



## From Vivaldi to Beatles and back: Predicting lateralized brain responses to music



Vinoo Alluri<sup>a,b,\*</sup>, Petri Toiviainen<sup>a,1</sup>, Torben E. Lund<sup>c,2</sup>, Mikkel Wallentin<sup>c,f,3</sup>, Peter Vuust<sup>c,d,2,3</sup>, Asoke K. Nandi<sup>e,b,5</sup>, Tapani Ristaniemi<sup>b,4</sup>, Elvira Brattico<sup>a,g,h,6</sup>

<sup>a</sup> Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland

<sup>b</sup> Department of Mathematical Information Technology, University of Jyväskylä, Finland

<sup>c</sup> Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus University, Denmark

<sup>d</sup> Royal Academy of Music, Aarhus/Aalborg, Denmark

<sup>e</sup> Department of Electronic and Computer Engineering, Brunel University, UK

<sup>f</sup> Center for Semiotics, Aarhus University, Denmark

<sup>g</sup> Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, Finland

<sup>h</sup> Brain & Mind Laboratory, Department of Biomedical Engineering and Computational Science (BECS), Aalto University, Espoo, Finland

### ARTICLE INFO

#### Article history:

Accepted 18 June 2013

Available online 28 June 2013

#### Keywords:

fMRI

Computational feature extraction

Naturalistic stimulus

Auditory cortex

Orbitofrontal cortex

Cross-validation

### ABSTRACT

We aimed at predicting the temporal evolution of brain activity in naturalistic music listening conditions using a combination of neuroimaging and acoustic feature extraction. Participants were scanned using functional Magnetic Resonance Imaging (fMRI) while listening to two musical medleys, including pieces from various genres with and without lyrics. Regression models were built to predict voxel-wise brain activations which were then tested in a cross-validation setting in order to evaluate the robustness of the hence created models across stimuli. To further assess the generalizability of the models we extended the cross-validation procedure by including another dataset, which comprised continuous fMRI responses of musically trained participants to an Argentinean tango. Individual models for the two musical medleys revealed that activations in several areas in the brain belonging to the auditory, limbic, and motor regions could be predicted. Notably, activations in the medial orbitofrontal region and the anterior cingulate cortex, relevant for self-referential appraisal and aesthetic judgments, could be predicted successfully. Cross-validation across musical stimuli and participant pools helped identify a region of the right superior temporal gyrus, encompassing the planum polare and the Heschl's gyrus, as the core structure that processed complex acoustic features of musical pieces from various genres, with or without lyrics. Models based on purely instrumental music were able to predict activation in the bilateral auditory cortices, parietal, somatosensory, and left hemispheric primary and supplementary motor areas. The presence of lyrics on the other hand weakened the prediction of activations in the left superior temporal gyrus. Our results suggest spontaneous emotion-related processing during naturalistic listening to music and provide supportive evidence for the hemispheric specialization for categorical sounds with realistic stimuli. We herewith introduce a powerful means to predict brain responses to music, speech, or soundscapes across a large variety of contexts.

© 2013 Elsevier Inc. All rights reserved.

### Introduction

The past two decades have witnessed a surge of neuroimaging studies exploring music perception. In the classical approach commonly

employed in these studies, specific hypotheses are tested in controlled conditions wherein the condition of interest is interspersed with a rest baseline or with other control conditions. As a result it fails to capture the modulations caused due to continuous information flow to our sensory modalities occurring in the real world. Moreover, these studies typically employ the General Linear Model approach wherein the regressors are usually binary (1 representing a condition of interest and 0 representing baseline) or consist at most of a small number of discrete values thereby not allowing for generalizations of the models in other contexts as there is no parameterization of the presented stimulus. Furthermore, analysis of such data also typically includes averaging operations across scans, and sometimes also across regions of interest (ROIs), and therefore suffers from severe underestimation of the “amount of information collected in a single fMRI measurement” (Haynes and Rees, 2006). Despite these reservations, such an approach provides us with

\* Corresponding author at: Department of Music, University of Jyväskylä, PL 35(M), 40014 Jyväskylä, Finland.

E-mail address: [vinoo.alluri@jyu.fi](mailto:vinoo.alluri@jyu.fi) (V. Alluri).

<sup>1</sup> Department of Music, University of Jyväskylä, PL 35(M), 40014 Jyväskylä, Finland.

<sup>2</sup> Royal Academy of Music, Aarhus Skovgaardsgade 2C, DK-8000 Aarhus C, Denmark.

<sup>3</sup> Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Nørrebrogade, 8000 Aarhus C, Denmark.

<sup>4</sup> Department of Mathematical Information Technology, University of Jyväskylä, P.O. Box 35 (Agora), FIN-40014, Finland.

<sup>5</sup> Department of Electronic and Computer Engineering, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

<sup>6</sup> Institute of Behavioral Sciences, P.O.B. 9, 00014, University of Helsinki, Finland.

a macro-level description, spatially and temporally, of the functionality of the regions in the brain thereby providing a foundation for formulating hypotheses for further investigations employing alternate techniques to study the temporal evolution of neural processing in a more fine-grained manner.

Recent studies, especially concerning the visual modality, have promoted a novel method of exploring brain functionality at a micro-level, namely employing voxel-based encoding. Naselaris et al. (2011) describe it as a technique that enables prediction of brain activity using stimulus-based features, at a voxel-level. Due to the possibility of representing the stimulus as features, the models created using such an approach can be tested on alternate datasets in order to assess their generalizability. However, this approach has been tested so far using only random selections of natural scenes (Kay et al., 2008; Naselaris et al., 2009) rather than continuous natural stimuli. To add to that, Wu et al. (2006) highlight the laboriousness involved in analyzing data obtained using natural stimuli. Nevertheless, studying neural processing as a continuous process is of vital importance in order to understand how the brain processes information, be it via any sensory modality, in the real world.

In the auditory modality, voxel-based encoding studies are scarce. One of the initial investigations was performed by Janata et al. (2002) in which they used fMRI and identified the rostral and medial part of the superior frontal gyrus or medial prefrontal cortex (MPFC) as the area that was engaged in tonality-tracking of an artificially created stimulus that modulated through major and minor keys systematically. In a subsequent study Janata (2009) examined music-evoked autobiographical memories and the interaction of musical features representing tonality. This time their stimuli comprised 30-second excerpts of natural music which were familiar or unfamiliar to the participants. The participants were asked to perform a behavioral rating task after each excerpt. The tonal structure in the music was captured by a perceptually validated toroidal model that is known to emulate the perceptual and cognitive processes involved in tonality processing. Results revealed the MPFC as the core area associated with processing music-evoked memories in addition to several frontal, secondary visual, and sub-cortical regions. Chapin et al. (2010) reported that the level of expressivity in a piano performance affected the evolution of neural activity. The level of expressivity was characterized by two features representing the evolution of tempo and sound intensity which were found to be processed by limbic and paralimbic regions of the brain as well as the dorsal MPFC. Particularly they report the involvement of several regions of the brain previously known to process pulse belonging to the primary motor and somatomotor areas.

