times, each time sampling a new set of 10 participants for each group. The probability of detecting a significant group difference for each shift amount was recorded, and the results showed that we were able to detect a significant group difference 80% of the time only when the two groups' median weights differed by 47%.

This power analysis suggests that with our sample size, we are only able to detect a relatively large difference in music component weight magnitude between the groups. With this in mind, we performed this analysis and found the strength of the music component was slightly higher in musicians compared with nonmusicians, but this difference did not reach significance (P = 0.11, two-tailed nonparametric test permuting subject groupings 10,000 times). A Bayesian independent-sample t test was inconclusive [$t(18)$ = 1.68, P = 0.11, BF_{10} = 1.03; prior on the effect size following a Cauchy distribution with $r = \sqrt{2}/2$], which suggests that the data are equally likely under the null hypothesis that groups do not differ in their music component weight magnitude vs. the alternative hypothesis that they do. Together, these results do not rule out differences between the strength of music selectivity in individuals at the two extremes of musical training, but they suggest that any such differences are not large.

As previously explained, we thought that restricting this analysis to the most music-selective voxels (i.e., the voxels with the highest music component weights) would be most likely to detect any group difference in weight magnitude. However, we also tried a simpler analysis in which we ran an ROI $\times$ hemisphere repeated-measures ANOVA on participants' weights for the component inferred from all 20 participants, and including "group" as a between-subjects factor. When we do this, we still find a significant main effect of ROI [$F(3,54)$ = 37.89, P = 2.56e-13], but no significant main effect of hemisphere [$F(1,18)$ = 0.84, P = 0.37], as was found in the corresponding analyses within each group that are reported in the main text. However, we also found no significant main effect of group [$F(1,18)$ = 2.81, P = 0.11] or any significant two-way or three-way interactions with group (all Ps > 0.05). Moreover, a Bayesian version of this analysis provides no evidence either for or against an effect of group (BF_{inc} = 1.11). This result is consistent with the power analysis, indicating that although we find no evidence of a difference between musicians' and nonmusicians' music component weights, we do not have enough statistical power to rule out the possibility that a small difference does exist.

Parametric Matrix Factorization Method

In addition to the nonparametric matrix factorization method reported throughout this paper, we repeated our analyses using a probabilistic parametric algorithm also developed and reported in Ref. 7, which did not constrain voxel weights to be uncorrelated. This parametric model assumed a skewed and sparse non-Gaussian prior (the Gamma distribution) on the distribution of voxel weights, which constrained them to be positive (unlike the nonparametric method). Because the components discovered using the nonparametric method showed different degrees of skewness and sparsity, the exact shape of the Gamma distribution prior was allowed to vary between components in this parametric analysis. Components were discovered by searching for response profiles and shape parameters that maximized the likelihood of the data, integrating across all possible voxel weights.

Due to the stochastic nature of the model optimization procedure (see 7 for details), the optimization procedure was repeated 25 times each for musicians and nonmusicians. For each group, we chose the set of component response profiles with the highest estimated log-likelihood (though all 25 iterations produced very similar results) and used the Hungarian algorithm to match them with the components inferred using the nonparametric method. The sets of components inferred with these two different methods were very highly correlated, with r values ranging from 0.83 to 0.998 for nonmusicians, and from 0.90 to 0.99 for musicians. However, the components did differ somewhat in the mean magnitude of the response profiles (see Fig. A10), plausibly due to the positivity constraint on the component voxel weights.

ACKNOWLEDGMENTS

We thank the McDermott lab for useful comments on an earlier version of this manuscript.

GRANTS

This work was supported by National Science Foundation Grant BCS-1634050 to J. McDermott and NIH grant DP1HD091947 to N. Kanwisher. S. Norman-Haignere was supported by a Life Sciences Research Fellowship from the Howard Hughes Medical Institute (HHMI). The Athinoula A. Martinos Imaging Center at Massachusetts Institute of Technology is supported by the NIH Shared Instrumentation Grant S10OD021569.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

D.B., S.N-H., J.M., and N.K. conceived and designed research; D.B. performed experiments; D.B. analyzed data; D.B., S.N-H., J.M., and N.K. interpreted results of experiments; D.B. prepared figures; D.B. drafted manuscript; D.B., S.N-H., J.M., and N.K. edited and revised manuscript; D.B., S.N-H., J.M., and N.K. approved final version of manuscript.

