NeuroImage 59 (2012) 3677–3689

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm

Vinoo Alluri [a,*], Petri Toiviainen [a,1], Iiro P. Jääskeläinen [c,d,2], Enrico Glerean [c,d,2], Mikko Sams [c,d,2], Elvira Brattico [a,b,3]

[a] Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland
[b] Cognitive Brain Research Unit, Institute of Behavioral Sciences, University of Helsinki, Finland
[c] Mind and Brain Laboratory, Department of Biomedical Engineering and Computational Science (BECS), Aalto University School of Science, Finland
[d] Advanced Magnetic Imaging Centre, Aalto University School of Science, Finland

ARTICLE INFO

Article history:
Received 21 June 2011
Revised 13 October 2011
Accepted 6 November 2011
Available online 12 November 2011

Keywords:
Music processing
fMRI
Computational feature extraction
Naturalistic stimulus
Temporal evolution

ABSTRACT

We investigated the neural underpinnings of timbral, tonal, and rhythmic features of a naturalistic musical stimulus. Participants were scanned with functional Magnetic Resonance Imaging (fMRI) while listening to a stimulus with a rich musical structure, a modern tango. We correlated temporal evolutions of timbral, tonal, and rhythmic features of the stimulus, extracted using acoustic feature extraction procedures, with the fMRI time series. Results corroborate those obtained with controlled stimuli in previous studies and highlight additional areas recruited during musical feature processing. While timbral feature processing was associated with activations in cognitive areas of the cerebellum, and sensory and default mode network cerebrocortical areas, musical pulse and tonality processing recruited cortical and subcortical cognitive, motor and emotion-related circuits. In sum, by combining neuroimaging, acoustic feature extraction and behavioral methods, we revealed the large-scale cognitive, motor and limbic brain circuitry dedicated to acoustic feature processing during listening to a naturalistic stimulus. In addition to these novel findings, our study has practical relevance as it provides a powerful means to localize neural processing of individual acoustical features, be it those of music, speech, or soundscapes, in ecological settings.

Introduction

Music is fundamental to humans across all cultures and is capable of eliciting intense emotions (Salimpoor et al., 2011). Uncovering the neural underpinnings of music processing has become a central theme in cognitive neuroscience in the past decade, as evidenced by the constantly increasing corpus of studies on this topic. The intrinsically multidimensional nature of music renders this task challenging. More specifically, music comprises several perceivable features of varying levels of abstraction, such as loudness, pitch (the organization of sounds along a scale from low to high), rhythm (the perceptual organization of sound events in time) and timbre (property that allows to distinguish between different instrument sounds having the same pitch and loudness). Perceiving polyphonic music involves automatic segregation of the musical information in the brain. For instance, when listening to a piece of music played by an orchestra we are able to distinguish one instrument's timbre from that of another, perceive the leading melody by extracting pitch height, and feel the beat (Bregman, 1990; Janata et al., 2002a, 2002b). In this process, domain-specific neural mechanisms for acoustic feature analysis and integration as well as domain-general neural circuits of attention and memory are required. In particular, hierarchical processing within the auditory cortex going from more simple to more complex features (Chevillet et al., 2011; Patterson et al., 2002), and hemispheric specialization (Samson et al., 2011; Zatorre et al., 2002) for spectral *vs.* temporal acoustic variations, have been identified as putative principles of functional organization of acoustic feature-related processing. Previous neuroimaging studies of music have attempted to identify brain structures involved in the perception of music-related perceptual features, such as pitch (Patterson et al., 2002), sensory dissonance (Blood et al., 1999; Koelsch et al., 2006), rhythm (Chen et al., 2008; Grahn and Rowe, 2009), timbre (Caclin et al., 2006; Halpern et al., 2004), and key (Janata et al., 2002a, 2002b). However, while these studies have successfully identified brain regions participating in processing of individual musical features they have relied on controlled auditory paradigms in which these features have been presented in isolation and manipulated artificially. Although a few studies have investigated brain responses during continuous listening to relatively simple musical stimuli (Janata et al., 2002a, 2002b; Schaefer et al., 2009), it has not been previously studied how the human brain processes, in parallel, the multitude of musical features

* Corresponding author at: Department of Music, University of Jyväskylä, PL 35(M), 40014 Jyväskylä, Finland.
E-mail address: vinoo.alluri@jyu.fi (V. Alluri).
[1] Department of Music, University of Jyväskylä, PL 35(M), 40014 Jyväskylä, Finland.
[2] BECS, Dept. of Biomedical Engineering and Computational Science, P.O.Box 12200, FI-00076 Aalto, Finland.
[3] Institute of Behavioral Sciences, P.O.B. 9, 00014, University of Helsinki, Finland.

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.11.019

when participants are listening to a record of real orchestra music during neuroimaging.

In the visual modality, recent evidence suggests that the brain processes visual stimuli presented in a more ecological setting differently than when presented in conventional controlled settings (Hasson et al., 2004). Assuming that this finding is generalizable across sensory modalities, one could expect that the majority of studies in the auditory modality, in which acoustic features were artificially manipulated, may have revealed an incomplete picture of brain function related to musical feature processing. Therefore, studying music listening as a continuous process using naturalistic stimuli could provide more accurate accounts of the processing of musical features in the brain.

We employed a stimulus-wise and task-wise more ecological setting in which participants freely listened to real music without performing any other task, in order to determine the neural mechanisms and structures responsible for musical feature processing under realistic conditions. To tackle the complexity of the problem, we introduced a novel interdisciplinary approach combining neuroimaging with computational acoustic feature extraction and behavioral psychology. As music stimulus we chose the modern tango *Adios Nonino* by Astor Piazzolla. The participants were scanned with fMRI while listening to this piece. Temporal evolutions of acoustic components representing timbral, tonal and rhythmic features of the stimulus were computationally extracted and validated via a perceptual experiment. Following this we performed correlation analyses of the time series of the individual acoustic components and the time series of the BOLD signal.

In light of previous studies (Samson et al., 2011), we hypothesized that timbral components would activate mainly sensory areas, such as the superior temporal gyrus (STG) and the Heschl's gyrus (HG). Moreover, we expected that the spectrally varying timbral components would activate particularly the caudolateral and anterior superior temporal regions, respectively (see Samson et al., 2011 for an overview). In addition, we predicted hemispheric lateralization, in particular, that right hemispheric regions would show larger areas involved in the processing of these features (Zatorre et al., 2002). Processing of tonality-related features was expected to recruit areas in the brain formerly known to be neural substrates of tonality processing, such as the rostromedial prefrontal cortex (Janata et al., 2002a, 2002b). In addition, as tonality processing draws on long-term knowledge of hierarchical tonality structures (Krumhansl, 1990), we expected the brain areas related to memory processing, such as the hippocampus (see Burianova et al., 2010 for an overview) to be activated. We hypothesized rhythm-related features to recruit, in addition to areas in the auditory cortex, cortical and subcortical areas related to motor processing, such as the premotor and supplementary motor areas, and subcortical structures involved in the processing of time intervals such as the basal ganglia (Harrington et al., 1998; Janata and Grafton, 2003; Rao et al., 2001; Schwartze et al., 2011). Furthermore, as tonal and rhythmic features are known to elicit expectations in listeners (Janata, 2005; Zanto et al., 2006), we hypothesized them to shape activations in the higher-order areas in the brain, such as the supplementary motor areas, which are known to be involved in perceptual tasks having an anticipatory component (Schubotz and von Cramon, 2002).

Materials and methods

Participants

Eleven healthy participants (with no neurological, hearing or psychological problems) with formal musical training participated in the study (mean age: 23.2 ± 3.7 SD; 5 females). We chose participants with formal musical training as it has been shown that musicians display stronger neural responses to various musical features in comparison to non-musicians (Pantev et al., 2001; Wong et al., 2007). Five participants were educated in and performed mainly classical music, two musicians were trained in folk and jazz music, and the rest were playing mainly pop/rock music. Four musicians played string instruments, three percussive instruments, two wind instruments, and two keyboard instruments. All participants, except one, were also able to play other instruments along with their main one. These musicians started to play their main instrument on average at 9.1 ± 3.4 SD years of age, and their second instrument at 10.5 ± 3.7 SD years, collecting a total amount of years of training equal, on average, to 16.1 ± 6 SD.

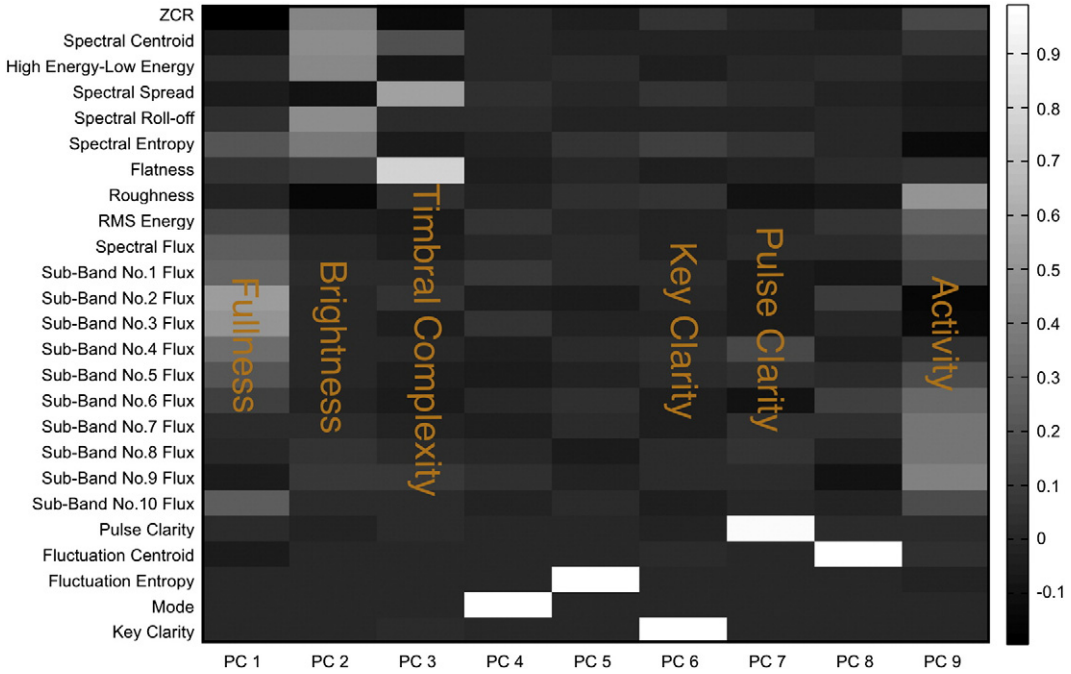

Fig. 1. Loadings of the features on the first 9 Principal Components (PC) as a result of principal component analysis with varimax rotation. The x-axis indicates principal component number in order of decreasing amount of variance explained. The shade of gray indicates the respective loading with white indicating the maximum and black the minimum. PCs 4, 5, and 8 are not labeled as they failed to correlate with the perceptual ratings and thus were excluded from subsequent analyses.