The only study to date that has investigated voxel-based continuous encoding of several timbral, rhythmic, and tonal features simultaneously in the brain was performed by Alluri et al. (2012). They introduced a novel approach for investigating naturalistic music processing in the brain. Ecological musical material (an instrumental modern tango) was used in a more realistic setting wherein musicians listened to the stimulus uninterrupted and without distraction by any experimental task unlike previous imaging studies that employed controlled settings with task requirements. Subsequently a comprehensive set of acoustic features was extracted from the stimuli and correlation analyses were performed with the fMRI time-series to determine where these features were processed. As a result they were able to localize the processing of the extracted musical features in the brain with timbral features recruiting auditory and somatomotor areas and higher-order rhythmic and tonality features involving also emotion-related limbic and paralimbic areas of the brain. However, the generalizability of these results is yet to be assessed. Here we extend this novel approach to other musical stimuli. In the present study, we aimed at predicting the temporal evolution of brain activity in relation to acoustic features extracted from musical pieces belonging to various genres, with and without lyrics. As generalizability was the

main goal, we also assessed the robustness of the hence created models across stimuli via cross-validation.

In light of previous studies that examined temporal evolution of musical features in the brain (Alluri et al., 2012; Chapin et al., 2010; Janata, 2009) we hypothesized that the encoding models of the voxels which were found to be associated with musical feature processing would possess a high degree of generalizability across musical stimuli. Specifically, we expected to successfully predict the activations in the auditory, default mode network (DMN)-related, limbic and paralimbic, and somatomotor-related brain areas as well as cerebellar cognitive areas, and dorsal and rostromedial regions of the prefrontal cortex. Furthermore, activations in regions belonging to the ventrolateral prefrontal cortex (VLPFC), extrastriate visual areas (cuneus), and cerebellum may also be predictable at more liberal significance thresholds (Janata, 2009). Furthermore, following the notion of lateralized processing of speech in the left hemisphere (e.g., Zatorre et al., 2002) and in line with a previous study comparing brain activity to music without lyrics and music including lyrics (Brattico et al., 2011), we anticipated that the presence of lyrics might shift the balance in processing of musical features more to the right-hemisphere, measured as the number of voxels whose activations can be modeled with high accuracy. Based on this, we predicted that voxel-based models of music without lyrics would have lower predictive accuracy for activations in the left auditory cortex than those models related to music containing lyrics. Previous evidence indicated that voice stimuli (and other sound stimuli containing fast spectrotemporal transitions; Zatorre and Gandour, 2009) are predominantly processed in the left primary and adjacent auditory cortices (Heschl's gyrus and superior temporal gyrus), whereas musical stimuli are bilaterally processed or else preferentially in the corresponding areas of the right hemisphere (Garza-Villarreal et al., 2011; Hickok and Poeppel, 2000; Samson et al., 2011; Tervaniemi and Hugdahl, 2003; Zatorre et al., 2002). To note, this evidence had been obtained with carefully controlled parametric studies or with artificial designs including direct contrasts between two sound categories. Our goal here is also to find support for hemispheric specialization for sound processing by utilizing realistic stimuli, a naturalistic listening condition, and novel data processing techniques.

One main advantage of employing the voxel-based encoding technique is the possibility of testing the models in alternate settings, which helps enhance their generalizability. To this end, we further extended the cross-validation procedure by using a third dataset, specifically the one obtained in Alluri et al. (2012), as it was comparable to the current data in terms of experimental setting. That dataset comprised continuous fMRI measurements of musicians' brains while listening to the tango, *Adios Nonino* by Astor Piazzolla. Such comparisons would then facilitate the determination of regions of the brain that are dedicated to musical feature processing across different participant pools as well.

## Materials and methods

### Stimuli

Two musical medleys, lasting approximately 15 min each, including pieces from various genres, with and without lyrics, were used in two separate scan sessions. One of the medleys comprised the B-side from *Abbey Road* by The Beatles (1969) and will be referred to as *Abbey Road*. The other consisted of four musical pieces without lyrics from different genres: Booker T and the MGs (*Green Onions*), Vivaldi (*The Four Seasons – Spring*), Miles Davis (*Straight, no chaser*), and The Shadows (*Apache*), and will be referred to as *Medley*. Stimuli were played in mono (sample frequency: 22,050 Hz) and delivered through pneumatic headphones from Avotec (Stuart, FL USA). Participants were instructed to listen carefully to the music. To maintain their attention, we inserted in four places of the stimulation the voice stimulus “nu” (Danish word

for “now” in English) and asked subjects to press a button whenever they heard it.

### Participants

Participants for the Abbey Road comprised eleven healthy individuals (mean age:  $26.4 \pm 5$  SD; 7 males) whereas for the Medley, twelve participants (mean age:  $25.9 \pm 5.7$  SD; 8 males) were scanned. Participants were selected without regard to their musical education. None reported any neurological, hearing or psychological problems. Eight participants were common between both sets of participants. All participants were right-handed. Permission for the study was obtained from the local ethics committee (Region Midtjylland, Denmark) and written informed consent was obtained from each participant. Each received a 100 DKK payment per hour for participation. All participants had Danish as their primary language.

### fMRI measurements

A 3 T General Electric Medical Systems (Milwaukee, WI USA) MR system with a standard head coil was used to acquire both T2-weighted gradient echo, echo-planar images (EPI) with Blood Oxygenation Level-Dependent (BOLD) contrast and T1-weighted structural images. 464 EPI volumes were acquired per participant for the Abbey Road stimulus and 410 for the Medley stimulus respectively. The first five volumes were discarded to allow for effects of T1 equilibrium. Whole brain coverage was achieved using 39 (Medley) and 42 (Abbey Road) axial slices of 3 mm thickness with an in-plane resolution of  $3 \times 3$  mm in a  $64 \times 64$  voxel matrix (FOV 192 mm). Images were obtained with a 2200 ms TR, a 30 ms TE and a  $90^\circ$  flip angle. A high-resolution 3D GR T1 anatomical scan was acquired for spatial processing of the fMRI data. It consisted of  $256 \times 256 \times 134$  voxels with a  $0.94 \text{ mm} \times 0.94 \text{ mm} \times 1.2 \text{ mm}$  voxel size, obtained with a 6.552 ms TR, a 2.824 ms TE and a  $14^\circ$  flip angle.

### fMRI preprocessing

Whole-brain image analysis was carried out using Statistical Parametric Mapping 8 (SPM8 – <http://www.fil.ion.ucl.ac.uk/spm>). For each subject the images were realigned, spatially normalized into the Montreal Neurological Institute template (12 parameter affine model, gray matter segmentation; realignment: translation components  $< 2$  mm, rotation components  $< 2^\circ$ ), and spatially smoothed (Gaussian filter with FWHM of 6 mm). Following this, 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), was used to detrend the fMRI responses. Next, temporal smoothing was performed as it provides a good compromise between efficiency and bias (Friston et al., 2000). The Gaussian 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. Finally, the effect of the participants' movements was removed by modeling the 6 movement parameters as regressors of no interest.