ENDNOTE

At the request of the authors, readers are herein alerted to the fact that additional materials related to this manuscript may be found at https://github.com/snormanhaignere/nonparametric-ica. These materials are not a part of this manuscript and have not undergone peer review by the American Physiological Society (APS). APS and the journal editors take no responsibility for these materials, for the website address, or for any links to or from it.

REFERENCES

1. **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.** Universality and diversity in human song. *Science* 366: eaax0868, 2019. doi:10.1126/science.aax0868.
2. **Trehub SE.** The developmental origins of musicality. *Nat Neurosci* 6: 669–673, 2003. doi:10.1038/nn1084.
3. **Fedorenko E**, **McDermott JH**, **Norman-Haignere SV**, **Kanwisher NG.** Sensitivity to musical structure in the human brain. *J Neurophysiol* 108: 3289–3300, 2012. doi:10.1152/jn.00209.2012.
4. **LaCroix AN**, **Diaz AF**, **Rogalsky C.** The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Front Psychol* 6: 1–19, 2015. doi:10.3389/fpsyg.2015.01138.
5. **Leaver AM**, **Rauschecker JP.** Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30: 7604–7612, 2010. doi:10.1523/JNEUROSCI.0296-10.2010.
6. **Norman-Haignere SV**, **Feather J**, **Boebinger D**, **Brunner P**, **Ritaccio A**, **McDermott JH**, **Schalk G**, **Kanwisher N.** Intracranial recordings from human auditory cortex reveal a neural population selective for musical song (Preprint). *bioRxiv* 696161, 2020. doi:10.1101/696161.
7. **Norman-Haignere SV**, **Kanwisher NG**, **McDermott JH.** Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88: 1281–1296, 2015. doi:10.1016/j.neuron.2015.11.035.
8. **Rogalsky C**, **Rong F**, **Saberi K**, **Hickok G.** Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J Neurosci* 31: 3843–3852, 2011. doi:10.1523/JNEUROSCI.4515-10.2011.
9. **Tierney A**, **Dick F**, **Deutsch D**, **Sereno M.** Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex* 23: 249–254, 2013. doi:10.1093/cercor/bhs003.
10. **Bigand E.** Perceiving musical stability: the effect of tonal structure, rhythm, and musical expertise. *J Exp Psychol* 23: 808–822, 1983. doi:10.1037/0096-1523.23.3.808.
11. **Bigand E**, **Pineau M.** Global context effects on musical expectancy. *Percept Psychophys* 59: 1098–1107, 1997. doi:10.3758/bf03205524.
12. **Bigand E**, **Poulin-Charronnat B.** Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100: 100–130, 2006. doi:10.1016/j.cognition.2005.11.007.
13. **Koelsch S**, **Gunter T**, **Friederici AD**, **Èger ES.** Brain indices of music processing: nonmusicians'' are musical. *J Cogn Neurosci* 12: 520–541, 2000. doi:10.1162/089892900562183.
14. **Tillmann B.** Implicit investigations of tonal knowledge in nonmusician listeners. *Ann NY Acad Sci* 1060: 100–110, 2005. doi:10.1196/annals.1360.007.
15. **Tillmann B**, **Bigand E**, **Bharucha JJ.** Implicit learning of tonality: a self-organizing approach. *Psychol Rev* 107: 885–913, 2000. doi:10.1037/0033-295x.107.4.885.
16. **Baker CI**, **Liu J**, **Wald LL**, **Kwong KK**, **Benner T**, **Kanwisher NG.** Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc Natl Acad Sci USA* 104: 9087–9092, 2007. doi:10.1073/pnas.0703300104.
17. **Dehaene S**, **Pegado F**, **Braga LW**, **Ventura P**, **Nunes Filho G**, **Jobert A**, **Dehaene-Lambertz G**, **Kolinsky R**, **Morais J**, **Cohen L.** How learning to read changes the cortical networks for vision and language. *Science* 330: 1359–1364, 2010. doi:10.1126/science.1194140.
18. **Dehaene-Lambertz G**, **Monzalvo K**, **Dehaene S.** The emergence of the visual word form: longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol* 16: 1–34, 2018. doi:10.1371/journal.pbio.2004103.