Furthermore, they declared to practice on average 2.5 ± 1.2 SD hours per day at the time of the experiment.

Stimulus

The stimulus used in the experiment was the tango *Adios Nonino* by Astor Piazzolla of a duration of 8 min and 32 s. This piece of music was selected due to its high range of variation in several musical features such as dynamics, timbre, tonality and rhythm, while having an appropriate duration for the experimental setting used.

fMRI measurements

The fMRI measurements were conducted with the 3-T scanner (3.0 T Signa VH/I General Electric) at the Advanced Magnetic Imaging (AMI) Centre of the Aalto University and were approved by the local ethical committee. To prevent postural adjustments and to attenuate the noise and vibration of the scanner, foam cushions were placed around the arms of the participants. Music was presented through audio headphones with about 30 dB of gradient noise attenuation. Further attenuation was achieved with cotton inserted in the headset. Thirty-three oblique slices covering the whole brain (field of view 200 × 200 mm; 64 × 64 matrix; slice thickness 4 mm; gap 0 mm) were acquired using a single-shot gradient echo-planar imaging (EPI) sequence (TR = 2 s; echo time, 32 ms; flip angle, 75°) sensitive to blood oxygenation level-dependent (BOLD) contrast.

During the fMRI measurement, participants listened to the stimulus presented at an average sound level of 80 dB. The participants were instructed to stay still and to relax while listening to the musical stimulus and to maintain their gaze on the screen. Subsequent to a short break after fMRI recording, anatomical T1 weighted MR images (field of view 260 × 260 mm; 256 × 256 matrix; thickness 1 mm; spacing 0 mm) were acquired.

fMRI preprocessing

Whole-brain image analysis was carried out using Statistical Parametric Mapping 5 (SPM5—http://www.fil.ion.ucl.ac.uk/spm). Images for each subject were realigned, spatially normalized into the Montreal Neurological Institute template (12 parameter affine model, gray matter segmentation; realignment: translation components <2 mm, rotation components <2°), and spatially smoothed (Gaussian filter with FWHM of 6 mm). fMRI responses were detrended using a high-pass filter with a cut-off frequency of .008 Hz, which conforms to the standards used to reduce the effects the scanner drift typically occurring at a timescale of 128 s (Smith et al., 1999). Following this, Gaussian smoothing was performed as it provides a good compromise between efficiency and bias (Friston et al., 2000). The smoothing kernel had a width of 5 s, which was found to maximize the correlation between the frequency response of the HRF and the smoothing kernel. The effect of the participants' movements was removed by modeling the 6 movement parameters as regressors of no interest.

Acoustic feature extraction and processing

This section focuses on the computational analyses performed on the audio stimulus. We chose acoustic features that broadly capture the timbral, tonal and rhythmic aspects of the stimulus (see Appendix A). The feature set, comprising twenty-five features, can be generally classified into two categories based on the duration of the analysis-window used during the extraction process, that is, short-term features and long-term features. The short-term features, which encapsulate timbral properties of the stimulus, were obtained by employing short-time analysis using a 25 ms window, which is in the order of the commonly used standard window length in the field of Music Information Retrieval (MIR) (Tzanetakis and Cook, 2002). These short-term features comprise the zero crossing rate, spectral centroid, high energy–low energy ratio, spectral spread, spectral roll-off, spectral entropy, spectral flatness (Wiener entropy), roughness, RMS energy, spectral flux, and Sub-Band Flux (10 coefficients).

The long-term features encapsulate context-dependent aspects of music, such as tonality and rhythm and were calculated using a longer time-window of analysis, that is, 3 s. This window length was chosen because it corresponds to typical estimates of the length of the auditory sensory memory (Fraisse, 1982). The long-term features include pulse clarity, fluctuation centroid, fluctuation entropy, mode and key clarity. All the features were extracted using the MIRToolbox (Lartillot and Toiviainen, 2007) in the MATLAB environment. After obtaining the time series of all the features, to make the data comparable to the fMRI data, we processed the data as discussed below.

First, to account for the lag caused due to the hemodynamic response in the fMRI data, the acoustic feature time series were convolved with a double-gamma HRF. The convolved acoustic feature time series were then filtered employing the detrending filter used in the post-processing stage of the fMRI data. This operation was performed to eliminate those low-frequency components whose eventual brain correlates were eliminated during the preprocessing stage of the fMRI time series. For subsequent analysis we chose to use features of only the part of the stimulus that contained music, therefore we excluded the last 24 s of the stimulus due to the presence of applause and the first 26 s corresponding to the length of the HRF in order to avoid any artifacts caused due to the convolution operation. Following this, all the features were downsampled by retaining one sample for every 160 samples to match the sampling rate of the fMRI data (.5 Hz). As a result, the length of the feature vectors was 231 samples corresponding to 7 min and 42 s.

In order to reduce the number of features in the feature set, we performed a principal component analysis (PCA). The first 9 PCs were included in subsequent analyses as they explained 95% of the variance. Fig. 1 displays the loadings for these 9 PCs.

Table 1
Mean inter-subject correlations, Cronbach alphas and number of participants included in subsequent analyses for each of the perceptual scales.

Perceptual scale	Mean Inter-subject *r*	Cronbach alpha	*N*
Fullness	.28	.87	19
Brightness	.27	.87	20
Timbral complexity	.34	.90	18
Rhythmic complexity	.30	.90	21
Key clarity	.54	.96	21
Pulse clarity	.55	.96	21
Event synchronicity	.25	.85	17
Activity	.61	.97	20
Dissonance	.55	.96	19

Table 2
Pearson's r correlation coefficients between perceptual ratings and respective PC scores.

Perceptual scale	*r*
Fullness	.80***
Brightness	.55**
Timbral complexity	.53**
Rhythmic complexity	.28
Key clarity	.53**
Pulse clarity	.51**
Event synchronicity	.30
Activity	.77***
Dissonance	.30

* $p<.05$. ** $p<.01$. *** $p<.001$.

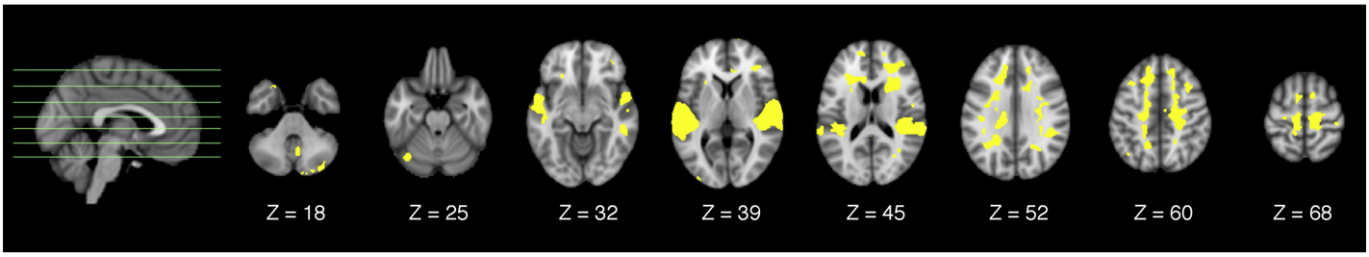

Fig. 2. Selected slices depicting the areas of the brain that displayed significant mean inter-subject correlation ($p<.001$).

As visible in Fig. 1, the first PC has the highest loadings from Subbands No. 2 and No. 3 Flux. The presence of spectral fluctuations in lower bands of the spectrum has been associated previously to perceptual 'Fullness' (Alluri and Toiviainen, 2010). The second PC, with the highest loadings from spectral centroid, spectral roll-off, high energy–low energy ratio and zero crosssing rate measures the perceptual 'Brightness' (Alluri and Toiviainen, 2010, 2012). PC3 represents the spread and flatness (Wiener entropy) of the spectrum. In perceptual terms, it can be regarded to represent timbral complexity. As can further be seen from Fig. 1, each of the PCs from 4 through 8 has high contributions from only one of the long-term features representing 'Mode', 'Rhythmic Complexity', 'Key Clarity', 'Pulse Clarity', and 'Event Synchronicity' respectively (see Appendix A for details). PC 9 has the highest loadings from roughness and flux of the high end of the spectrum. This can be regarded to represent 'Activity' (Alluri and Toiviainen, 2010, 2012). The loudness feature of the stimulus, characterized by RMS energy, failed to load highly onto any of these PCs suggesting that none of the PCs correlated notably with RMS energy.

To validate the perceptual relevance of the labels given to the acoustic components, we conducted a perceptual listening experiment where participants were asked to rate selected segments representing varying levels of the acoustic components on various perceptual scales. The procedure is explained in detail as follows.