### Acoustic feature extraction and processing

For the present study, we used an approach similar to that of Alluri et al. (2012) for acoustic feature selection and extraction. Twenty-five acoustic features capturing timbral, rhythmical and tonal properties were extracted from both stimulus sets using the MIRtoolbox (Lartillot and Toivainen, 2007). The features were extracted using a frame-by-frame analysis approach commonly used in the field of Music Information Retrieval (MIR). The duration of the frame was 25 ms with a 50% overlap between two adjacent frames for the timbral features, while it was 3 s with a 67% overlap for the rhythmical

and tonal features respectively (see Alluri et al., 2012 for a comprehensive overview). All the operations were performed in the MATLAB environment. Following extraction, a series of post-processing operations similar to those carried out in Alluri et al. (2012) were performed on the extracted features to make the data comparable to the fMRI data which are described in brief below.

The lag caused in the fMRI data due to the hemodynamic response was accounted for in the acoustic features by means of a convolution operation performed on them using a double-gamma Hemodynamic Response Function (HRF). Next, the convolved acoustic feature time-series were subjected to the same detrending operation in the post-processing stage of the fMRI data in order 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 excluded the first 26 s corresponding to the length of the HRF in order to avoid any artifacts due to the convolution operation. Following this, all the features were downsampled to match the sampling rate of the fMRI data (.45 Hz). The processed feature sets will be henceforth referred to as  $F_A$  and  $F_M$  for Abbey Road and Medley respectively.<sup>7</sup>

### Statistical analysis

First the mean inter-subject correlation (ISC) was estimated as a measure of inter-subject consistency of the fMRI data. The significance of the observed correlations was estimated by a Monte Carlo approach. Then, the brain responses were averaged across participants for each dataset. The process of modeling and predicting the temporal evolution of the brain responses was carried out on the averaged brain responses by employing principal component regression (PCR) modeling. Finally, the performance of these models was evaluated in a cross-validation setting in order to assess their generalizability. These steps are explained in detail in the following.

#### Principal component regression (PCR) modeling

In order to model the responses of every voxel, we chose to perform linear regression using the acoustic features as independent variables. However, it is common for acoustic features extracted from any stimulus to covary. Hence, to deal with the collinearity problem we performed a PCR analysis. This process is described in what is to follow.

Principal Component Analysis (PCA) was first performed on the acoustic features. Nine principal components (PCs) were chosen as they proved to be sufficient in capturing the major proportion of the variance (96% for Abbey Road and 94% for Medley). The PC projections were subsequently obtained by the following equations:

$$\mathbf{PC}_A = \mathbf{F}_A \mathbf{L}_A \quad (1)$$

$$\mathbf{PC}_M = \mathbf{F}_M \mathbf{L}_M \quad (2)$$

where  $\mathbf{PC}_x$  represents the principal component scores matrix and  $\mathbf{L}_x$  represents the loading matrix for the first nine PCs for the respective datasets. Following this, linear regression was performed on a voxel-by-voxel basis for each dataset using the principal component scores time-series as independent variables and voxel time-series as dependent variables. Eqs. (3) and (4) describe the process of estimating the beta coefficients per voxel using PCR.

$$V_A^i \approx \mathbf{PC}_A \mathbf{b}_A^i \quad (3)$$

$$V_M^i \approx \mathbf{PC}_M \mathbf{b}_M^i \quad (4)$$

where  $V^i$  represents the voxel time-series,  $\mathbf{b}^i$  a vector of nine beta coefficients of the  $i$ th voxel. The correlation value between the predicted

<sup>7</sup> From now on subscripts A represents data, be it features, beta coefficients, loading matrix, etc. corresponding to the Abbey Road dataset and M represents the same for the Medley dataset.

voxel time-series and the observed time-series was used as a measure of the goodness of fit for the models. The significance of the correlations was estimated by means of a Monte Carlo approach. 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 used the Monte Carlo simulation approach described in Alluri et al. (2012) originally proposed by Ledberg et al. (1998).

#### Cross-validation of models

Following the modeling stage, in order to estimate the generalizability of the models across stimuli, we assessed their predictive power on different datasets. To this end, we used a cross-validation approach wherein the brain responses to the Medley were predicted by the regression models created using the Abbey Road as stimulus and vice versa. The prediction process is described in the equations below.

$$\tilde{V}_A^i = \mathbf{F}_A \mathbf{L}_M \mathbf{b}_M^i \quad (5)$$

$$\tilde{V}_M^i = \mathbf{F}_M \mathbf{L}_A \mathbf{b}_A^i \quad (6)$$

$\tilde{V}_A^i$  represents the predicted voxel time-series at the  $i$ th voxel in response to the Abbey Road using the regression models corresponding to the Medley dataset, and  $\tilde{V}_M^i$ , the predicted voxel time-series at the  $i$ th voxel in response to the Medley using the regression models corresponding to the Abbey Road dataset.

Following this, we performed Pearson correlation analyses voxel-by-voxel between the predicted brain responses  $\tilde{V}_X^i$  and observed responses  $V_X^i$  respectively as shown in Eqs. (7) and (8).

$$r_A = \text{Corr}(\tilde{V}_A^i, V_A^i) \quad (7)$$

$$r_M = \text{Corr}(\tilde{V}_M^i, V_M^i) \quad (8)$$

where *Corr* represents the Pearson correlation operation.

This operation resulted in correlation maps  $\mathbf{r}_A$  and  $\mathbf{r}_M$  for each dataset. The significance of the correlations was estimated employing a Monte Carlo approach. The correlation maps were then thresholded at a significance level of  $p < .001$ .

The thresholded  $\mathbf{r}$  maps were converted to their respective Z-maps using the Fisher's Z-transformation. In order to obtain the regions in the brain where the activation can be predicted significantly using both models, we performed an intersection operation between the Z-maps  $\mathbf{Z}_A$  and  $\mathbf{Z}_M$ . The intersection operation serves the purpose of obtaining the regions in the brain where the activation can be predicted at a significance level of  $p < .001$  using either model. Following this, the minimum Z statistic of the regions that are obtained after the intersection is reported. This approach of choosing the minimum statistical test for conjunction of effects is proposed by Friston et al. (1999), albeit in more conventionally controlled settings where an effect is examined by interspersing the condition of interest with a rest baseline. Jordan and Krishnamoorthy (1996) propose methods for combining confidence intervals based on  $t$ -statistics. In light of their methods, the intersection operation we perform provides a conservative threshold. Therefore the results at a more liberal threshold will also be presented. The resulting significance maps are corrected to minimize Type I errors as described in the previous section.

#### Cross-validation across participant pools

Following the cross-validation operation, which aimed at assessing the generalizability of the models across stimuli, we further decided to examine the strength of the models across entirely different

participant pools as well. To this end we chose a dataset that was investigated previously by Alluri et al. (2012) as the experimental setting was comparable to the current setting. Their data comprised continuous fMRI measurements of musicians listening to the piece Adios Nonino by Astor Piazzolla. Pair-wise cross-validation across stimuli was then performed as described in the previous section. The results obtained are discussed in the next section.

