Supporting Online Material

Subjects

Although there is compelling evidence that non-musicians possess mental representations of tonal structures, we reasoned that in an initial experiment we would be most likely to succeed in identifying the cortical loci of these structures in musically trained individuals. The ages of 4 female and 4 male listeners ranged from 20–45 years (26 ± 8.9, mean ± s.d.). One listener was left handed. Two listeners reported possessing absolute pitch. Although a test showed that these listeners indeed possessed the ability to label discrete pitches, their functional activation data did not stand apart from the rest of the listeners so they were retained as part of the cohort. The range of formal musical training was 7–19 years (12.9 ± 4.2, mean ± s.d.). Prior to the experiment, all listeners provided informed consent after reviewing forms approved by the Committee for Protection of Human Subjects at Dartmouth College.

Stimuli & Tasks

A detailed description and behavioral validation of the stimulus is provided elsewhere (*S1*). In brief, an original melody was composed that formed an endless loop and modulated through all 24 major and minor keys in the following order: C, a, E, c#, Ab, f, c, G, e, B, g#, Eb, Bb, g, D, b, F#, eb, bb, F, d, A, f#, Db. Most Western tonal music is written in the major and minor modes. Major keys are labeled with an uppercase letter. The labels of minor keys begin with a lowercase letter. The symbol, #, and lowercase letter b replace the words "sharp" and "flat", respectively.

A harmonic progression was defined that allowed the melody to dwell in each key for ~14.4 s and move smoothly to the next over a period of ~4.8 s. Thus, a new tonal

center was established every 19.2 s. This amount of time allowed for a hemodynamic response to develop fully in areas that might be sensitive to the particular key that the melody was centered around within the 19.2 s window. The notes of the chords defining the harmonic progression were arpeggiated and presented in 6/8 meter with a note onset asynchrony of 200 ms. Six melodies, for use in each task, were derived from the original by temporally shifting the original so that it would start in a different key. The starting key was varied in order to avoid a confound of tonality sensitive responses with effects associated with the amount of time elapsed from the beginning of the functional scan. Overall there were seven different starting keys: Gb, B, Bb, Ab, E, D, or Eb.

In the tonality violation task the melody began in one of six keys, and three different test tones were used: A (220 Hz) for the keys of Bb and Ab; C3 (262 Hz) for the keys of Gb and B; and Eb3 (311 Hz) for the keys of E and D. Test tones occurred every 4 seconds on average and represented 4% of the notes in the melody. Because the test tones would blend into some keys and pop out in others, listeners' rates of responding fluctuated in this task (Fig. 1D). Note that during each run, the melody modulated through all 24 keys; all that varied from run to run was the starting key and the identity of the test tone. Our task is a variant on the traditional probe-tone task in which listeners how well a discrete probe tone fits into a preceding tonal context (*S2*). Recently, the probe-tone task has been implemented as a continuous monitoring task in order to obtain moment-to-moment tonality estimates (*S3*).

Timbral deviance detection task. Flute deviants occurred every 4 seconds on average and constituted 4% of the notes in the melody. Listeners detected deviants quickly (M = 437 ms, S.E.M = 20 ms) and accurately (M= 87%, S.E.M.= 6%). In

contrast to the tonality violation task, the timbral deviance of notes played by the flute was equisalient in all keys. Thus, rates of responding were constant in the timbre deviance detection task.

During each session, listeners heard four of the twelve melodies and performed each task twice in alternation. Over all the sessions they heard all of the melodies. The order in which the tasks were performed and the melodies heard were counterbalanced across sessions and listeners. Thus, if a listener received the timbre deviance detection task during the first run of the first session, she received the tonality violation task as the first run of the second session. During a 30 minute session prior to the first fMRI scanning session, listeners were familiarized with the melody and tasks. Listeners found the tasks challenging but had no trouble performing them.

Stimulus preparation. The melodies were played via MIDI (Performer 6.01, Mark of the Unicorn) from an iMac (Apple Computer, Cupertino). The sounds were rendered with the "clarinet" patch of an FM tone generator (TX802, Yamaha) and recorded to disk (SoundEdit 16, Macromedia). Each melody was stored in one channel of an audio file. A magnet trigger pulse and assorted event markers were added to the other channel. The file for each melody was burned to a separate CD track.