Perceptual experiment

Stimuli

We selected nine sets of stimuli comprising 30 6-second segments each, obtained from the stimulus used in the fMRI experiment. For each of the sets, the excerpts were chosen to represent varying levels of one of the acoustic components. The acoustic component representing mode was excluded, because according to initial analysis of the fMRI data it failed to display any significantly correlating areas. Additionally, we wanted to limit the duration of the experiment to avoid fatigue effects. To obtain the excerpts, for each acoustic component, the raw acoustic features were multiplied by the respective PC loadings. Subsequently, we averaged the hence obtained PC score time series within 6-second moving windows using a hop-size of 1 s. Next, each of these averaged time series were rank ordered and sampled ($n=30$) equidistantly in order to obtain segments that represented varying levels of the respective PC scores and spanned the entire range of variation.

Participants

Twenty-one musicians (15 females, age $M=23.7$ years, $SD=3.5$) participated in rating the music excerpts. All reported having formal music education ($M=15.8$ years, $SD=6.1$ years, minimum $=3$ years). They practiced on average 11.2 ($SD=9$) hours per week. Seven reported familiarity with the piece *Adios Nonino*. All of the participants reported listening to music the average being 12 h/week. None of these participants took part in the fMRI experiment. However, the two groups shared similar demographic properties in terms of their ages as well as their musical and cultural backgrounds.

Procedure

The participants were given written instructions, following which the listening experiment took place in a silent room. To present the stimuli and obtain the ratings, a graphical interface was developed in Max/MSP. The experiment was divided into nine sections, in each of which the participants were asked to rate the stimuli according to one of the scales. Each scale was divided into 9 levels with the extremes of the scale indicating low and high values of the perceptual scale (e.g: 1 = Low Pulse Clarity, 9 = High Pulse Clarity), from which the subject could choose the level that best described the music excerpt presented. The participants were able to listen to each excerpt as many times as they wished. The order of presentation of each of the sections and the music excerpts in each section was randomized. Prior to the actual experiment, the participants were allowed to familiarize themselves with the working of the interface.

Results

The perceptual ratings were initially checked for inconsistencies and outliers. First, for each scale, the participants with a negative mean inter-subject correlation were eliminated. Following this, participants with a mean inter-subject correlation two SDs below the overall mean inter-subject correlation were eliminated. As a result, one to four participants were excluded for each scale. As visible in Table 1, Cronbach's alpha revealed high agreement between the participants' ratings, indicating fairly consistent opinions among listeners with respect to the perceptual scales. Therefore, for subsequent analysis, the individual ratings for each scale were averaged across the participants. As can be seen from Table 2, significant correlations were observed between the rating scales of Fullness, Brightness, Timbral Complexity, Key Clarity, Pulse Clarity and Activity and the respective acoustic components. The remaining scales of Rhythmic Complexity and Event Synchronicity failed to correlate significantly with the corresponding acoustic components and were hence excluded from further analyses. Therefore, the temporal evolutions of only the corresponding principal component scores of these six PCs were used for further analysis and will henceforth be referred to as acoustic components.

Statistical analysis

Initially, the inter-subject consistency of the fMRI data was checked using mean inter-subject correlation as a measure. The significance of the correlations was estimated by means of a Monte Carlo approach. Following this, we employed the correlation analysis approach utilized by He et al. (2008). First, for each participant, the Pearson correlation coefficients (r) were obtained per voxel and per acoustic component. The r maps were then converted to respective Z-score maps using Fisher's Z transformation, normalized by the factor $1/\sqrt{(df-3)}$, where df represents the estimated number of effective degrees of freedom. The effective degrees of freedom were estimated using the approach described by Pyper and Peterman (1998) (see Appendix B for details). As a result, the effective degrees of freedom varied from a minimum of 29 (correction factor $=7.97$) to a maximum of 68 (correction factor $=3.40$) across the acoustic components.

For each acoustic component, in order to obtain the group maps we employed the combining tests procedure described by Lazar (2008). In this approach, individual Z-score maps were first converted to p-maps and were then pooled using the Fisher's p-value technique (Fisher, 1950) to create the group maps. The individual p-values of each voxel were pooled using the equation below to obtain a T-statistic (Eq. (1)) that is modeled as a Chi-square distribution with $2k$ degrees of freedom where k represents the number of participants.

$$T = -2\sum_{i=1}^{k} \log p_i. \quad (1)$$

The group maps hence obtained for each component were thresholded at a significance level of $p<.001$. Following this, in order to minimize Type I errors, we corrected for multiple comparisons using cluster size thresholding. To determine the thresholds for multiple comparisons correction, we performed a Monte Carlo simulation of the approach described by Ledberg et al. (1998) (see Appendix B). As result, we obtained a cluster size threshold of 22 voxels for $p<.001$ ($Z=3.29$).

Results

Fig. 2 displays the results of the correlation analysis performed to test the consistency between the participants' fMRI responses. As can be seen, relatively large areas of the brain were found to display

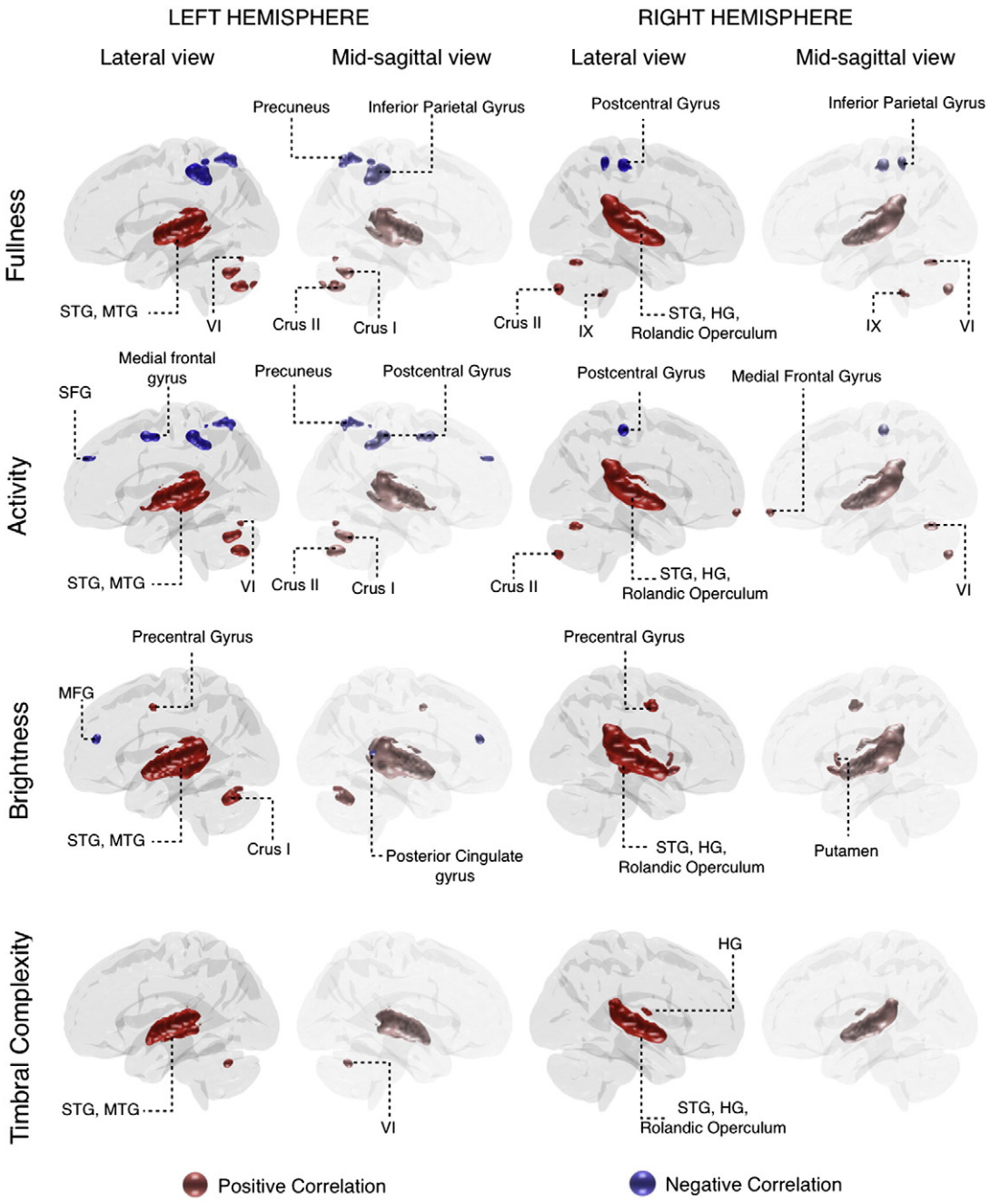

Fig. 3. Lateral and mid-sagittal views of the left and right hemispheres of the brain showing regions correlating significantly with timbral components. The significance threshold for the correlations was set at $p=.001$. Cluster correction was performed at a significance level of $p=.001$ ($Z=3.29$) corresponding to a cluster size of 22 voxels. The exact Z-values are listed in Table 3. The areas indicated in red and blue correspond to the brain areas that correlated positively and negatively to each of the timbral components, respectively.

Table 3
Correlation results between acoustic components and brain activity using a significance threshold p<.001. The clusters were obtained using the 18-connectivity scheme employed in SPM. The significant voxels were cluster corrected at Z>3.29 (p<.001 and cluster size>22). The coordinates are in Talairach space and indicate the location of the global maximum within each cluster. For each cluster, ROI analysis was performed using MarsBar. Each cluster is numbered and named after the ROI containing the voxel with the maximal Z-value. The sub-clusters obtained from the ROI analysis are displayed under each main cluster. The coordinates for the sub-clusters are in Talairach space and indicate the location of the local maximum within each cluster.