## Results

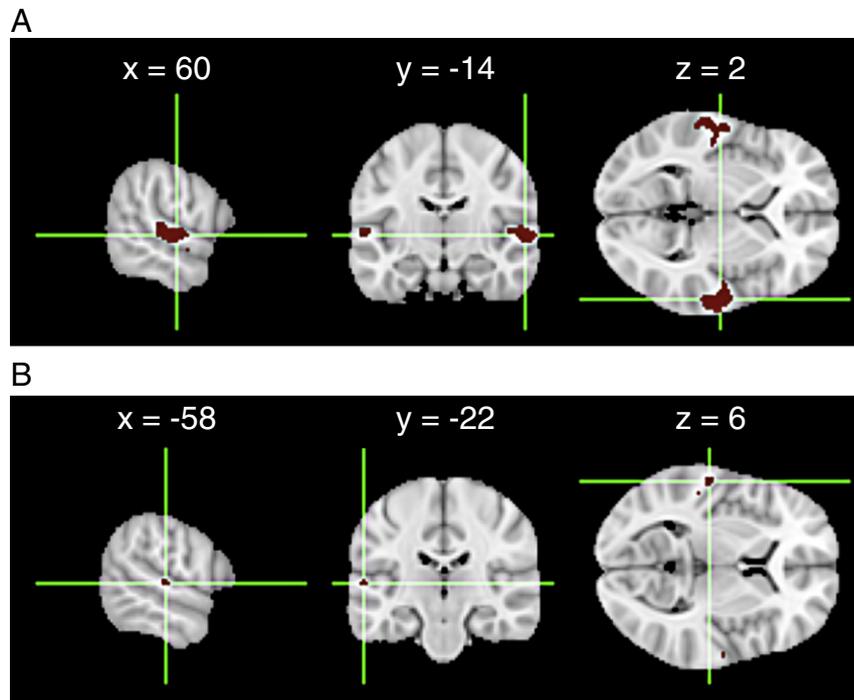
### ISC and PCR models

Fig. 1 displays the results of the correlation analysis performed on the participants' fMRI responses. As can be seen, significant mean inter-subject correlations were demonstrated in the auditory cortices with the maximum value in the superior temporal gyrus (STG) ( $r = .34$ ,  $p < .01$  for Abbey Road and  $r = .28$ ,  $p < .05$  for Medley).

Subsequently, we performed PCR analyses on each of the datasets as described in the Principal component regression (PCR) modeling section. Fig. 2 demonstrates the correlation between the predicted and observed fMRI time-series for each of the datasets reported at a significance level of  $p < .0001$  (see Supplementary Tables S.1 and S.2 for list of areas).

Results showed that activations in large areas of the brain including auditory, limbic, motor, somatosensory, and frontal areas could be predicted with significant accuracy. However, the Abbey Road model predicts the activations in a larger proportion of the voxels than the Medley model. Both the models were able to predict activations in the bilateral STG and Heschl's gyrus. However, the Abbey Road model predicted a larger proportion of voxel activations in the right hemispheric primary and non-primary regions of the auditory cortex (including Heschl's gyrus, the planum polare and planum temporale of the STG and extending even to the Rolandic operculum and the insular cortex). In contrast, the Medley model predicted fewer right hemispheric voxel activations mainly located inside or in the vicinity of the Heschl's gyrus. Also, activations in broader left-hemispheric clusters encompassing primary and non-primary regions of the auditory cortex, the Rolandic operculum and the temporal pole were predicted by the Abbey Road model in comparison to the few voxels in the left hemisphere predicted by the Medley model. Statistical tests indeed confirmed significant hemispheric differences in the accuracy of prediction between the auditory cortex regions, and particularly the STGs, for the two stimuli. In particular, for the Abbey Road the STG volume for which activation could be significantly predicted at  $p < .001$  was 7.3 cm<sup>3</sup> in the left hemisphere and 12.0 cm<sup>3</sup> in the right hemisphere, whereas for the Medley the volume which was significantly predicted was 5.9 cm<sup>3</sup> in the left STG and 4.8 cm<sup>3</sup> in the right STG (see Fig. 3). Similarly, for the Abbey Road the proportion of STG volume for which activation could be significantly predicted was 40% in the left hemisphere and 51% in the right hemisphere, whereas for the Medley it was 32% in the left STG and 21% in the right STG. Finally, for the Abbey Road the mean  $z$  value was 2.97 in the left STG and 3.22 in the right STG, whereas for the Medley it was 2.77 in the left STG and 2.54 in the right STG. According to permutation tests, all these values were significantly greater in the right STG than in the left STG for the Abbey Road ( $p < .001$ ), and significantly smaller in the right STG than in the left STG for the Medley ( $p < .001$ ).

In the limbic areas, especially in the cingulate gyrus, the voxels whose activations could be predicted were more posterior for the Abbey Road model and more anterior for the Medley model. In addition, right-hemispheric amygdala, hippocampus and putamen activations could be predicted with high accuracy by the Abbey Road model and not by the Medley model. Furthermore, both models could predict activations in parts of the postcentral gyrus and right hemispheric supplementary motor area. However, only the Abbey Road model was



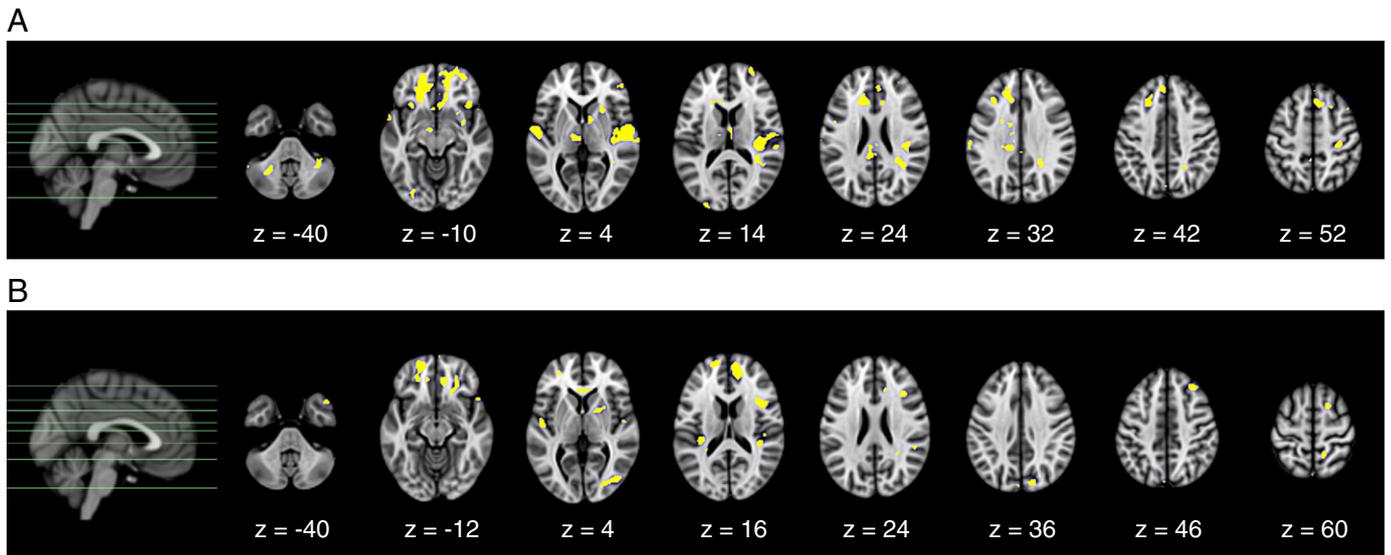
**Fig. 1.** A. Sagittal, coronal and horizontal views of the areas in the brain that displayed maximal mean inter-subject correlation for the Abbey Road dataset ( $p < .05$ ). The coordinates indicate the location in mm in MNI space. B. Sagittal, coronal and horizontal views of the areas in the brain that displayed maximal mean inter-subject correlation for the Medley dataset ( $p < .05$ ). The coordinates indicate the location in mm in MNI space.

capable of predicting activations in the bilateral precentral gyrus and the cerebellum.