Scanning procedures

Continuous whole-brain BOLD signal was acquired with a 1.5 T GE Signa MRI scanner using the following echoplanar imaging (EPI) pulse sequence parameters: TE: 35 ms, TR: 3s, 27 slices, slice thickness: 5.0 mm, slice skip: 0 mm, interleaved slice acquisition, field-of-view (FOV) = 240 x 240 mm; flip angle = 90°; matrix size = 64 x 64; in-plane resolution = 3.75 x 3.75 mm. In each scanning session we also obtained a T1-weighted image with the same slice orientation as the EPI images. The stimuli were

delivered to the listeners via pneumatic headphones (ER-30, Etymotic Research) at ~90 dB SPL. All listeners reported being able to clearly segregate the melody from the background pinging.

An event marker on the stimulus CD triggered EPI acquisition on each run. Each run began with the acquisition of 2 volumes (6 s) of dummy images that were discarded, followed by 60 s of rest. Three high-pitched warning tones were sounded 6 s prior to the onset of the melody. The melody lasted 7 min 40.8 s, and listeners responded to test tones by pressing a button with their right thumbs. An additional 60 s rest period followed the end of the melody, whereupon collection of images ceased. Thus, a total of 194 images volumes were collected during each run. An additional file was recorded during each run with the signals from chest bellows that monitored respiration, thresholded output from a pulse oxymeter, the magnet's receiver-unblank output for each acquired slice, event markers from the stimulus CD, and listener responses. These signals were sampled at 250 Hz and were used for assessing behavioral performance and determining the timing of events during construction of the design matrix.

fMRI analysis procedures

Image preprocessing. Translational and rotational motion parameters were estimated for the functional runs of each session using SPM99 (http://www.fil.ion.ucl.ac.uk/spm; *S4*). These estimates were used to reslice the EPI images. We performed no further spatial adjustments (realignment or normalization) or spatial smoothing prior to analyzing the data because of the slice-specific design matrices that we employed. Each voxel's time-series was standardized within each run.

Design matrix construction. A separate design matrix was constructed for each slice through the image volume (Fig. S1B). In order to remove variance that was not directly modeled by task, stimulus, or listener response parameters, we included the

following set of nuisance parameters: the aforementioned motion estimates, the respiratory signal, phase of the cardiac cycle, linear trends, run means, and linear trend by run interactions. Regressors of interest included the spherical harmonic time-series that modeled the moment-to-moment tonality surface (see "Tonality surface estimation" below), listener responses modeled as Dirac impulses located at the onsets of button presses that were then convolved with the SPM canonical HRF, the onset of the alerting cue convolved with the HRF, two task regressors (described below), and task regressor by response interaction terms.

We first performed an omnibus F-test to identify voxels whose activity was significantly predicted by the overall model (Fig. S1A). Of those voxels exceeding a nominal threshold of $p < 0.05$, the mean proportion of variance explained (R^2) ranged from 0.40–0.61 (mean= 0.48 ± 0.08 s.d.). These voxels entered into a second analysis in which the increment in the proportion of variance explained by the set of stimulus, task, and response regressors above the variance explained by the nuisance parameters was tested for significance ($p < 0.05$). 71 ± 8% of the voxels passed this test. The mean R^2 for these voxels ranged from 0.09–0.18 (mean=0.11 ± 0.03 s.d.). These voxels then entered into two separate analyses of the increments in R^2 explained by the tonality regressors and the task regressors, respectively. For these analyses we set a stricter criterion ($p < 0.001$) for considering the fluctuations in a voxel's BOLD signal to be task and/or tonality related.

In the first analysis we tested the main effect of task using contrast coding (boxcars) of two task regressors: 1) the epochs during which the tasks were performed as the melody played relative to rest, and 2) the two tasks relative to each other. Note,

detailed analyses of each of the task effects, while of great interest, are beyond the scope of this paper so we restricted our analysis to the main effect of task. Across listeners, the maximum R^2 for significant voxels (49 ± 10% of analyzed voxels) ranged from 0.28 to 0.61. Averaged across listeners, the mean R^2 was 0.05 ± 0.01 s.d..