Right hemisphere	BA	k	x	Y	z	Z-value	Left hemisphere	BA	k	x	y	z	Z-value
(*A*) *Fullness*													
Positive correlation													
1. Superior temporal gyrus	22	2325	51	−14	1	7.35	5. Superior temporal gyrus	22	2283	−50	−21	3	7.34
Superior temporal gyrus		1646	51	−14	1	7.35	Superior temporal gyrus		1137	−50	−21	3	7.34
Heschl's gyrus		203	50	−15	4	6.82	Heschl's gyrus		80	−53	−15	8	5.20
Insula		17	50	−4	0	4.70	Middle temporal gyrus		890	−50	−19	1	6.94
Postcentral gyrus		8	63	−15	14	3.78							
Rolandic operculum		193	40	−28	16	6.13	Rolandic operculum		64	−44	−28	14	6.23
Supramarginal gyrus		100	50	−34	24	4.96	Supramarginal gyrus		13	−51	−25	14	4.71
Superior temporal pole		20	53	2	−5	4.38							
Cerebellum													
2. Inferior semi-lunar lobule		76	30	−78	−35	4.25							
VIIB		1	26	−78	−36	3.58							
Crus II		75	30	−78	−35	4.25							
3. Declive		39	32	−63	−17	3.76	6. Declive		39	−14	−73	−13	3.75
VI		39	32	−63	−17	3.76	VI		39	−14	−73	−13	3.75
4. Tonsil		40	6	−45	−38	3.60							
IX		23	16	−47	−40	3.41							
							7. Uvula		79	−28	−63	−24	4.33
							VI		28	−28	−61	−24	4.18
							Crus I		50	−28	−63	−24	4.33
							VIIB		22	−8	−71	−30	3.95
							VIII		2	−8	−69	−30	3.50
							8. Pyramis		66	−28	−68	−34	4.10
							VIIB		10	−28	−68	−34	4.10
							VIII		2	−26	−66	−34	3.45
							Crus II		54	−30	−70	−34	4.09
							9. Pyramis		23	−20	−83	−31	4.01
							Crus II		23	−20	−83	−31	4.01
Negative correlation													
1. Postcentral gyrus	2	67	44	−23	49	3.66	3. Postcentral gyrus	3	191	−36	−31	48	4.19
Postcentral gyrus		67	44	−23	49	3.66	Postcentral gyrus		83	−36	−31	48	4.19
2. Inferior parietal lobule	40	31	40	−38	55	3.81	Inferior parietal gyrus		95	−32	−41	43	3.70
Inferior parietal gyrus		12	40	−38	52	3.41							
Superior parietal gyrus		9	40	−38	53	3.66							
Postcentral gyrus		10	40	−38	55	3.81							
							4. Middle frontal gyrus	11	27	−26	38	−19	4.09
							Inferior orbitofrontal gyrus		11	−30	42	−17	3.54
							5. Superior parietal lobule	7	165	−18	−61	56	3.71
							Superior parietal gyrus		75	−18	−61	56	3.71
							Postcentral gyrus		18	−24	−42	56	3.58
							Precuneus		72	−16	−61	58	3.53
(*B*) *Brightness*													
Positive correlation													
1. Superior temporal gyrus	22	3166	50	−19	3	8.13	6. Superior temporal gyrus	22	3245	−55	−15	3	8.13
Superior temporal gyrus		2070	50	−19	3	8.13	Superior temporal gyrus		1516	−55	−15	3	8.13
Heschl's gyrus		225	48	−15	4	8.13	Heschl's gyrus		153	−53	−15	8	7.73
Insula		51	50	−4	0	6.16	Insula		13	−46	−8	−1	4.50
Postcentral gyrus		27	63	−15	14	5.01	Postcentral gyrus		3	−63	−21	16	3.41
Rolandic operculum		274	40	−28	16	6.80	Rolandic operculum		109	−46	−28	14	6.89
Supramarginal gyrus		99	48	−34	24	4.95	Supramarginal gyrus		16	−63	−23	14	5.37
Middle temporal gyrus		90	46	−33	3	4.17	Middle temporal gyrus		1192	−59	−12	−3	8.13
Temporal pole		74	51	2	−7	5.90	Temporal pole		11	−57	5	−9	4.40
2. Precentral Gyrus	6	64	55	0	41	4.05	7. Precentral gyrus	6	28	−55	0	42	4.39
Precentral gyrus		53	55	0	41	4.05	Precentral gyrus		17	−55	0	42	4.39
Middle frontal gyrus		11	51	−1	50	3.69	Postcentral gyrus		10	−55	0	41	4.21
3. Putamen		38	24	11	−6	3.63							
Putamen		36	24	11	−6	3.63							
4. Putamen		30	14	10	1	3.38							
Putamen/pallidum		7	14	6	−4	3.33							
Putamen		10	22	12	5	3.26							
Cerebellum													
5. Declive		186	−32	−63	−22	4.34							
VI		102	−28	−59	−22	4.28							
CrusI		84	−32	−63	−22	4.34							
Negative correlation													
							1. Medial frontal gyrus	9	26	−20	42	15	3.68
							Middle frontal gyrus		7	−20	43	14	3.63
							Superior frontal gyrus		2	−20	42	18	3.31
							2. Posterior cingulate	29	26	−6	−44	10	3.50
							Calcrine		4	−6	−44	8	3.44

Table 3 (*continued*)

Right hemisphere	BA	k	x	Y	z	Z-value	Left hemisphere	BA	k	x	y	z	Z-value
							Posterior cingulate		5	−6	−40	8	3.36
							Precuneus		9	−6	−44	10	3.50
(*C*) *Activity*													
Positive correlation													
1. Superior temporal gyrus	22	2368	51	−14	1	7.92	5. Superior temporal gyrus	22	2378	−50	−21	3	7.48
Superior temporal gyrus		1632	51	−14	1	7.92	Superior temporal gyrus		1185	−50	−21	3	7.48
Heschl's gyrus		211	50	−15	4	7.31	Heschl's gyrus		91	−53	−15	8	5.58
Rolandic operculum		209	40	−28	16	6.26	Rolandic operculum		73	−44	−28	14	5.87
Supramarginal gyrus		101	50	−34	24	4.63	Supramarginal gyrus		11	−53	−25	14	4.36
Temporal pole		24	51	2	−5	4.67	Temporal pole		15	−57	7	−9	3.78
Insula		24	50	−4	0	5.45	Middle temporal gyrus		893	−50	−19	1	7.09
Postcentral gyrus		8	63	−15	14	3.87							
2. Medial frontal gyrus	11	22	8	61	−15	3.69							
Medial frontal gyrus		22	8	61	−15	3.69							
Cerebellum													
3. Inferior semi-lunar lobule		70	30	−79	−35	4.42	6. Uvula		91	−30	−63	−24	4.13
VIIB		1	26	−78	−36	3.60	VI		19	−28	−61	−24	3.96
CrusII		69	30	−79	−35	4.42	CrusI		72	−30	−63	−24	4.13
4. Declive		50	32	−63	−15	3.89	7. Pyramis		106	−32	−70	−32	4.11
VI		50	32	−63	−15	3.89	VIIB		8	−28	−68	−34	3.86
							VIII		1	−26	−66	−34	3.40
							CrusII		97	−32	−70	−32	4.11
							8. Declive		57	−16	−73	−13	3.91
							VI		57	−16	−73	−13	3.91
Negative correlation													
1. Postcentral gyrus	3	46	42	−25	49	3.51	2. Postcentral gyrus	3	137	−36	−31	48	4.26
Postcentral gyrus		46	42	−25	49	3.51	Postcentral gyrus		72	−36	−31	48	4.26
							Inferior parietal lobule		56	−38	−35	44	3.64
							3. Superior parietal lobule	7	105	−18	−53	58	3.61
							Superior parietal lobule		46	−18	−53	58	3.61
							Postcentral gyrus		9	−24	−42	56	3.43
							Precuneus		50	−16	−53	58	3.47
							4. Middle frontal gyrus	6	62	−32	8	46	3.63
							Middle frontal gyrus		24	−32	8	46	3.63
							Precentral gyrus		38	−32	6	44	3.60
							5. Superior frontal gyrus	9	24	−20	52	25	3.48
							Superior frontal gyrus		24	−20	52	25	3.48
							Middle frontal gyrus		19	−22	52	27	3.48
(*D*) *Timbral complexity*													
Positive correlation													
1. Superior temporal gyrus	22	1804	53	−8	−1	7.05	2. Superior temporal gyrus	21	1787	−55	−12	−1	6.93
Superior temporal gyrus		1418	53	−8	−1	7.05	Superior temporal gyrus		973	−55	−12	−1	6.93
Heschl's gyrus		163	51	−13	4	6.41	Heschl's gyrus		64	−50	−17	8	5.06
Rolandic operculum		104	65	−5	8	4.83	Rolandic operculum		22	−42	−28	14	4.20
Middle temporal gyrus		4	69	−25	0	3.71	Middle temporal gyrus		659	−57	−14	−1	6.78
Insula		9	50	−4	0	4.50							
Postcentral gyrus		9	65	−11	13	3.66							
Temporal pole		43	53	2	−7	4.90							
Cerebellum													
							3. Declive		27	−26	−61	−22	4.19
							VI		26	−26	−61	−22	4.19
							CrusI		1	−30	−63	−22	3.13
(*E*) *Key clarity*													
Negative correlation													
1. Precentral gyrus	6	306	59	−4	28	4.67							
Postcentral gyrus		256	59	−4	28	4.67							
Precentral gyrus		44	63	3	22	4.09							
Rolandic operculum		1	61	−3	13	3.25							
Supramarginal gyrus		1	55	−12	26	3.31							
2. Postcentral gyrus	3	82	48	−15	58	4.29	6. Postcentral gyrus	3	142	−44	−18	58	4.12
Postcentral gyrus		24	50	−15	56	4.14	Postcentral gyrus		69	−44	−18	58	4.12
Precentral gyrus		58	48	−15	58	4.29	Precentral gyrus		72	−42	−17	58	4.03
3. Postcentral gyrus	43	80	50	−14	21	4.29	7. Postcentral gyrus	43	86	−46	−9	17	4.23
Insula		15	34	−17	17	4.29	Heschl's gyrus		4	−46	−11	10	3.39
Rolandic operculum		62	50	−14	21	4.29	Postcentral gyrus		31	−53	−13	21	3.67
							Rolandic operculum		45	−46	−9	17	4.23
4. Superior frontal gyrus	6	29	12	−12	67	3.83	8. Superior frontal gyrus	9	57	−8	52	31	3.82
Superior frontal gyrus		3	14	−12	67	3.23	Superior frontal gyrus		31	−10	52	31	3.74
Paracentral lobule		1	8	−18	69	3.13	Medial frontal gyrus		26	−8	52	31	3.82
Precentral gyrus		3	14	−14	67	3.52							
Supplementary motor area		22	12	−12	67	3.83							
							9. Claustrum		22	−34	−11	4	3.79
							Insula		12	−36	−13	4	3.68