19. **Weinberger NM**, **Javid R**, **Lepan B.** Long-term retention of learning-induced receptive-field plasticity in the auditory cortex. *Proc Natl Acad Sci USA* 90: 2394–2398, 1993. doi:10.1073/pnas.90.6.2394.
20. **Blake DT**, **Strata F**, **Churchland AK**, **Merzenich MM.** Neural correlates of instrumental learning in primary auditory cortex. *Proc Natl Acad Sci USA* 99: 10114–10119, 2002. doi:10.1073/pnas.092278099.
21. **Recanzone GG**, **Schreiner CE**, **Merzenich MM.** Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci* 13: 87–103, 1993. doi:10.1523/JNEUROSCI.13-01-00087.1993.
22. **Bieszczad KM**, **Weinberger NM.** Learning strategy trumps motivational level in determining learning-induced auditory cortical plasticity. *Neurobiol Learn Mem* 93: 229–239, 2010. doi:10.1016/j.nlm.2009.10.003.
23. **Polley DB**, **Steinberg EE**, **Merzenich MM.** Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26: 4970–4982, 2006. doi:10.1523/JNEUROSCI.3771-05.2006.
24. **Ahissar E**, **Abeles M**, **Ahissar M**, **Haidarliu S**, **Vaadia E.** Hebbian-like functional plasticity in the auditory cortex of the behaving monkey. *Neuropharmacology* 37: 633–655, 1998. doi:10.1016/s0028-3908(98)00068-9.
25. **Ahissar E**, **Vaadia E**, **Ahissar M**, **Bergman H**, **Arieli A**, **Abeles M.** Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. *Science* 257: 1412–1415, 1992. doi:10.1126/science.1529342.
26. **Fritz J**, **Elhilali M**, **Shamma S.** Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res* 206: 159–176, 2005. doi:10.1016/j.heares.2005.01.015.
27. **Ohl FW**, **Scheich H.** Learning-induced plasticity in animal and human auditory cortex. *Curr Opin Neurobiol* 15: 470–477, 2005. doi:10.1016/j.conb.2005.07.002.
28. **Bakin JS**, **Weinberger NM.** Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proc Natl Acad Sci USA* 93: 11219–11224, 1996. doi:10.1073/pnas.93.20.11219.
29. **David SV**, **Fritz JB**, **Shamma SA.** Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci USA* 109: 2144–2149, 2012. doi:10.1073/pnas.1117717109.
30. **Bao S**, **Chan VT**, **Merzenich MM.** Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412: 92–95, 2001. doi:10.1038/35083586.
31. **Blake DT**, **Heiser MA**, **Caywood M**, **Merzenich MM.** Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. *Neuron* 52: 371–381, 2006. doi:10.1016/j.neuron.2006.08.009.
32. **Kilgard MP**, **Pandya PK**, **Vazquez J**, **Gehi A**, **Schreiner CE**, **Merzenich MM.** Sensory input directs spatial and temporal plasticity in primary auditory cortex. *J Neurophysiol* 86: 326–338, 2001. doi:10.1152/jn.2001.86.1.326.
33. **Rutkowski RG**, **Weinberger NM.** Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. *Proc Natl Acad Sci USA* 102: 13664–13669, 2005. doi:10.1073/pnas.0506838102.
34. **Bieszczad KM**, **Weinberger NM.** Extinction reveals that primary sensory cortex predicts reinforcement outcome. *Eur J Neurosci* 35: 598–613, 2012. doi:10.1111/j.1460-9568.2011.07974.x.
35. **Reed A**, **Riley J**, **Carraway R**, **Carrasco A**, **Perez C**, **Jakkamsetti V**, **Kilgard MP.** Cortical map plasticity improves learning but is not necessary for improved performance. *Neuron* 70: 121–131, 2011. doi:10.1016/j.neuron.2011.02.038.
36. **Blood AJ**, **Zatorre RJ.** Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci USA* 98: 11818–11823, 2001. doi:10.1073/pnas.191355898.
37. **Salimpoor VN**, **Benovoy M**, **Larcher K**, **Dagher A**, **Zatorre RJ.** Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat Neurosci* 14: 257–264, 2011. doi:10.1038/nn.2726.
38. **Salimpoor VN**, **Van Den Bosch I**, **Kovacevic N**, **McIntosh AR**, **Dagher A**, **Zatorre RJ .** Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340: 216–219, 2013. doi:10.1126/science.1231059.