The second analysis estimated the main effect of the moment-to-moment activation of the tonality surface irrespective of the task that was performed. Across listeners, the maximum R^2 for significant voxels (24 ± 8% of analyzed voxels) ranged from 0.22 to 0.36. Averaged across listeners, the mean R^2 was 0.10 ± 0.03 s.d.. The final criterion for considering a voxel to be task or tonality sensitive was that the voxel exceed the $p < 0.001$ significance threshold in all of the scanning sessions for a listener. In order to compare statistical maps across scanning sessions, we computed affine transformation matrices as follows. The mean of the resliced EPI images was coregistered with a mutual information algorithm (*S5*) with the T1-weighted coplanar anatomical image that was acquired prior to the functional runs in each session. The coplanar images were then coregistered with the average of two T1-weighted high resolution structural images that were obtained in two of the sessions for each listener. The affine transformation parameters for the latter coregistration step were propagated to the mean EPI image. Thus, the statistical maps from all sessions could be transformed into the space of the first session which was arbitrarily chosen as the reference session. For those voxels exhibiting a significant main effect of the tonality regressors across sessions, we obtained β estimates for use in reconstructing the voxel tonality sensitivity surfaces as follows. We first removed the variance associated with all other variables in the model, and then fit the tonality regressors to the residuals. We reconstructed the tonality surface, as described

below, only for voxels in clusters of five or more voxels that were considered to be tonality sensitive.

Tonality surface estimation

The scheme for using moment-to-moment tonality estimates of the actual stimuli to identify tonality sensitive regions of the cortex is shown in Fig. S2. The moment-to-moment tonality surface activation patterns were estimated for each version of the stimulus by passing the stimulus audio files through a computational model of the auditory periphery coupled to a self-organizing map (SOM) neural network. Values of the SOM outputs comprised the tonality surface activation. Previous research has shown that SOM neural networks can be used to recover the topology of key relationships predicted by music theory and cognitive psychology (*S3, S6–S7*). We implemented the auditory model in several stages using the IPEM Toolbox (http://www.ipem.rug.ac.be; *S8*). The first stage estimated auditory nerve firing patterns. The second stage extracted periodicity pitch estimates by cross-correlating the auditory nerve patterns in 38 ms time windows. The third stage temporally filtered the periodicity pitch images with a 2 s time constant. The filtered pitch images served as the input to the SOM.

The SOM was implemented using the Finnish SOM Toolbox (http://www.cis.hut.fi/projects/somtoolbox/) and consisted of a single input layer fully connected to 192 output units arranged as a hexagonal grid of 12 by 16 units. The distances among output units were defined such that the top and bottom rows of units were neighbors as were the left and right columns. Thus the output surface topology of the SOM was a toroidal surface. The SOM was trained for 200 iterations using the standard batch training procedures described in the toolbox. Relevant parameters were the use of a gaussian neighborhood with an initial radius of 3 and final radius of 1. Weights were initialized to random values between 0 and 1. The SOM was trained using the original version of the melody which contained no test tones or timbral deviants.

However, because each stimulus melody will give rise to a different activation time-course on the toroidal surface, we used the SOM output arising from each stimulus melody to construct the tonality regressors for estimating the sensitivity of cortical areas to different tonalities. The construction of the regressors is described in detail below.