(*continued on next page*)

Table 3 (*continued*)

Right hemisphere	BA	k	x	Y	z	Z-value	Left hemisphere	BA	k	x	y	z	Z-value
							Putamen		1	−32	−11	4	3.15
5. Cingulate gyrus	24	50	14	−6	44	3.66							
Middle cingulate gyrus		27	14	−6	44	3.66							
Supplementary motor area		22	14	−4	44	3.54							
(*F*) *Pulse clarity*													
Positive correlation													
1. Superior temporal gyrus	22	173	51	−8	0	4.30	3. Superior temporal gyrus	22	97	−61	−17	3	3.73
Superior temporal gyrus		153	51	−8	0	4.30	Superior temporal gyrus		74	−61	−17	3	3.73
Heschl's gyrus		19	53	−10	4	3.73	Middle temporal gyrus		23	−55	−16	1	3.63
Insula		1	50	−4	0	3.11							
2. Superior temporal gyrus	41	26	40	−30	16	3.37							
Heschl's gyrus		2	38	−30	16	3.23							
Rolandic operculum		3	40	−30	18	3.26							
Superior temporal gyrus		21	40	−30	16	3.37							
Negative correlation													
							5. Claustrum		78	−32	−8	−8	4.15
							Amygdala		5	−28	−8	−10	3.91
							Hippocampus		7	−30	−10	−10	3.76
							Insula		6	−36	−8	−8	3.42
							Putamen		11	−30	−10	−6	3.75
1. Cingulate gyrus	32	50	2	6	40	3.72	6. Cingulate gyrus	32	38	−2	6	40	4.01
Cingulum mid		40	2	6	40	3.72	Cingulum mid		18	−2	6	40	4.01
Supplementary motor area		10	4	6	42	3.40	Supplementary motor area		20	−2	8	40	3.94
2. Precuneus	7	23	12	−48	59	3.58							
Precuneus		15	12	−48	59	3.58							
Superior parietal lobule		8	14	−48	59	3.42							
3. Insula	13	22	38	10	9	3.46	7. Insula	13	28	−38	12	5	3.57
Insula		13	38	10	9	3.46	Insula		28	−38	12	5	3.57
Inferior frontal operculum		9	40	10	11	3.44							
4. Inferior temporal gyrus		97	48	−48	−18	4.10							
Fusiform gyrus		7	44	−47	−16	3.41							
Inferior temporal gyrus		90	48	−48	−18	4.10							

significant mean inter-subject correlations with the maximum found in the auditory cortices ($r = .64$, $p < .0001$). Following this, we performed correlation analysis between the fMRI data and acoustic components. As described in Perceptual experiment, perceptual validation of the acoustic components resulted in a compact set of six acoustic components that represented the temporal evolution of the main timbral, tonal, and rhythmic features in the stimulus. Timbral components comprised perceived Fullness, Activity, Brightness, and Timbral complexity of the stimulus. Tonal and Rhythmic components comprised Key Clarity and Pulse Clarity respectively (audio excerpts depicting examples of low levels and high levels of all six acoustic components can be found as Supplementary material: AudiomaterialS1). First-level analysis comprised correlating these six components with fMRI time series at an individual level. Following this, second-level analysis involved pooling individual results to obtain group maps for each acoustic component. The results are discussed below.

Timbral feature processing in the brain

Correlation analyses revealed that the presence of high values in all the timbral features namely Fullness, Brightness, Timbral Complexity and Activity, was associated with increased neuronal activation in the bilateral STG (BA 22) (see Fig. 3). Additionally, a lateralization effect was found wherein the right hemisphere displayed positive correlations with timbral components in larger proportions of the HG, rolandic operculum, supramarginal gyrus, and superior temporal pole than the left hemisphere (see Table 3). In contrast, the left hemispheric middle temporal gyrus had a larger proportion displaying such a correlation than its right hemispheric counterpart (see Table 3 for details).

Another group of activations was found in the cerebellum. In particular, high values of Fullness and Activity in the stimulus were associated with increased activation in the declive, uvula, and pyramis (lobule VI, Crus I and II). Increase in the timbral components of Brightness and Timbral Complexity was associated with increased activation in the declive only.

Outside of the auditory cortex and the cerebellum, we observed cerebrocortical areas correlating negatively with the timbral components. In particular, decreased Activity and Fullness was found to be associated with increased activations in the bilateral postcentral gyrus (BA 2 and 3), and the left precuneus (BA 7). In addition, low levels of Fullness were associated with increased activations in the bilateral inferior parietal gyrus (BA 40), and those of Activity were associated with increased activations in the left superior frontal gyrus (BA 9) and left medial frontal gyrus (BA 6). Increased activations in the right medial frontal gyrus (BA 11) were found to be associated with increasing Activity.

Furthermore, increase in Brightness recruited the bilateral precentral gyrus (BA 6), and the right putamen. Reduced levels of Brightness in the stimulus, on the other hand were associated with increased activations in two left hemispheric clusters of the medial frontal gyrus (BA 9) and the posterior cingulate cortex (BA 29). No negative correlations were found for Timbral Complexity.

Rhythmic and tonal feature processing in the brain

High levels of Pulse Clarity were linked with increased activation in the bilateral STG (BA 22), as well as the right primary auditory cortex (BA 41) (see Fig. 4 and Table 3). In contrast, decreased Pulse Clarity recruited the right inferior temporal gyrus (ITG, BA 37) and precuneus (BA 7). In addition, decreased Pulse Clarity was associated with high levels of activation in several subcortical limbic areas including the left hemispheric amygdala, hippocampus and putamen, the bilateral mid-cingulate gyrus (BA 32) in the vicinity of the supplementary motor area, and the bilateral insula (BA 13).

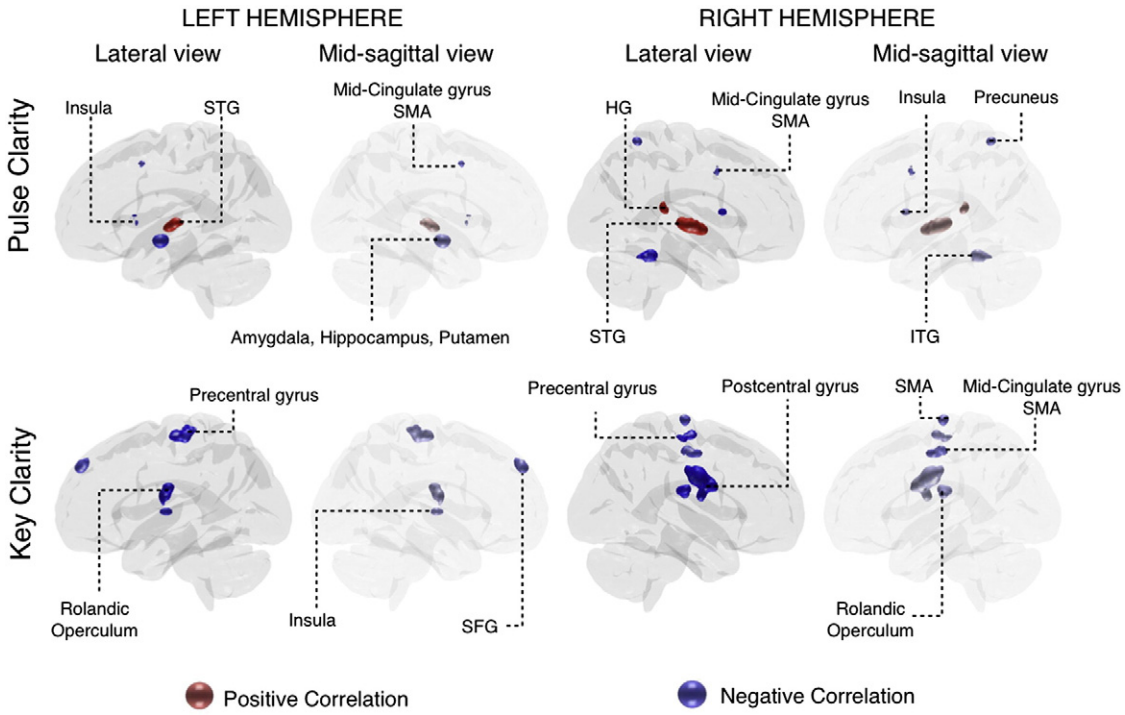

Fig. 4. Lateral and mid-sagittal views of the left and right hemispheres of the brain showing regions correlating significantly with Pulse Clarity and Key Clarity components. The significance threshold for the correlations was set at $p = .001$. Cluster correction was performed at a significance level of $p = .001$ ($Z = 3.29$) corresponding to a cluster size of 22 voxels. The exact Z-values are listed in Table 3. The areas indicated in red and blue correspond to the brain areas that correlated positively and negatively to each of the components, respectively.

The presence of unclear key (i.e., low levels of Key Clarity) was found to be associated with increased activation in cortical and subcortical areas such as parts of the bilateral precentral gyrus (BA 3), the right mid-cingulate gyrus (BA 24) in the vicinity of the supplementary motor area and right postcentral (BA 6) gyrus, and the left hemispheric superior frontal gyrus (BA 9), the left insula and the bilateral rolandic operculum (see Fig. 4). No positive correlations were observed for Key Clarity.