39. **Golestani N**, **Price CJ**, **Scott SK.** Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *J Neurosci* 31: 4213–4220, 2011. doi:10.1523/JNEUROSCI.3891-10.2011.
40. **Teki S**, **Kumar S**, **von Kriegstein K**, **Stewart L**, **Rebecca Lyness C**, **Moore BCJ**, **Capleton B**, **Griffiths TD.** Navigating the auditory scene: an expert role for the hippocampus. *J Neurosci* 32:12251–12257, 2012. doi:10.1523/JNEUROSCI.0082-12.2012.
41. **Ohnishi T**, **Matsuda H**, **Asada T**, **Aruga M**, **Hirakata M**, **Nishikawa M**, **Katoh A**, **Imabayashi E.** Functional anatomy of musical perception in musicians. *Cereb Cortex* 11: 754–760, 2001. doi:10.1093/cercor/11.8.754.
42. **Pantev C**, **Roberts LE**, **Schulz M**, **Engelien A**, **Ross B.** Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport* 12: 959–965, 2001. doi:10.1097/00001756-200101220-00041.
43. **Shahin A**, **Bosnyak DJ**, **Trainor LJ**, **Roberts LE.** Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci* 23: 5545–5552, 2003.
44. **Fujioka T**, **Trainor LJ**, **Ross B**, **Kakigi R**, **Pantev C.** Musical training enhances automatic encoding of melodic contour and interval structure. *J Cogn Neurosci* 16: 1010–1021, 2004. doi:10.1162/0898929041502706.
45. **Fujioka T**, **Trainor LJ**, **Ross B**, **Kakigi R**, **Pantev C.** Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J Cogni Neurosci* 17: 1578–1592, 2005. doi:10.1162/089892905774597263.
46. **Besson M**, **Schön D**, **Moreno S**, **Santos A**, **Magne C.** Influence of musical expertise and musical training on pitch processing in music and language. *Restor Neurol Neurosci* 25: 399–410, 2007.
47. **Wong PCM**, **Skoe E**, **Russo NM**, **Dees T**, **Kraus N.** Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10: 420–422, 2007. doi:10.1038/nn1872.
48. **Dick F**, **Ling Lee H**, **Nusbaum H**, **Price CJ.** Auditory-motor expertise alters "speech selectivity" in professional musicians and actors. *Cereb Cortex* 21: 938–948, 2011. doi:10.1093/cercor/bhq166.
49. **Lee H**, **Noppeney U.** Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc Natl Acad Sci* 108: E1441–E1450, 2011. doi:10.1073/pnas.1115267108.
50. **Ellis RJ**, **Norton AC**, **Overy K**, **Winner E**, **Alsop DC**, **Schlaug G.** Differentiating maturational and training influences on fMRI activation during music processing. *NeuroImage* 60: 1902–1912, 2012. doi:10.1016/j.neuroimage.2012.01.138.
51. **Ellis RJ**, **Bruijn B**, **Norton AC**, **Winner E**, **Schlaug G.** Training-mediated leftward asymmetries during music processing: a cross-sectional and longitudinal fMRI analysis. *NeuroImage* 75: 97–107, 2013. doi:10.1016/j.neuroimage.2013.02.045.
52. **Angulo-Perkins A**, **Aubé W**, **Peretz I**, **Barrios FA**, **Armony JL**, **Concha L.** Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. *Cortex* 59: 126–137, 2014. doi:10.1016/j.cortex.2014.07.013.
53. **Doelling KB**, **Poeppel D.** Cortical entrainment to music and its modulation by expertise. *Proc Natl Acad Sci* 112: E6233–E6242, 2015. doi:10.1073/pnas.1508431112.
54. **Lappe C**, **Lappe M**, **Pantev C.** Differential processing of melodic, rhythmic and simple tone deviations in musicians-an MEG study. *NeuroImage* 124: 898–905, 2016. doi:10.1016/j.neuroimage.2015.09.059.
55. **Norman-Haignere SV**, **McDermott JH.** Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biol* 16: e2005127, 2018. doi:10.1371/journal.pbio.2005127.
56. **Penhune VB.** Sensitive periods in human development: evidence from musical training. *Cortex* 47: 1126–1137, 2011. doi:10.1016/j.cortex.2011.05.010.
57. **Barratt W.** *The Barratt Simplified Measure of Social Status (BSMSS)*. Terre Haute, IN: Indiana State University, 2006.
58. **Hall DA**, **Haggard MP**, **Akeroyd MA**, **Palmer AR**, **Summerfield AQ**, **Elliott MR**, **Gurney EM**, **Bowtell RW.** "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp* 7: 213–223, 1999. doi:10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N.