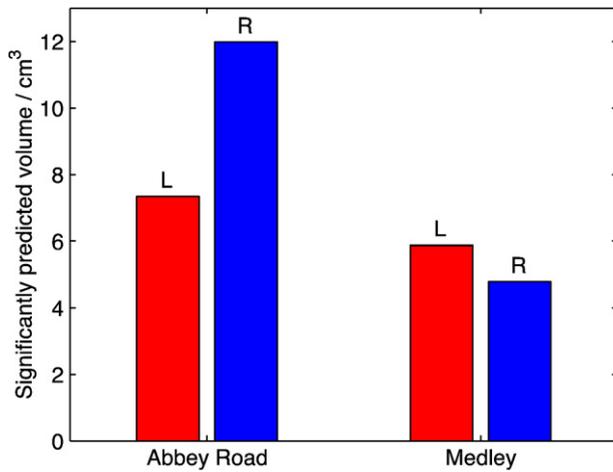
Activations in several other right hemispheric structures could be predicted for the Abbey Road dataset including the supplementary motor area, and parts of the cerebellum extending towards the inferior temporal gyrus and fusiform gyrus. In contrast, the Medley model could predict activations in several occipital areas including

the left hemispheric middle and inferior occipital gyrus, and bilateral cuneus. Activations in the parietal areas, specifically in the bilateral precuneus could be predicted relatively well with the Medley model whereas the Abbey Road model performance was more restricted in those areas.

Finally, both Medley and Abbey Road models (although with larger clusters again for the Abbey model) predicted activations of voxels



**Fig. 2.** A. Selected slices depicting the areas of the brain that displayed significant correlation between the predicted and observed fMRI time-series for the Abbey Road dataset ( $p < .0001$ ). Cluster correction was performed at a significance level of  $p = .001$  ( $Z = 3.29$ ) corresponding to a cluster size of 15 voxels. The slices are presented as seen from the top, therefore the left side of the viewer denotes the left hemisphere and the right side denotes the right hemisphere. The coordinates indicate the location of the slice along the z-axis (in mm) in MNI space. B. Selected slices depicting the areas of the brain that displayed significant correlation between the predicted and observed fMRI time-series for the Medley dataset ( $p < .0001$ ). Cluster correction was performed at a significance level of  $p = .001$  ( $Z = 3.29$ ) corresponding to a cluster size of 15 voxels. The slices are presented as seen from the top, therefore the left side of the viewer denotes the left hemisphere and the right side denotes the right hemisphere. The coordinates indicate the location of the slice along the z-axis (in mm) in MNI space.



**Fig. 3.** The volume in the left and the right STG, in  $\text{cm}^3$ , for which activation could be significantly predicted at  $p < .001$  for the Abbey Road and Medley.

in the medial and lateral parts of the orbitofrontal cortex, comprising the gyrus rectus, the olfactory sulcus, the orbital parts of the middle and superior frontal gyri, the pars orbitalis of the inferior frontal gyrus, the caudate, and for the Abbey model reaching ventrally even in the cingulate cortex up to the middle cingulate gyrus.

#### Cross-validation of Abbey Road and Medley models

Following this, the created models were assessed in a cross-validation setting as described in the [Cross-validation of models](#) section. [Fig. 4](#) shows the areas obtained as a result of the intersection operation. [Table 1](#) summarizes these results.

As can be seen, activations in the right hemispheric STG, HG and insula can be predicted significantly ( $p < .001$ ). However, at a more liberal threshold ( $p < .01$ ), activations in the bilateral auditory cortices including the Rolandic operculum, cuneus, precuneus, and superior medial frontal gyrus can also be significantly predicted, albeit with larger areas in the right hemisphere. Furthermore the activations in the right hemispheric putamen and middle cingulate gyrus, and left hemispheric postcentral gyrus could be predicted at this liberal threshold.

#### Cross-validation of Abbey Road, Medley, and Piazzolla models

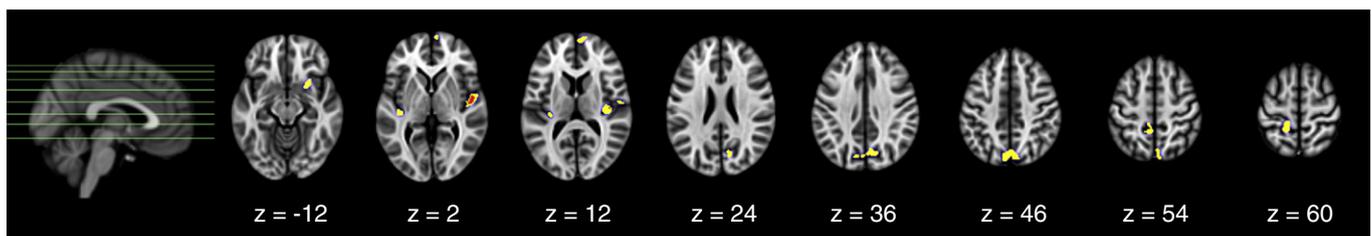
Cross-validation performed to assess the generalizability across participant pools as described in the [Cross-validation across participant pools](#) section revealed interesting results. [Fig. 5](#) depicts the regions in the brain where the predicted and the observed activations correlated significantly ( $p < .001$ ) as a result of cross-validation between the Abbey Road and Piazzolla, and between the Medley and Piazzolla models.

As can be seen in [Fig. 5A](#), activations in only the right hemispheric STG and HG could be predicted. At a more liberal threshold ( $p < .01$ ), we observe the emergence of the left precuneus as a region whose activations can be predicted. Interestingly cross-validation between the Medley and Piazzolla models, as is visible in [Fig. 5B](#), revealed that in the bilateral auditory cortices, specifically the left STG, and right STG and HG, the observed activations correlated significantly ( $p < .001$ ) with the predicted activations. In addition, activations in parts of the right hemispheric structures such as the superior parietal gyrus and postcentral gyrus could also be predicted with the same accuracy. Application of a more liberal threshold ( $p < .01$ ) revealed additional areas whose activations could be predicted. These include mainly the left hemispheric regions around the superior parietal gyrus, postcentral gyrus, and the left precuneus, in addition to primary and supplementary motor areas.