Discussion

In the present study, we investigated the neural correlates of timbral, tonal, and rhythmic feature processing of a naturalistic music stimulus. To this end we employed a novel paradigm combining neuroimaging, computational acoustic feature extraction and behavioral psychology. Participants were scanned using fMRI while they freely listened to the musical piece *Adios Nonino* by Astor Piazzolla. First, inter-subject consistency on a voxel-by-voxel basis was evaluated using mean inter-subject correlation as a measure. Following this, the evolution of musical features in the piece was obtained using sophisticated acoustic feature extraction procedures. Based on a perceptual test we selected a set of six acoustic components representing the main timbral, tonal, and rhythmic features present in this piece. Following this, the neural underpinnings of these acoustic components were investigated by correlating their time series with the time series of the BOLD signal. Overall, our results highlighted the brain structures responsible for the processing of an extensive set of timbral, tonal and rhythmic features. The results corroborate findings reported in previous neuroimaging studies, which have used artificial and acoustically reduced conditions to investigate musical feature processing. Moreover, they also highlight additional brain structures involved in musical feature processing. Timbral features activated mainly perceptual and resting-state or default mode areas of the cerebrum and cognitive areas of the cerebellum. In contrast, for tonal and rhythmic features, we observed for the first time during listening to a naturalistic stimulus, activations in subcortical emotion-related areas along with activations in cognitive and somatomotor cerebrocortical areas.

Timbre-related acoustic components correlated positively with activations in large areas of the temporal lobe (STG, HG, and MTG) (see Fig. 3). These areas were mostly activated during moments of high Fullness, Activity, Brightness and Timbral Complexity, which were often associated with quicker passages in the stimulus with several pitches usually played by several instruments. While there exists a dearth of studies regarding neural correlates of timbre processing in a musical context, or polyphonic timbre processing, evidence from neural studies on monophonic timbre has repeatedly pointed at the involvement of the bilateral STG and HG (Caclin et al., 2006; Halpern et al., 2004). The current findings suggest that the same brain areas are recruited in polyphonic timbre processing. As hypothesized based on Samson et al. (2011), we found the caudolateral and anteriolateral parts of the STG, specifically in the right hemisphere to be involved in timbral feature processing. Furthermore, this finding also supports our hypothesis of interhemispheric specialization in the auditory cortices with regard to the processing of timbral features, with the right temporal lobe displaying larger areas with significant correlations with these features.

Negative correlations between timbral features of Activity and Fullness and brain activity were observed in cerebrocortical regions in the vicinity of the left superior frontal gyrus (BA 9), the left precuneus and surrounding parietal areas, and the ventral medial prefrontal cortex. These areas are known to be part of the default mode network (DMN). The DMN is a neural circuit constantly monitoring the sensory environment and displaying high activity during lack of focused attention on external events (Fox et al., 2009; McAvoy et al., 2008). As low values in Activity and Fullness were mostly associated with sections in the stimulus with sparse texture played by the piano, thereby resulting in lower levels of auditory–cognitive load, the activation of the DMN during these moments is in line with previous results (Levitin and Menon, 2003; Pallesen et al., 2009; Uddin et al., 2009). In the visual modality, a network comprising several parietal areas such as the precuneus and the supramarginal gyrus

has been found to be highly activated with reduced visual attention or load (Hahn et al., 2006). Similarly, the left posterior cingulate, one of the central structures of the DMN, was observed to be deactivated during moments in the stimulus with high Brightness, which were associated with the presence of several instruments playing simultaneously. This finding appears to be in line with those discussed by Levitin and Menon (2003) regarding the deactivations in the posterior cingulate cortex in response to taxing auditory and visual cognitive tasks. Activations in the right putamen and the bilateral precentral gyrus as well as the right medial frontal gyrus, which appear to be related to movement (Grahn and Rowe, 2009), were found to correlate positively with Brightness and Activity, respectively. Uddin et al. (2009) showed that these same motor-control areas are functionally connected to the anticorrelated DMN associated with the posterior cingulate cortex. Further research, however, is called for to clarify the link between Brightness and motor-related brain activity.

In addition to the DMN, negative correlations for timbral features of Activity and Fullness were observed in the somatosensory areas (BA 2, 3 and 40). Previous studies investigating the mirror neuron system (Keysers et al., 2010; Koelsch et al., 2006) found that the primary somatosensory (BA 2) and secondary somatosensory areas were recruited while hearing sounds resulting from other people's actions. As nine out of the eleven participants reported playing the piano as their primary or secondary instrument, one could postulate that listening to the piano parts may have activated somatosensory areas that would be active during actual piano playing.

Interestingly, areas of the cerebellum, including lobule VI, Crus I and II, were found to be involved in processing timbre-related acoustic components. According to a meta-analysis by Stoodley and Schmahmann (2009), these areas are known to be the cognitive regions of the cerebellum. Additionally, the involvement of Crus I and II in conjunction with the superior parietal lobule, lateral prefrontal cortex, and dorsal premotor cortex, in processing high cognitive load in an auditory task with chords was found by Salmi et al. (2010). As mentioned earlier, high levels of Fullness and Activity in the stimulus were associated with high levels of auditory-related cognitive load. Hence, our results for the first time demonstrate the role of the cerebellum in cognitive processing while listening to a naturalistic music stimulus.

Previously, the scarce attempts at identifying the neural structures responsible for tonality processing have utilized chord cadences especially composed or simple and repetitive melody lines (Janata et al., 2002a, 2002b). In the present study, this implicit complex cognitive skill has for the first time been investigated using continuous expressive real music. Janata et al. (2002a, 2002b) identified the rostromedial prefrontal cortex as a possible brain substrate for tonal encoding. While, contrary to our expectations, we failed to find significant correlations either in that area or the hippocampus, we observed that the time series for key clarity negatively correlated with activations in the superior frontal gyrus (BA 9), previously related to beauty judgments of sound patterns (Kornysheva et al., 2010), and several brain areas related to somatomotor processing, particularly concentrated in BA 6 and 3 (the precentral and postcentral gyri and the supplementary motor area) (see Fig. 4).

Importantly, we further found that decreasing clarity in the key of the musical stimulus activated inner encephalic structures related to emotion processing such as the claustrum and anterior cingulate (Etkin et al., 2010). Among the auditory areas, only the left HG and the bilateral rolandic operculum were activated during processing music with unclear key. The rolandic operculum of the premotor cortex is known to play an important role in speech articulation and phonological rehearsal (Brown et al., 2005). Previous studies have further revealed that the rolandic operculum in conjunction with the insular cortex is known to play an important role in overt and covert singing (Jeffries et al., 2003). This activation, coupled with the ones in the precentral and postcentral gyri, support the link between spontaneous singing and key processing. However, further investigations are required to ascertain this link. Koelsch et al. (2006) in their fMRI study contend that the rolandic operculum in conjunction with the anterior superior insula, and ventral striatum are recruited in non-musicians while processing pleasant music, in contrast to unpleasant highly atonal counterparts (which were created electronically by pitch-shifting thereby causing them to sound very unnatural or far from 'real world' music). Based on our findings we could postulate that the participants found the tonally less clear parts of the stimulus to be more pleasant than the tonally more clear parts. This finding could be explained in light of the theory advocated by Meyer (1956) according to which one important device for eliciting pleasurable emotions in music listening is the tension evoked by the violation of expectations, such as when a key is unclear due to the presence of complex harmonies (Sloboda, 1991). However, more studies with an ecological listening setting are needed to clarify the relation between key clarity and emotion induction related to violated expectation. Furthermore, the network of these brain areas in conjunction with activations in the medial frontal brain regions (BA 9), has been previously postulated to be involved in the processing of affect-related evaluation and classification in music (Khalfa et al., 2005) and with internal self-monitoring and evaluative processing, especially in an aesthetic context. For instance, activity in this region was observed when asking subjects to judge the beauty of a painting or a black-and-white abstract shape or even a rhythmic drum sequence (Jacobsen et al., 2006; Kornysheva et al., 2010). These results further hint at a link between key processing and aesthetic appreciation.

In line with our hypothesis, for the rhythmic component, Pulse Clarity, we found correlations in the auditory cortices (BA 22 and 41), the motor cortex, basal ganglia structures (putamen), and several regions of the limbic system (cingulate gyrus, insula) and the right ITG (BA 37). While the activations in the areas of the auditory cortex displayed positive correlations, the remaining areas correlated negatively.

The negative correlation of Pulse Clarity with the putamen indicates that it was highly activated during segments of the music with unclear pulse, suggesting the participants were internally generating the pulse during those moments (Grahn and Rowe, 2009). Importantly, the activations in interior brain structures belonging to the limbic system in the vicinity of the amygdala, and the middle cingulate gyrus in the vicinity of the supplementary motor area, have not been observed to be active in previous studies where pulse was artificially manipulated or where participants were asked to tap and extract the pulse within complex rhythmic patterns. As Low Pulse Clarity implies low temporal predictability, the present finding is in line with that of Engel and Keller (2011), who report similar activations when comparing temporally unpredictable improvizations with their more predictable imitations. Blood and Zatorre (2001) found that limbic and reward areas are recruited during music-induced chills. In addition, they emphasize the similarity of these results to those obtained from euphoria-related and pleasant emotion-related brain imaging studies. Moreover, recent evidence suggests that these inner brain structures are recruited especially during intense pleasurable sensations to music (Salimpoor et al., 2011). In light of these findings, one could postulate that lack of clarity in the perceived pulse causes tension, which could be pleasurable thereby resulting in the activation of reward circuits in their brains. The present study thus demonstrates for the first time a neural link between rhythm perception and activity of brain structures associated with emotions in a realistic listening condition.