59. **Greve DN**, **Fischl B.** Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 48: 63–72, 2009. doi:10.1016/j.neuroimage.2009.06.060.
60. **Jenkinson M**, **Smith S.** A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5: 143–156, 2001. doi:10.1016/s1361-8415(01)00036-6.
61. **Dale AM**, **Fischl B**, **Sereno MI.** Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* 9: 179–194, 1999. doi:10.1006/nimg.1998.0395.
62. **Moddemeijer R.** On estimation of entropy and mutual information of continuous distributions. *Signal Processing* 16: 233–248, 1989. doi:10.1016/0165-1684(89)90132-1.
63. **Genovese CR**, **Lazar NA**, **Nichols T.** Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15: 870–878, 2002. doi:10.1006/nimg.2001.1037.
64. **Glasser MF**, **Coalson TS**, **Robinson EC**, **Hacker CD**, **Harwell J**, **Yacoub E**, **Ugurbil K**, **Andersson J**, **Beckmann CF**, **Jenkinson M**, **Smith SM**, **Van Essen DC.** A multi-modal parcellation of human cerebral cortex. *Nature* 536: 171–178, 2016. doi:10.1038/nature18933.
65. **Kuhn HW.** The Hungarian method for the assignment problem. *Nav Res Logist Q* 2: 83–97, 1995. doi:10.1002/nav.3800020109.
66. **Chi T**, **Ru P**, **Shamma SA.** Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 1181: 887–906, 2005. doi:10.1121/1.1945807.
67. **Da Costa S**, **Van Der Zwaag W**, **Marques JP**, **Frackowiak RSJ**, **Clarke S**, **Saenz M.** Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 14067–14075, 2011. doi:10.1523/JNEUROSCI.2000-11.2011.
68. **Herdener M**, **Esposito F**, **Scheffler K**, **Schneider P**, **Logothetis NK**, **Uludag K**, **Kayser C.** Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex* 49: 2822–2833, 2013. doi:10.1016/j.cortex.2013.04.003.
69. **Hullett PW**, **Hamilton LS**, **Mesgarani N**, **Schreiner CE**, **Chang EF.** Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *J Neurosci* 36: 2014–2026, 2016. doi:10.1523/JNEUROSCI.1779-15.2016.
70. **Humphries C**, **Liebenthal E**, **Binder JR.** Tonotopic organization of human auditory cortex. *NeuroImage* 50: 1202–1211, 2010. doi:10.1016/j.neuroimage.2010.01.046.
71. **Santoro R**, **Moerel M**, **De Martino F**, **Goebel R**, **Ugurbil K**, **Yacoub E**, **Formisano E.** Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol* 10: e1003412, 2014. doi:10.1371/journal.pcbi.1003412.
72. **Schönwiesner M**, **Zatorre RJ.** Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc Natl Acad Sci USA* 106: 14611–14616, 2009. doi:10.1073/pnas.0907682106.
73. **Hickok G**, **Poeppel D.** The cortical organization of speech processing. *Nat Rev Neurosci* 8: 393–402, 2007. doi:10.1038/nrn2113.
74. **Overath T**, **Mcdermott JH**, **Zarate JM**, **Poeppel D.** The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat Neurosci* 18: 903–911, 2015. doi:10.1038/nn.4021.
75. **Scott SK**, **Blank CC**, **Rosen S**, **Wise RJS.** Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123: 2400–2406, 2000. doi:10.1093/brain/123.12.2400.
76. **Armony JL**, **Aubé W**, **Angulo-Perkins A**, **Peretz I**, **Concha L.** The specificity of neural responses to music and their relation to voice processing: an fMRI-adaptation study. *Neurosci Lett* 593: 35–39, 2015. doi:10.1016/j.neulet.2015.03.011.
77. **Margulis EH**, **Mlsna LM**, **Uppunda AK**, **Parrish TB**, **Wong PCM.** Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Hum Brain Mapp* 30: 267–275, 2009. doi:10.1002/hbm.20503.
78. **JASP Team**. JASP (Version 0.13.1), 2020.
79. **Lee MD**, **Wagenmakers EJ.** *Bayesian Data Analysis for Cognitive Science: A Practical Course*. New York, NY: Cambridge University Press, 2013.
80. **Darwin C.** *The Descent of Man and Selection in Relation to Sex*. London: John Murray, 1871.
81. **Kell AJE**, **Yamins DLK**, **Shook EN**, **Norman-Haignere SV**, **McDermott JH.** A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 98: 630–644.e16, 2018. doi:10.1016/j.neuron.2018.03.044.
82. **Carlson NL**, **Ming VL**, **DeWeese MR.** Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. *PLoS Comput Biol* 8: e1002594, 2012. doi:10.1371/journal.pcbi.1002594.