#### Discussion

The work presented here aimed at predicting the temporal evolution of brain responses to musical pieces with and without lyrics. To this end, we employed the paradigm introduced by [Alluri et al. \(2012\)](#) wherein participants were scanned continuously while listening to pop songs by the Beatles and to a medley of instrumental pieces belonging to various genres. Following this, acoustic features were extracted from the stimuli and used to model the brain responses using principal component regression. Subsequently these models were evaluated in a cross-validation setting in order to assess the robustness of the results. Finally, to improve further the generalizability of the results to neural processing of different kinds of musical material and levels of musical expertise, the cross-validation setting was extended to a third dataset, that is the one used by [Alluri et al. \(2012\)](#), which employed only musicians as participants and an Argentinean tango as the stimulus. The results are discussed below both in the light of previous neuroimaging studies that used the same paradigm and also those that have used artificial and acoustically reduced conditions to investigate musical feature processing.

Regression analyses revealed that activations in several regions in the brain including the frontal, auditory, limbic, somatosensory, and occipital areas could be predicted with significant accuracy by the models. The number of voxels whose activations could be successfully predicted was larger in the Abbey Road dataset most likely due to the higher mean inter-subject correlation observed when compared to the Medley dataset. Notably, emotion-related activations in amygdala and hippocampus were visible only with the Abbey Road dataset. Previous studies including movie watching ([Hasson et al., 2004](#); [Nummenmaa et al., 2012](#)) suggest that higher inter-subject correlation might be a prerequisite for mutual understanding between individuals in social situation. Our findings might hence be interpreted as suggesting similar brain processing for similar music, whereas the divergence of musical genres, associated with less powerful models of limbic activity, might signify a divergence of understanding and relating to emotional aspects of the



**Fig. 4.** Selected slices depicting the areas of the brain that displayed significant correlation between the predicted and observed fMRI time-series for Abbey Road and Medley datasets in the cross-validation setting at two significant levels ( $p < .001$  represented by orange and  $p < .01$  represented by yellow, cluster size corrected). The slices are presented as seen from the top, therefore the left side of the viewer denotes the left hemisphere and the right side denotes the right hemisphere. The coordinates indicate the location of the slice along the z-axis (in mm) in MNI space.

**Table 1**

Results of the cross-validation performed among all three datasets taken pairwise. The significant voxels were obtained at a threshold of  $p < .01$  (cluster corrected). The asterisk represents the significance level of  $p < .001$ . The clusters were obtained using the 18-connectivity scheme employed in SPM8. For each cluster, ROI analysis was performed using MarsBar. Each entry contains the sub-clusters obtained from the ROI analysis.  $k$  represents the number of voxels for the respective cluster. The coordinates are in MNI space and indicate the location of the global maximum within each cluster.

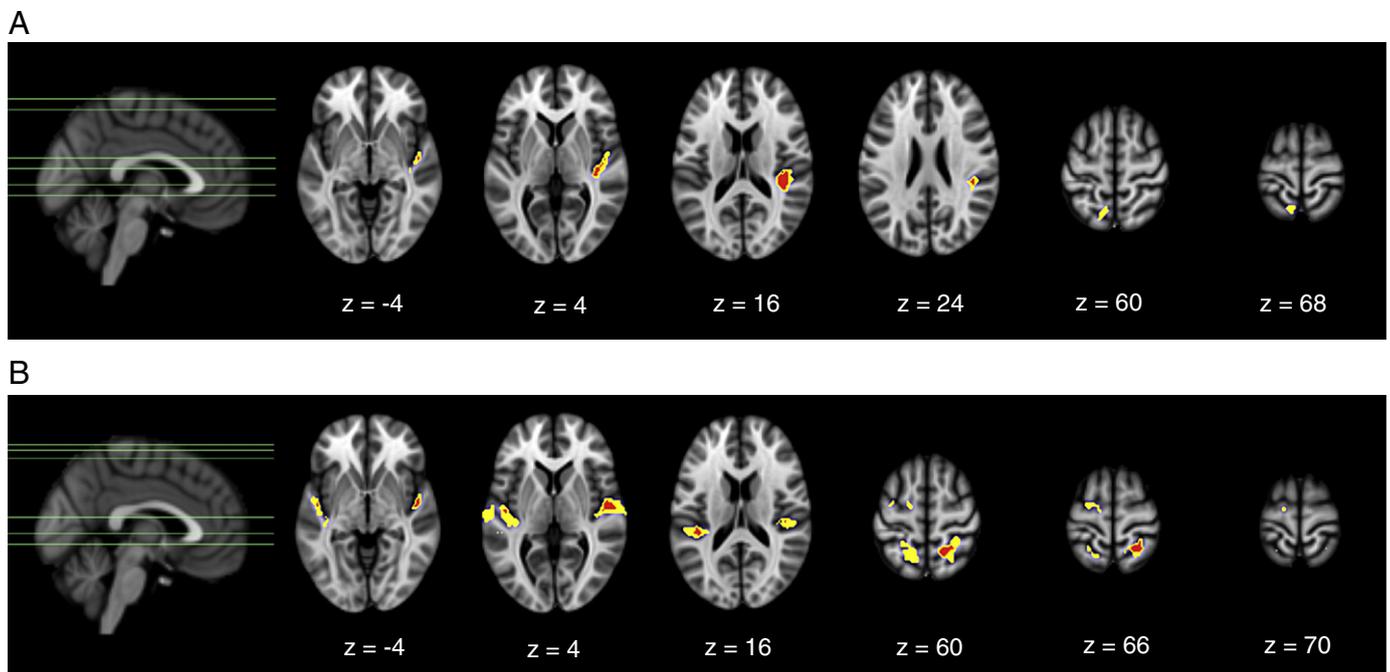
Left hemisphere	$k$	$x$	$y$	$z$	Z-value	Right hemisphere	$k$	$x$	$y$	$z$	Z-value
<i>Abbey Road &amp; Medley</i>											
Superior temporal gyrus, Heschl's gyrus	58	-38	-24	6	3.56	Insula, superior temporal gyrus, Heschl's gyrus, Rolandic operculum	291	50	-8	0	4.97*
Precuneus, median cingulate and paracingulate gyrus, postcentral gyrus, paracentral lobule	142	-8	-46	52	4.00	Putamen	42	26	10	-10	3.80
Medial part of superior frontal gyrus	28	2	60	18	3.58	Medial part of superior frontal gyrus	128	6	62	16	4.05
Precuneus, cuneus	170	-4	-74	44	4.02	Precuneus, cuneus	275	8	-70	36	3.90
<i>Abbey Road &amp; Piazzolla</i>											
Precuneus	90	-10	-58	68	3.74	Insula, superior temporal gyrus, Heschl's gyrus, supramarginal gyrus, Rolandic operculum	565	40	-32	20	4.86*
<i>Medley &amp; Piazzolla</i>											
Superior temporal gyrus, Heschl's gyrus, Rolandic operculum, middle temporal gyrus	693	-40	-30	12	4.39*	Superior temporal gyrus, Heschl's gyrus, Rolandic operculum, insula	585	48	-10	-2	4.53*
Precuneus, superior parietal gyrus, postcentral gyrus	226	-18	-54	60	3.86	Superior parietal gyrus, postcentral gyrus, precuneus	318	20	-48	68	4.19*
Precentral gyrus, superior frontal gyrus, supplementary motor area, paracentral lobule	68	-16	-10	62	3.52						

music. Future studies should control more closely the familiarity of subjects with a particular musical genre and test the hypothesis of background knowledge affecting emotional brain synchronization to music.