Interestingly, we found that the right ITG (BA 37) to be activated during moments of Low Pulse Clarity. While the ITG (BA 37) has been associated previously with visual processing, Levitin and Menon (2003) found the left ITG to be more activated during scrambled music listening conditions in comparison to their unscrambled counterparts. These findings hint at the possible role of the ITG in

processing predictability of temporal order. However, this hypothesis is open to question and requires further investigation.

With regard to inter-subject consistency, one could expect the voxels that correlate significantly with the acoustic components to display overall significant mean inter-subject correlations. Indeed this was found to be true (see Appendix Table C.1). In particular the timbral and rhythmic components displayed on average significant mean inter-subject correlations. However, this was not true for Key Clarity. This finding could be attributed to the fact that key can be considered a high level musical concept, the processing of which is based on schematic knowledge that emerges as a result of prior exposure to music (Krumhansl, 1990). Consequently, one could assume the presence of relatively large inter-individual variability in the processing of Key Clarity (see Brattico, 2011; Brattico and Jacobsen, 2009).

Conclusions

To sum up, the current study introduced a new paradigm to investigate and predict the neural mechanisms related to the processing of timbral, tonal, and rhythmic features while listening to a naturalistic stimulus. A notable result of this novel naturalistic approach employed is that, in addition to corroborating findings from previous controlled settings, it revealed additional brain areas involved in music feature processing. First, cognitive areas of the cerebellum as well as sensory and DNM-related cortical areas were found to be involved in timbral feature processing. Second, the results demonstrate the recruitment of limbic and reward areas in the processing of musical pulse. Finally, processing of tonality was found to involve cognitive and emotion-related regions of the brain. These findings advocate the use of more ecological paradigms in future studies in order to obtain a more comprehensive picture of music processing in the brain.

As can be seen in Fig. 3, there exists considerable overlap in the areas of the brain that correlate significantly with the timbral components, especially Activity and Fullness. This overlap can be attributed to the covariance in the acoustic components representing timbre (see Appendix Table C.2). Covariance in acoustic features is inevitable especially in naturalistic stimulus. Nevertheless, it is noteworthy that there hardly exists overlap between timbral, rhythmic and tonal components.

In addition to the aforementioned areas, high inter-subject consistency was observed in some areas that failed to correlate significantly with the present set of acoustic components. This finding suggests that additional acoustic components are needed to account for these stimulus-related activations. This calls for an expansion of the acoustic feature set to obtain a more comprehensive picture of processing of the musical elements in the brain. A potential confounding factor in the present paradigm is the covariance between the acoustic components. The presence of between-component covariance would render it more demanding to tease out the unique individual contributions of the components to evoked activations. However the correlation between the acoustic components representing timbral, rhythmic and tonal aspects was found to be at most moderate suggesting that they represented mutually relatively independent musical aspects of the stimulus. Conducting further experiments with larger sets of naturalistic stimuli representing, for instance, different genres, as well as with musically untrained listeners, would allow for generalizations concerning the brain networks involved in processing musical features in real time. In addition, the results obtained in the present study using the data-driven paradigm could serve as a basis for future hypothesis-driven studies.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.11.019.

Acknowledgments

The authors wish to thank Geoff Luck for his help. This research was supported by the Academy of Finland (project numbers 7118616, 130412 and 138145) and by the aivoAALTO project of the Aalto University.

Appendix A

Acoustic features

The features were extracted from the stimulus on a frame-by-frame basis (see Alluri and Toiviainen (2010) for more details). A window length of 25 ms with a 50% overlap was used to extract the short-term (timbral) features and a frame size of 3 s with a 33% overlap was used to extract the long-term (tonal and rhythmic) features.

A brief description of each of the acoustic features is presented below. A detailed explanation can be found in the user manual of the MIRToolbox (Lartillot and Toiviainen, 2007).

Loudness

Root Mean Square Energy: measure of instantaneous energy contained in the signal, obtained by taking the square root of sum of the squares of the amplitude.

Timbral features

Zero crossing rate (ZCR): number of time-domain zero crossings of the signal per time unit.

Spectral centroid: geometric center on the frequency scale of the amplitude spectrum.

High energy–low energy ratio: ratio of energy content below and above 1500 Hz.

Spectral entropy: the relative Shannon entropy (1948) calculated using the equation Eq. (A.1)

$$H_t = -\frac{\sum_{n=1}^{N} A_t[n] \log A_t[n]}{\log N} \tag{A.1}$$

where A_t is the amplitude spectrum of audio frame at time t and N is the number of frequency bins in the amplitude spectrum. The relative Shannon entropy indicates whether the spectrum contains predominant peaks or not. For example, a single sine tone has minimal entropy and white noise maximal.

Spectral roll-off: frequency below which 85% of the total energy exists.

Spectral flux: measure of temporal change in the spectrum, obtained by calculating the Euclidian distance between subsequent window-based amplitude spectra.

Spectral spread: standard deviation of the spectrum.

Spectral flatness: Wiener entropy of the spectrum, defined as the ratio of its geometric mean to its arithmetic mean.

Sub-Band Flux (10 features in total): measure of fluctuation of frequency content in ten octave-scaled sub-bands of the spectrum (Alluri and Toiviainen, 2010).

Roughness: estimate of sensory dissonance (Sethares, 1998).

Tonal features

Mode: strength of major of minor mode (Saari et al., in press).

Key clarity: measure of the tonal clarity (Gómez, 2006; Krumhansl, 1990; Saari et al., in press).

Rhythmic features

Fluctuation centroid: geometric mean of the fluctuation spectrum representing the global repartition of rhythm periodicities within the

range of 0–10 Hz, indicating the average frequency of these periodicities (Pampalk et al., 2002).

Fluctuation entropy: Shannon entropy of the fluctuation spectrum (Pampalk et al., 2002) representing the global repartition of rhythm periodicities. Fluctuation entropy is a measure of the noisiness of the fluctuation spectrum. For example, a noisy fluctuation spectrum can be indicative of several co-existing rhythms of different periodicities, thereby indicating a high level of rhythmic complexity.

Pulse clarity: estimate of clarity of the pulse (Lartillot et al., 2008).

Appendix B

Estimation of effective degrees of freedom

Due to the presence of serial correlation, the number of effective degrees of freedom in the fMRI and acoustic component time series is less than the total number of sample points. To estimate the effective degrees of freedom, we employed an approach similar to that adopted by Pyper and Peterman (1998). To this end, we performed a Monte Carlo simulation using the approximation shown in the Eq. (B.1) to estimate the effective degrees of freedom (df).

$$\frac{1}{df} \approx \frac{1}{N} + \frac{2}{N}\sum_{j}\frac{N-j}{N}\rho_{XX}(j)\rho_{YY}(j) \quad \text{B.1}$$

where $\rho_{XX}(j)$ is the normalized autocorrelation of the signal of N observations at lag j. The maximal lag j chosen was $N/5$ because it is known to yield relatively accurate results in terms of error rates (Pyper and Peterman, 1998).

Estimation of cluster size threshold to correct for multiple comparisons

To obtain a cluster size threshold for multiple comparisons, we used the approach proposed by Ledberg et al. (1998). This method aims at finding an estimation of the distribution of cluster sizes from which one can estimate the cluster size threshold to be used to correct for multiple comparisons at any particular significance level. In this approach, first, pseudo noise-Statistical Images (pn-SIs), which contain the same spatial spectral properties as the signal-Statistical Image (signal-SI) but contain no stimulus-dependent activations, are used to obtain an estimate of the autocorrelation function (ACF) kernel K. The pseudo noise-SI is created as the Z-score statistic image obtained by randomly choosing a subject's data, correlating it with the reversed and circularly time-shifted acoustic component time series. Then, the ACF kernel K is determined according to Eq. (B.2).

$$K = IFT|FFT(P)| \quad \text{B.2}$$

where P represents the pseudo noise-Statistical Image. To account for the variance in the estimate of the ACF, K is determined as an average of all the ACF kernels obtained via a Monte Carlo simulation. In order to create a Noise-Statistical Image (noise-SI), a randomly generated image U is convolved with the ACF kernel, K, using Eq. (B.3)

$$SI = U * K \quad \text{B.3}$$

To estimate the distribution of the cluster sizes we generated 1000 Noise-SIs using the equation above. As a result, we obtained a cluster size threshold of 22 voxels for $p<.001$ ($Z=3.29$).

Appendix C

Appendix Table C.1
Summary of mean inter-subject correlations within the areas correlating significantly with each of the acoustic components. Columns two and three indicate the means and standard deviations of the mean inter-subject correlation values across the voxels displaying significant correlations with each of the acoustic components. Columns four through six indicate, for the voxels correlating significantly with each acoustic component, the percentage of voxels with significant mean inter-subject correlation (evaluated at $p<.05$, $p<.01$, $p<.001$).

	Inter-subject correlation		Overlap		
	Mean	STD	$p<.05$	$p<.01$	$p<.001$
Fullness	.23***	.14	86%	82%	78%
Brightness	.22***	.13	92%	86%	79%
Timbral complexity	.30***	.12	99%	97%	95%
Key clarity	.05	.03	36%	17%	3%
Pulse clarity	.24***	.24	49%	47%	47%
Activity	.23***	.14	88%	84%	80%

*** $p<.001$.

Appendix Table C.2
Pearson's correlation coefficients (r) between principal component scores of the six perceptually validated acoustic components. The significance values have been calculated based on the effective degrees of freedom.

	Fullness	Brightness	Timbral Complexity	Key Clarity	Pulse Clarity
Brightness	.51**				
Timbral complexity	.33*	−.21			
Key clarity	−.11	−.12	.02		
Pulse clarity	.49**	.24	−.26	−.15	
Activity	.92***	.64***	−.27	−.12	.50**

* $p<.05$. ** $p<.01$. *** $p<.001$.

References

Alluri, V., Toiviainen, P., 2010. Exploring perceptual and acoustic correlates of polyphonic timbre. Music Percept. 27 (3), 223–241.