83. **Lewicki MS.** Efficient coding of natural sounds. *Nat Neurosci* 5: 356–363, 2002. doi:10.1038/nn831.
84. **Młynarski W**, **McDermott JH.** Ecological origins of perceptual grouping principles in the auditory system. *Proc Natl Acad Sci USA* 116: 25355–25364, 2019. doi:10.1073/pnas.1903887116.
85. **Jacoby N**, **McDermott JH.** Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr Biol* 27: 359–370, 2017. doi:10.1016/j.cub.2016.12.031.
86. **Jacoby N**, **Undurraga EA**, **McPherson MJ**, **Valdés J**, **Ossandón T**, **McDermott JH.** Universal and non-universal features of musical pitch perception revealed by singing. *Curr Biol* 29: 3229–3243, 2019. doi:10.1016/j.cub.2019.08.020.
87. **McDermott JH**, **Schultz AF**, **Undurraga EA**, **Godoy RA.** Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535: 547–550, 2016. doi:10.1038/nature18635.
88. **McPherson MJ**, **Dolan SE**, **Durango A**, **Ossandon T**, **Valdés J**, **Undurraga EA**, **Jacoby N**, **Godoy RA**, **McDermott JH.** Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nat Commun* 11: 1–14, 2020. doi:10.1038/s41467-020-16448-6.
89. **Ding N**, **Patel AD**, **Chen L**, **Butler H**, **Luo C**, **Poeppel D.** Temporal modulations in speech and music. *Neurosci Biobehav Rev* 81:181–187, 2017. doi:10.1016/j.neubiorev.2017.02.011.
90. **Albouy P**, **Benjamin L**, **Morillon B**, **Zatorre RJ.** Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* 367: 1043–1047, 2020. doi:10.1126/science.aaz3468.
91. **McDermott JH**, **Simoncelli EP.** Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71: 926–940, 2011. doi:10.1016/j.neuron.2011.06.032.
92. **Zuk NJ**, **Teoh ES**, **Lalor EC.** EEG-based classification of natural sounds reveals specialized responses to speech and music. *NeuroImage* 210: 116558, 2020. doi:10.1016/j.neuroimage.2020.116558.
93. **Brett M**, **Grahn JA.** Rhythm and beat perception in motor areas of the brain. *J Cogn Neurosci* 19: 893–906, 2007. doi:10.1162/jocn.2007.19.5.893.
94. **Janata P**, **Birk JL**, **Van Horn JD**, **Leman M**, **Tillmann B**, **Bharucha JJ.** The cortical topography of tonal structures underlying western music. *Science* 298: 2167–2170, 2002. doi:10.1126/science.1076262.
95. **Lee YS**, **Janata P**, **Frost C**, **Hanke M**, **Granger R.** Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage* 57: 293–300, 2011. doi:10.1016/j.neuroimage.2011.02.006.
96. **Matthews TE**, **Witek MAG**, **Lund T**, **Vuust P**, **Penhune VB.** The sensation of groove engages motor and reward networks. *NeuroImage* 214: 116768, 2020. doi:10.1016/j.neuroimage.2020.116768.
97. **Norman-Haignere SV**, **Long LK**, **Devinsky O**, **Doyle W**, **Irobunda I**, **Merricks EM**, **Feldstein NA**, **McKhann GM**, **Schevon CA**, **Flinker A**, **Mesgarani NA.** Multiscale integration organizes hierarchical computation in human auditory cortex (Preprint). *bioRxiv* 321687, 2020. doi:10.1101/2020.09.30.321687.
98. **Fritz J**, **Shamma S**, **Elhilali M**, **Klein D.** Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6: 1216–1223, 2003. doi:10.1038/nn1141.
99. **Norman-Haignere SV**, **Albouy P**, **Caclin A**, **McDermott JH**, **Kanwisher NG**, **Tillmann B.** Pitch-responsive cortical regions in congenital amusia. *J Neurosci* 36: 2986–2994, 2016. doi:10.1523/JNEUROSCI.2705-15.2016.
100. **Albouy P**, **Caclin A**, **Norman-Haignere SV**, **Lévêque Y**, **Peretz I**, **Tillmann B**, **Zatorre RJ.** Decoding task-related functional brain imaging data to identify developmental disorders: the case of congenital amusia. *Front Neurosci* 13: 1–13, 2019. doi:10.3389/fnins.2019.01165.