Interestingly, activations in music-related emotion-processing regions of the brain such as the amygdala, hippocampus, and parahippocampal gyrus (Blood et al., 1999; Brown et al., 2004; Koelsch, 2005; Chapin et al., 2010) were predictable by the Abbey Road model but not by the Medley model. This finding could be attributed to the fact that the presence of lyrics in the Abbey Road stimulus may engender stronger limbic responses, visible then in the statistical

maps, when compared to the non-lyrical pieces that comprised the Medley stimulus.

Furthermore, for the Abbey Road we were also able to predict the activations in the posterior cingulate, fusiform gyrus and cerebellar areas which in conjunction with the parahippocampal gyrus were previously reported to process songs with lyrics (Jeffries et al., 2003). Although the acoustic features were not selected particularly to represent the lyrical content of the stimulus, there is evidence from several studies in the field of Music Information Retrieval (MIR) corroborating the usefulness of acoustic features in building systems that classify



**Fig. 5.** A. Selected slices depicting the areas of the brain that displayed significant correlation between the predicted and observed fMRI time-series in the cross-validation between the Abbey Road and Piazzolla datasets at two significant levels ( $p < .001$  represented by orange and  $p < .01$  represented by yellow, cluster size corrected). The slices are presented as seen from the top, therefore the left side of the viewer denotes the left hemisphere and the right side denotes the right hemisphere. The coordinates indicate the location of the slice along the  $z$ -axis (in mm) in MNI space. B. Selected slices depicting the areas of the brain that displayed significant correlation between the predicted and observed fMRI time-series in the cross-validation between the Medley and Piazzolla datasets at two significant levels ( $p < .001$  represented by orange and  $p < .01$  represented by yellow, cluster size corrected). The slices are presented as seen from the top, therefore the left side of the viewer denotes the left hemisphere and the right side denotes the right hemisphere. The coordinates indicate the location of the slice along the  $z$ -axis (in mm) in MNI space.

music based on semantic content, which usually reflect the lyrical content (Turnbull et al., 2006, 2007). In contrast, for the Medley cuneus activations, which were previously found to be modulated by the intensity of chills evoked by listening to pleasurable music (Blood and Zatorre, 2001), were highly predictable.

The Medley model could predict activations in other occipital areas including the left hemispheric middle and inferior occipital gyri successfully. These areas were also previously associated with emotion processing during congruent presentation of musical and visual stimuli in contrast to only visual stimuli (Baumgartner et al., 2006). Alternatively, in the current context, activations in visual and associative areas to the Medley only might be explained by the higher default mode activity and mind-wandering during listening to instrumental music as compared to the more attention-catching songs from the Abbey Road. Indeed, in addition to the precuneus, the cuneus and the middle and inferior occipital gyri have been identified as crucial nodes in the default mode network (Buckner et al., 2008; Raichle et al., 2001).

The right-hemispheric asymmetry observed for the Abbey Road dataset when compared to the Medley dataset, especially in the STG, supports our hypothesis based on previous neuroimaging findings (Zatorre and Gandour, 2009; Zatorre et al., 2002; Samson et al., 2011; Tervaniemi and Hugdahl, 2003) that the presence of lyrics seems to shift the balance of processing more to the right-hemispheric STG. In a previous study, we observed that affective listening of 18-s music excerpts containing lyrics recruited activations in the left-hemispheric STG (among other areas) when contrasted with listening to 18-s excerpts of instrumental music, whereas the opposite contrast did not evidence any difference in activations of the auditory cortex (Brattico et al., 2011). Our current findings provide further support for the hypothesis of category-based specialization for music containing lyrics in the left auditory cortex, even during the naturalistic condition of continuous listening.

Both models also predicted the activations in the orbitofrontal cortex (OFC, including gyrus rectus, olfactory sulcus and orbital parts of the superior and middle frontal gyri) and ventromedial prefrontal cortex successfully. This result is in line with that of Janata (2009) where robust MPFC responses to tonal variation were observed for popular stimuli. The findings in the ventral and dorsal prefrontal regions might be related to the familiarity of the participants with some of the pieces included in the stimuli. For instance, in the Medley stimulus, Vivaldi's Four Seasons is very familiar to everybody and so must be some of Beatles' songs in the Abbey Road stimulus. Furthermore, converging evidence highlights the role of the orbitofrontal cortex as one that relates to processing pleasurable emotions connected to reward (Kringelbach, 2005; Liu et al., 2011). A recent meta-analysis by Liu et al. (2011) identifies the medial OFC as the locus for attribution of positive valence to a reward stimulus (such as food or money) during a decision-making process such when opting for food or when betting to gain or not lose money. Indeed, the medial part of the orbitofrontal cortex has been associated to appraisal based on the hedonic content of a stimulus, even in an artistic context (Brown et al., 2011; Ishizu and Zeki, 2011). These findings allow us to speculate that the participants found both stimuli pleasurable and give grounds for novel investigations of musical enjoyment using naturalistic listening to music. In the above context, positive rewarding appraisal of the musical pieces and their familiarity may have interacted to generate an orbitofrontal response, since familiarity has proven to be a crucial factor determining limbic and reward activations when passively listening to 15-s excerpts of popular music (Pereira et al., 2011). Future studies should monitor with continuous measures the emotional valence of the musical pieces as well as their familiarity, in order to disentangle their relations and their neural correlates in a naturalistic listening condition.

As hypothesized, we found that activations in parietal areas (precuneus) which were previously found to be associated with processing timbre (Alluri et al., 2012), could be predicted with significant

accuracy by the individual models. However, activations in the primary motor cortex in conjunction with cerebellar areas, previously found to be associated with processing various polyphonic timbral dimensions, namely perceived "brightness", "activity" and "fullness" (Alluri et al., 2012), could be predicted successfully by the Abbey Road model only. This further reflects the difference between the consistencies in each of the datasets. More specifically, the higher mean inter-subject correlation of the Abbey Road dataset allowed for better prediction of activations than the Medley dataset. Furthermore, activations in areas associated with processing of clarity in rhythmic and tonal content in a naturalistic context (Alluri et al., 2012), specifically in the somatosensory cortex and supplementary motor areas were also successfully predicted. The somatosensory cortex was also previously found to be associated with processing timbral features in naturalistic conditions (Alluri et al., 2012). It must be noted that due to the covarying nature of musical features in compositions, it may be challenging to tease apart the causal relationships between musical elements and brain activations, especially in such naturalistic listening scenarios. One solution to overcoming the problem of teasing out the activations caused due to individual musical elements would be to use a more controlled approach, especially in terms of musical content, wherein each musical element can be artificially manipulated. However, such an approach would again paint but an incomplete picture of music processing in the brain and in turn hamper the ecological validity of the results. The current results call for further studies delving further into the influence of musical background on brain activations in naturalistic music listening conditions.