Because the initial weights in the SOM are set to random values, the absolute spatial organization of the different keys on the toroidal surface differs for each SOM training session. Given that multiple SOM models will yield as many different topographic maps, one cannot simply average the final output surfaces to determine whether the training procedures result in stable tonality classification behavior. Therefore, we projected each toroidal surface to a 24-element vector corresponding to the individual keys as follows. For each time window from the 2nd through 6th measures of the 8 measures that were nominally assigned a single tonality, we determined the most activated output unit on the toroidal surface. We then tallied the number of times each output unit was activated while the melody was in that key. The tally for each key then served as a weighting function for mapping the activity on the output surface at any given moment to the corresponding key unit in the 24-element vector. The temporal activation patterns on the 24-element vector corresponded well to the known tonal location of the melody. In other words, when it was known that the melody was in g-minor, the g-minor unit was activated most strongly. To assess the stability of the SOM classification approach, we trained 10 networks. Despite slight variation in the topographical relationships among the keys on the output surface, and differences in the absolute locations of a key from one SOM surface to the next, very little variation was observed in the activity pattern of the 24-element key vector across individual SOMs (Fig. S3). Thus, the first SOM was arbitrarily chosen to simulate the activation of the tonality surface by each of the stimulus melodies.

Tonality sensitive regions of the cortex are defined as those areas whose fluctuations in BOLD signal are correlated with movement of the activation locus on the tonality surface. Consequently, the tonality regressors are a model of the moment-to-moment fluctuations in activation patterns on the tonality surface that are then used to identify tonality sensitive regions . Rather than introduce the time series from all of the 192 SOM output units into the design matrix as regressors, we reduced the number of regressors that were needed to describe the moment-to-moment activation of the toroidal tonality surface by decomposing the toroidal surface at each time point into its component spherical harmonics (Eq. 1, *S9*),

$$f(\theta,\phi) = \sum_{m,n}^{\infty} a_{mn}^{cc} \cos(m\theta)\cos(n\phi) + \sum_{m,n}^{\infty} a_{mn}^{cs} \cos(m\theta)\sin(n\phi) + \sum_{m,n}^{\infty} a_{mn}^{sc} \sin(m\theta)\cos(n\phi) + \sum_{m,n}^{\infty} a_{mn}^{ss} \sin(m\theta)\sin(n\phi) \quad \text{(Eq. 1)}$$

where the harmonic indices for *m* and *n* ranged from 0 to 2 (*m*) and 0 to 3 (*n*). This resulted in 48 amplitude (*a)* parameter estimates for the toroidal surface at each time point. The superscripts *cc*, *cs*, etc. simply identify amplitude parameters as belonging to the cos-cos, cos-sin, etc. terms, but do not assume numerical values. Even though the highest spatial frequencies were note estimated because the maximum number of harmonics along each dimension was set to a value below the Nyquist frequency, reconstructions of the toroidal surfaces for the stimulus melodies using the reduced set of estimated parameters explained over 98% of the variance in the original surfaces. The number of regressors was further reduced to 35 because amplitude parameter estimates for sin terms containing either *m* or *n* equal to zero are necessarily zero. The time-series of the spherical harmonic parameter estimates for each stimulus melody were then low-pass filtered with the canonical hemodynamic response function (HRF) and entered into the fMRI design matrix. The HRF is a generalized approximation of the BOLD signal

change in response to a stimulus event. It is a composite of two gamma functions, peaks at 6 s from stimulus onset, and serves as a low-pass filter.

For those voxels meeting the criteria for tonality surface reconstruction described above, the β estimates of the tonality regressors associated with each spherical harmonic were first scaled to the spherical harmonic's original time-series by multiplying by the standard deviation and adding the mean of the original time-series. They were then entered as the amplitude coefficients in the spherical harmonic expansion (Eq. 1) to obtain the voxel's tonality sensitivity surface (TSS). In order to assign a voxel to a specific tonality, we correlated its TSS with the mean activation surface for each key (Fig. 1A) as well as the surface obtained by averaging all the surfaces across the course of the melody. Given the strong correlations in the tonality surfaces among related keys (Fig. 1C), voxels exhibiting a preferred tonality (rather than the average tonality) were further classified into one of three groups of keys (Fig. 1B).

Spatial normalization

The average T1-weighted high resolution image for each listener was spatially normalized to the International Consortium for Brain Mapping's average 152 brain T1 weighted image using default procedures in SPM99. The normalization parameters were applied to those statistical images that were entered into between-listener conjunction images used to generate Fig. 2.

Supporting Online References

S1. P. Janata, J.L Birk, B. Tillmann, J.J. Bharucha, *Music Perception* (in press).