101. **Brainard DH.** The psychophysics toolbox. *Spat Vis* 10: 433–436, 1997.
102. **Spiegel MF**, **Watson CS.** Performance on frequency-discrimination tasks by musicians and nonmusicians. *J Acoust Soc Am* 76: 1690–1695, 1984. doi:10.1121/1.391605.
103. **Kishon-Rabin L**, **Amir O**, **Vexler Y**, **Zaltz Y.** Pitch discrimination: are professional musicians better than non-musicians? *J Basic Clin Physiol Pharmacol* 12: 125–144, 2001. doi:10.1515/jbcpp.2001.12.2.125.
104. **Micheyl C**, **Delhommeau K**, **Perrot X**, **Oxenham AJ.** Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res* 219: 36–47, 2006. doi:10.1016/j.heares.2006.05.004.
105. **Levitt H.** Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49: 467–477, 1971. doi:10.1121/1.1912375.
106. **Repp BH.** Sensorimotor synchronization: a review of the tapping literature. *Psychon Bull Rev* 12: 969–992, 2005. doi:10.3758/bf03206433.
107. **Repp BH.** Sensorimotor synchronization and perception of timing: effects of music training and task experience. *Hum Mov Sci* 29: 200–213, 2010. doi:10.1016/j.humov.2009.08.002.
108. **Bailey JA**, **Penhune VB.** Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Exp Brain Res* 204: 91–101, 2010. doi:10.1007/s00221-010-2299-y.
109. **Polak R**, **Jacoby N**, **Fischinger T**, **Goldberg D**, **Holzapfel A**, **London J.** Rhythmic prototypes across cultures: a comparative study of tapping synchronization. *Music Percept* 36: 1–23, 2018. doi:10.1525/mp.2018.36.1.1.
110. **McDermott JH**, **Keebler MV**, **Micheyl C**, **Oxenham AJ.** Musical intervals and relative pitch: frequency resolution, not interval resolution, is special. *J Acoust Soc Am* 128: 1943–1951, 2010. doi:10.1121/1.3478785.
111. **McPherson MJ**, **McDermott JH.** Diversity in pitch perception revealed by task dependence. *Nat Hum Behav* 2: 52–66, 2018. doi:10.1038/s41562-017-0261-8.
112. **Temperley D.** A probabilistic model of melody perception. *Cogn Sci* 32: 418–444, 2008. doi:10.1080/03640210701864089.
113. **Rauschecker JP**, **Tian B**, **Hauser M.** Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111–114, 1995. doi:10.1126/science.7701330.
114. **Baumann S**, **Petkov CI**, **Griffiths TD.** A unified framework for the organisation of the primate auditory cortex. *Front Syst Neurosci* 7: 1–19, 2013. doi:10.3389/fnsys.2013.00011.
115. **Singh NC**, **Theunissen FE.** Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114: 3394, 2003. doi:10.1121/1.1624067.
116. **Rodríguez FA**, **Read HL**, **Escabí MA.** Spectral and temporal modulation tradeoff in the inferior colliculus. *J Neurophysiol* 103: 887–903, 2010. doi:10.1152/jn.00813.2009.
117. **Patterson RD**, **Uppenkamp S**, **Johnsrude IS**, **Griffiths TD.** The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36: 767–776, 2002. doi:10.1016/s0896-6273(02)01060-7.
118. **Penagos H**, **Melcher JR**, **Oxenham AJ.** A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* 24: 6810–6815, 2004. doi:10.1523/JNEUROSCI.0383-04.2004.
119. **Norman-Haignere SV**, **Kanwisher NG**, **McDermott JH.** 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, 2013. doi:10.1523/JNEUROSCI.2880-13.2013.
120. **Hamilton LS**, **Edwards E**, **Chang EF.** A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr Biol* 28: 1860–1871, 2018. doi:10.1016/j.cub.2018.04.033.
121. **Spearman C.** The proof and measurement of association between two things. *Am J Psychol* 15: 72–101, 1904. doi:10.1093/ije/dyq191.