Alluri, V., Toiviainen, P., 2012. Effect of enculturation on the semantic and acoustic correlates of polyphonic timbre. Music Percept. 29 (3).

Blood, A.J., Zatorre, R.J., 2001. Intensely pleasurable responses to music correlates with activity in brain regions implicated in reward and emotion. Proc. Natl. Acad. Sci. 98, 11818–11823.

Blood, A.J., Zatorre, R.J., Bermudez, P., Evans, A.C., 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. Nat. Neurosci. 2, 382–387.

Brattico, E., 2011. Pitch perception: a physiological and individual act. In: Chouvel, J.-M.-., Hascher, X. (Eds.), Esthétique et Cognition. Publications de la Sorbonne, Paris, pp. 1–12.

Brattico, E., Jacobsen, T., 2009. Subjective appraisal of music. Ann. N. Y. Acad. Sci. 1169, 308–317.

Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge (MA).

Brown, S., Ingham, R.J., Ingham, J.C., Laird, A., Fox, P.T., 2005. Stuttered and fluent speech production: an ale meta-analysis of functional neuroimaging studies. Hum. Brain Mapp. 25, 105–117.

Burianova, H., McIntosh, A.R., Grady, C.L., 2010. A common functional brain network for autobiographical, episodic, and semantic memory retrieval. Neuroimage 49, 865–874.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.H., McAdams, S., 2006. Separate neural processing of timbre dimensions in auditory sensory memory. J. Cogn. Neurosci. 18, 1959–1972.

Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008. Moving on time: brain network for auditory–motor synchronization is modulated by rhythm complexity and musical training. J. Cogn. Neurosci. 20, 226–239.

Chevillet, M., Riesenhuber, M., Rauschecker, J.P., 2011. Functional correlates of the anterolateral processing hierarchy in human auditory cortex. J. Neurosci. 31 (25), 9345–9352.

Engel, A., Keller, P., 2011. The perception of musical spontaneity in improvised and imitated jazz performances. Front. Psychol. 2, 83. doi:10.3389/fpsyg.2011.00083.

Etkin, A., Egner, T., Kalisch, R., 2010. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci.. doi:10.1016/j.tics.2010.11.004

Fisher, R.A., 1950. Statistical Methods for Research Workers, 11th edn. Oliver and Boy, London.

Fox, M.D., Zhang, D., Snyder, A.Z., Raichle, M.E., 2009. The Global Signal and Observed Anticorrelated Resting State Brain Networks. J. Neurophysiol. 101 (6), 3270–3283.
Fraisse, P., 1982. Rhythm and tempo. In: Deutsch, D. (Ed.), Psychol. Music. Academic Press, New York.
Friston, K.J., Josephs, O., Zarahn, E., Holmes, A.P., Rouquette, S., Poline, J.-B., 2000. To smooth or not to smooth? Neuroimage 12, 196–208.
Gómez, E., 2006. Tonal description of music audio signal. PhD Thesis, Universitat Pompeu Fabra, Barcelona.
Grahn, J.A., Rowe, J.B., 2009. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. J. Neurosci. 29, 7540–7548.
Hahn, B., Ross, T.J., Stein, E.A., 2006. Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. Neuroimage 32, 842–853.
Halpern, A.R., Zatorre, R.J., Bouffard, M., Johnson, J.A., 2004. Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia 42, 1281–1292.
Harrington, D.L., Haaland, K., Hermanowicz, N., 1998. Temporal processing in the basal ganglia. Neuropsychology 12, 3–12.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.
He, B.J., Snyder, A.Z., Zempel, J.M., Smyth, M.D., Raichle, M.E., 2008. Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. Proc. Natl. Acad. Sci. 105, 16039–16044.
Jacobsen, T., Schubotz, R.I., Hofel, L., Cramon, D.Y., 2006. Brain correlates of aesthetic judgment of beauty. Neuroimage 29 (1), 276–285.
Janata, P., 2005. Brain networks that track musical structure. Ann. N. Y. Acad. Sci. 1060, 111–124.
Janata, P., Grafton, S., 2003. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. Nat. Neurosci. 6, 682–687.
Janata, P., Birk, J.L., van Horn, J.D., Leman, M., Tillmann, B., Bharucha, J.J., 2002a. The cortical topography of tonal structures underlying western music. Science 298, 2167–2170.
Janata, P., Tillmann, B., Bharucha, J.J., 2002b. Listening to polyphonic music recruits domain-general attention and working memory circuits. Cogn. Affect. Behav. Neurosci. 2, 121–140.
Jeffries, K., Fritz, J., Braun, A., 2003. Words in melody: an $H_2^{15}O$ PET study of brain activation during singing and speaking. Neuroreport 14, 749–754.
Keysers, C., Kaas, J.H., Gazzola, V., 2010. Somatosensation in social perception. Nat. Rev. Neurosci. 11, 417–428.
Khalfa, S., Schon, D., Anton, J.L., Liegeois-Chauvel, C., 2005. Brain regions involved in the recognition of happiness and sadness in music. Neuroreport 16, 1981–1984.
Koelsch, S., Fritz, T., von Cramon, D.Y., Muller, K., Friederici, A.D., 2006. Investigating emotion with music: an fMRI study. Hum. Brain Mapp. 27, 239–250.
Kornysheva, K., von Cramon, D.Y., Jacobsen, T., Schubotz, R.I., 2010. Tuning-in to the beat: aesthetic appreciation of musical rhythms correlates with a premotor activity boost. Hum. Brain Mapp. 31 (1), 48–64.
Krumhansl, C.L., 1990. Cognitive Foundations of Musical Pitch. Oxford University Press, Oxford.
Lartillot, O., Toiviainen, P., 2007. MIR in Matlab (II): a toolbox for musical feature extraction from audio. In: Dixon, S., Bainbridge, D., Typke, Rainer (Eds.), Proc. Intl. Conf. Music Inform. Retrieval, pp. 237–244.
Lartillot, O., Eerola, T., Toiviainen, P., Fornari, J., 2008. Multi-feature modeling of pulse clarity: design, validation, and optimization. Proc. ISMIR, Philadelphia.
Lazar, N.A., 2008. The statistical analysis of functional MRI data. Springer, New York.
Ledberg, A., Akerman, S., Roland, P.E., 1998. Estimation of the probabilities of 3D clusters in functional brain images. Neuroimage 8, 113–128.
Levitin, D.J., Menon, V., 2003. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann area 47 in temporal coherence. Neuroimage 20, 2142–2152.
Mcavoy, M., Larson-Prior, L., Nolan, T.S., Vaishnavi, S.N., Raichle, M.E., d'Avossa, G., 2008. Resting states affect spontaneous BOLD oscillations in sensory and paralimbic cortex. J. Neurophysiol. 100 (2), 922–931.
Meyer, L.B., 1956. Emotion and meaning in music. University of Chicago Press, Chicago.
Pallesen, K.J., Brattico, E., Bailey, C.J., Korvenoja, A., Gjedde, A., 2009. Cognitive and emotional modulation of brain default operation. J. Cogn. Neurosci. 21 (6), 1065–1080.
Pampalk, E., Rauber, A., Merkl, D., 2002. Content-based organization and visualization of music archives. Proc. IEEE ACM-MM, pp. 570–579.
Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12, 169–174.
Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.
Pyper, B.J., Peterman, R.M., 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. Can. J. Fish. Aquat. Sci. 55, 2127–2140.
Rao, S.M., Mayer, A.R., Harrington, D.L., 2001. The evolution of brain activation during temporal processing. Nat. Neurosci. 4, 317–323.
Saari, P., Eerola, T., Lartillot, O., 2011. Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music. IEEE Transactions on Audio, Speech, and Language Processing 19 (6), 1802–1812. doi:10.1109/TASL.2010.2101596.
Salimpoor, V., Benovoy, M., Larcher, K., Dagher, A., Zatorre, R.J., 2011. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nat. Neurosci. 14 (2), 257–262.
Salmi, J., Pallesen, K.J., Neuvonen, T., Brattico, E., Korvenoja, A., Salonen, O., et al., 2010. Cognitive and motor loops of the human cerebro-cerebellar system. J. Cogn. Neurosci. 22, 2663–2676.
Samson, F., Zeffiro, A.T., Toussaint, A., Belin, P., 2011. Stimulus complexity and categorical effects in human auditory cortex: an activation likelihood estimation meta-analysis. Front. Psychol. 1, 1–23.
Schaefer, R.S., Desain, P., Suppes, P., 2009. Structural decomposition of EEG signatures of melodic processing. Biol. Psychol. 82, 253–259.
Schubotz, R.I., von Cramon, D.Y., 2002. Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. Neuroimage 15, 787–796.
Schwartze, M., Keller, P.E., Patel, A.D., Kotz, S.A., 2011. The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. Behav. Brain Res. 216 (2), 685–691.
Sethares, W.A., 1998. Tuning, Timbre, Spectrum, Scale. Springer-Verlag.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423.
Sloboda, J., 1991. Music structure and emotional response: some empirical findings. Psychol. Music. 19, 110–120.
Smith, A.M., Lewis, B.K., Ruttimann, U.E., Ye, F.Q., Sinnwell, T.M., Yang, Y., Duyn, J.H., Frank, J.A., 1999. Investigation of low frequency drift in fMRI signal. Neuroimage 9, 526–533.
Stoodley, C.J., Schmahmann, J.D., 2009. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. Neuroimage 44, 489–501.
Tzanetakis, G., Cook, P., 2002. Music genre classification of audio signals. Proc. IEEE T Acoust Speech. 10, pp. 293–302.
Uddin, L.Q., Kelly, A.M.C., Biswal, B.B., Xavier Castellanos, F., Milham, M.P., 2009. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. Hum. Brain Mapp. 30, 625–637.
Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422.
Zanto, T.P., Snyder, J.S., Large, E.W., 2006. Neural correlates of rhythmic expectancy. Adv. Cogn. Psychol. 2, 221–231.
Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.