S2. C. L. Krumhansl, *Cognitive Foundations of Musical Pitch* (Oxford University Press, New York, 1990).

S3. C. L. Krumhansl, P. Toiviainen, paper presented at the 6th International Conference on Music Perception and Cognition, Keele, United Kingdom, 9 August 2000.

S4. K. J. Friston, et al., *Human Brain Mapping* **2**, 165-189 (1995).

S5. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, *IEEE Transactions on Medical Imaging* **16**, 187-98 (1997).

S6. B. Tillmann, J. J. Bharucha, E. Bigand, *Psychol. Rev.* **107**, 885 (2000).

S7. M. Leman, *Music and Schema Theory: Cognitive Foundations of Systematic Musicology* (Springer-Verlag, Berlin, Heidelberg, 1995).

S8. L. M. Vanimmerseel, J. P. Martens, *J. Acoust. Soc. Amer.* **91**, 3511 (1992).

S9. J. P. Boyd, *Chebyshev and Fourier Spectral Methods* (Dover, New York, ed. 2nd, 2001).

Supporting Online Tables

Table S1. Distributions of key membership of tonality sensitive voxels thoughout the brain. #clusters refers to the number of clusters with 5 or more significant voxels. Total #voxels is the total number of voxels in the clusters. Group refers to the key groups in Figure 1B.

Listener	#clusters	total #voxels	Session	Average	Group 1	Group 2	Group 3
1	4	44	1	2	24	3	15
			2	6	24	7	7
			3	3	5	32	4
2	6	64	1	2	26	22	14
			2	6	29	5	24
			3	6	18	19	21
			4	4	23	10	27
3	14	105	1	4	38	25	38
			2	9	26	26	44
			3	9	37	13	46
			4	2	9	45	49
4	2	20	1	2	5	4	9
			2	0	4	13	3
			3	0	6	2	12
5	31	440	1	5	169	79	187
			2	19	152	137	132
			3	27	157	95	161
6	7	47	1	6	18	19	4
			2	8	7	14	18
			3	1	29	12	5
7	9	139	1	7	62	37	33
			2	6	56	47	30
			3	11	59	30	39
8	6	81	1	4	43	11	23
			2	7	29	25	20
			3	4	13	32	32

Table S2. Anatomical distribution of tonality sensitive voxels for each listener. SFG, superior frontal gyrus; IFG, inferior frontal gyrus, STS, superior temporal sulcus

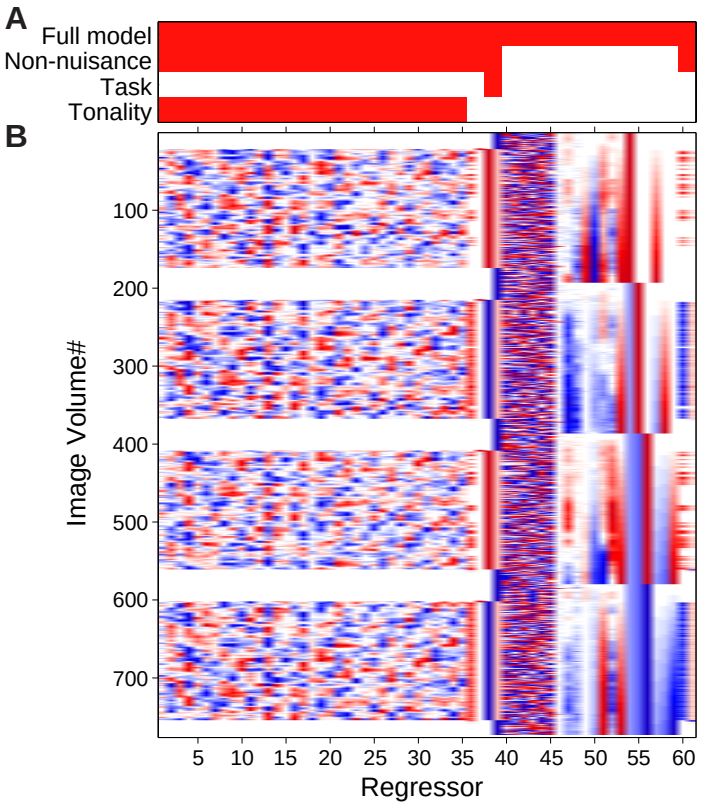
			Number of tonality sensitive voxels							
Lobe	Hemisphere	Region	L1	L2	L3	L4	L5	L6	L7	L8
Frontal										
	Bilateral									
		rostromedial SFG and frontopolar gyri	24	39	6	11	140	5	79	37
		supplementary motor area					17			
	Left									
		rostral, dorsal SFG	12		14					
		orbital gyrus					6			7
		inferior frontal sulcus					10			
		middle frontal gyrus					5			
	Right									
		supraorbital sulcus					6			
		frontomarginal sulcus		7						
		orbital gyrus			6					18
		rostral inferior frontal sulcus					11			
		IFG, pars opercularis								9
		IFG, pars triangularis		5	6				10	
		IFG, pars orbitalis							10	
		middle frontal gyrus			15				9	
		superior frontal sulcus			5		8			
		SFG					12			
		precentral gyrus					6			
Temporal										
	Left									
		temporal pole					12			
		anterior STG					8			
		fusiform gyrus						6		
		collateral sulcus						5		
	Right									
		temporal pole		7	7					
		superior temporal sulcus			7		5			
		fusiform gyrus			7		8			
Parietal										
	Bilateral									
		precuneus	8			9	7			
	Left									
		precuneus								10
		superior parietal gyrus					6			
		posterior STS					8			
	Right									
		supramarginal gyrus					25			
		posterior cingulate sulcus					5			
		superior parietal gyrus			6					
		intraparietal sulcus					35		11	
		posterior STS					6		15	
Limbic										
	Bilateral									
		anterior cingulate gyrus							5	
		posterior cingulate gyrus					10			
	Right									
		hippocampus					9			
Occipital										
	Left									
		posterior lingual gyrus			8		6	11		
		calcarine sulcus						6		
	Right									
		calcarine sulcus			9					
		superior occipital gyrus					9			
Other										
	Left									
		cerebellum			9			10		
		ventral basal ganglia					7			
	Right									
		cerebellum					38			

Supporting Online Figure Captions

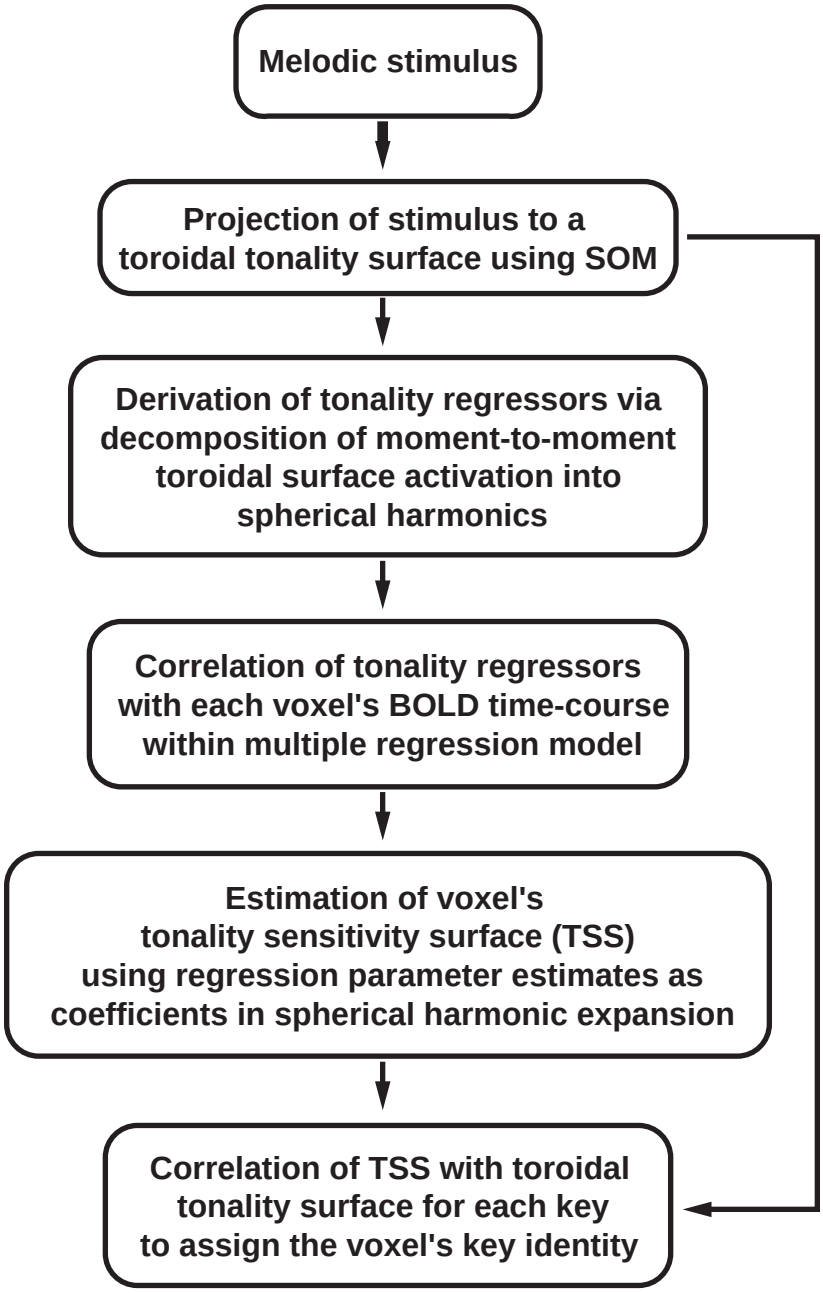
Figure S1. Design and reduced model matrices. A) Reduced model matrix. Each row indicates in red the regressors that were entered into an *F*-test for the significance of the proportion of overall variance explained by those regressors. B) A design matrix for one slice through the image volumes collected during one scanning session consisting of four runs. Run onsets occur at volume numbers 1, 195, 389, and 583. For purposes of display, values in each column have been normalized to the maximum absolute value in that column. Thus, the values range from -1 (blue) to +1 (red). The mapping between regressor groups and column numbers is as follows. Tonality surface (1–35), Response (36), Alerting Cue (37), Task (38–39), Cardiac Cycle (40–45), Respiration (46), Motion (47–52), Linear Trend (53), Run Offset (54–56), Run X Linear Interaction (57–59), Response X Task Interaction (60–61).

Figure S2. Data analysis flowchart showing the relationship of the tonality surface of the SOM and the estimated tonality sensitivity surfaces of fMRI voxels.

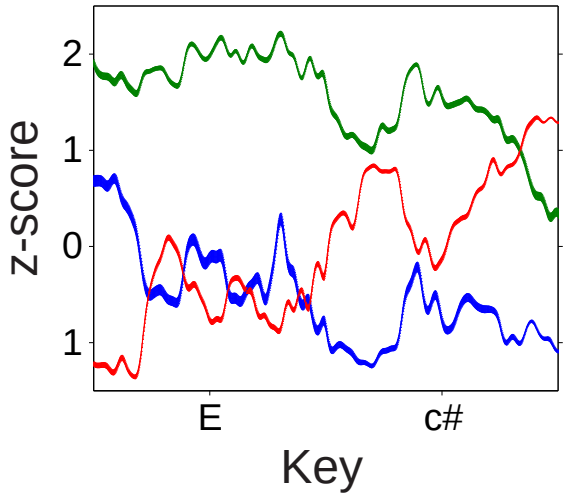
Figure S3. Consistency of tonality classification by ten trained SOM networks. The trace shows an excerpt of the time-varying magnitudes of units in the 24-element key vector corresponding to C major (blue), E major (green), and Ab major (red). The width of each trace indicates the standard error of the mean across the ten networks. The traces are shown for a period of time when the melody resided in E major and c# minor.



Janata et al., Figure S1



Janata et al., Figure S2



Janata et al., Figure S3