As the ulterior motive of the work presented here is generalizability, we investigated the predictive power of these models across stimuli and entirely different participant pools. The cross-validation between these two datasets and an additional dataset with only musicians as participants helped narrow down the core areas that process acoustic features. The comparison between the Abbey Road and Medley models revealed a region of the right superior temporal gyrus (STG), encompassing the planum polare and the Heschl's gyrus as the core structure for processing complex acoustic features within musical pieces from various genres. This finding is in accordance with results of previous controlled studies that report the function of the superior temporal gyrus anterior to the Heschl's gyrus as one of encoding and classification of auditory objects (Leaver and Rauschecker, 2010; Rauschecker and Scott, 2009; Samson et al., 2011; Schirmer et al., 2012). The findings we obtained when lowering the original conservative threshold revealed that the activations in additional areas belonging to the parietal areas, specifically around the precuneus, basal ganglia structures (putamen), and several regions of the limbic system (cingulate gyrus and insula) and the motor cortex can be predicted with significant accuracy. The above reported areas were previously associated with processing timbre and rhythm (Alluri et al., 2012) and thereby corroborate our hypothesis specifically concerning the music features representing timbre and rhythm and their generalizability across musical genres and musical knowledge. However, the models were unable to generalize over regions of the brain associated with tonality processing in a continuous context. One reason could be that unlike the musical piece by Piazzolla, which possessed large tonal variations, the Abbey Road and Medley were quite homogeneous in terms of their tonal qualities, as they contained more sparse key modulations. In order to quantify the tonal variations in each stimulus we used the harmonic change detection function (Harte et al., 2006). By doing so the Piazzolla stimulus was found to have larger tonal variations (mean harmonic change of 1.22) followed by Abbey Road (1.03) and Medley (.90). Hence the small variations of the feature representing key clarity were likely to be masked or averaged out due to the smoothing effect introduced by hemodynamic response.

Interestingly, the activations in ventral MPFC regions could be predicted at the liberal threshold. As discussed earlier, this finding can be attributed to the familiarity of the participants with the stimuli.

Due to the popularity of the stimuli of the current study, it is highly likely that they were also memory evoking for many participants. As hypothesized, the activations in the secondary visual area, the cuneus, were also predictable at the liberal threshold. In previous studies activation of secondary visual areas by familiar popular music stimuli has been interpreted as related to music-evoked mental imagery (Janata, 2009; Janata et al., 2007).

The cross-validation performed with the current datasets and the one used by Alluri et al. (2012) revealed interesting findings. Results confirmed that the highest predictability was observed in the right auditory cortex again reiterating its role as the core structure involved in processing music across various genres. Cross-validation of the Abbey Road and Piazzolla models allowed us to conclude that high generalizability was observed in some of the areas previously found to process timbral features, specifically the left-hemispheric precuneus and right-hemispheric primary and secondary auditory areas (Alluri et al., 2012). The left precuneus was also found by Chapin et al. (2010) to process expressivity in music, which in their study was characterized by tempo. Interestingly, the cross-validation of the models based on purely instrumental music, that is the Medley and Piazzolla, revealed that in addition to the bilateral auditory areas, activations in parietal timbre-processing related areas, parts of the motor cortex could be predicted successfully. This finding is in line with that of Chapin et al. (2010) where they report the involvement of the abovementioned areas in the processing of tempo.

The acoustic features representing timbre are considered low-level features, and are extracted using very short analysis windows of the order of tens of milliseconds. On the other hand the acoustic features representing rhythmic content are considered high-level features encapsulating context-dependent aspects of music and are therefore extracted using a longer analysis window of the order of a few seconds. Processing timbral features thus can be considered as a sensory process as opposed to rhythmic feature processing that requires higher cognitive processes. As a result, it is not surprising that activations in timbre-related areas could be predicted by all the models in the cross-validation approach. Furthermore, the MPFC activations found in the cross-validation of the Abbey Road and Medley stimuli and in Janata (2009) could not be predicted including the Piazzolla model in the cross-validation. This finding can be attributed to the fact that the Piazzolla stimulus was not as familiar as the other two stimuli. Disentangling the role of familiarity from the processing of tonality using the current approach calls for a study of its own.

## Conclusions

It was demonstrated that using continuous fMRI recording during naturalistic stimulus presentation, brain activity in response to acoustic feature processing could be predicted in the auditory, limbic, and motor regions of the brain with significant accuracy. For the first time, it is demonstrated that orbitofrontal regions, previously associated with evaluative judgments and aesthetic appraisal, are also recruited while passively listening to full musical pieces and their activations can be predicted significantly. Despite differences between the musical pieces with respect to genre and presence of lyrics, the cross-validation of brain activations helped identify a region of the right STG, encompassing the planum polare and the Heschl's gyrus, as the core structure for processing complex acoustic features within musical pieces from various genres. The right-hemispheric dominance displayed in the results suggests that the presence of lyrics might shift the balance of processing of musical features more to the right STG, hence providing novel support for hemispheric specialization in processing of sound categories by using naturalistic stimulation and listening conditions. Models based on purely instrumental music revealed that in addition to the bilateral auditory areas, bilateral parietal, somatosensory, and left hemispheric primary and supplementary motor areas were recruited for musical feature processing. In sum,

our novel approach reveals neural correlates of music processing during naturalistic listening across a large variety of musical contexts.

## Acknowledgments

This research was supported by the Academy of Finland (project number 7118616), TEKES (Finland) grant 40334/10 'Machine Learning for Future Music and Learning Technologies', MindLab grants from the Danish Ministry of Science, Technology and Innovation, the Academy of Finland (Post-Doctoral Researcher; project number 133673), and the University of Helsinki (Three-year Grant; project number 490083). Asoke K. Nandi would like to thank TEKES for the funding of the Finland Distinguished Professorship. V. Alluri would like to thank Iballa Burunat for her help in preparing the Supplementary material.

## Conflict of interest

There is no conflict of interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2013.06.064>.

## References

- Alluri, V., Toiviainen, P., Jääskeläinen, I., Sams, M., Gleason, E., Brattico, E., 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage* 59, 3677–3689.
- Baumgartner, T., Lutz, K., Schmidt, C.F., Jancke, L., 2006. The emotional power of music: how music enhances the feeling of affective pictures. *Brain Res.* 1075, 151–164.
- 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., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., Tervaniemi, M., 2011. A functional MRI study of happy and sad emotions in music with and without lyrics. *Front. Psychol.* 2. <http://dx.doi.org/10.3389/fpsyg.2011.00308> (Article 308).
- Brown, S., Martinez, M.J., Parsons, L.M., 2004. Passive music listening spontaneously engages limbic and paralimbic systems. *NeuroReport* 15 (13), 2033–2037.
- Brown, S., Gao, X., Tisdelle, L., Eickhoff, S.B., Liotti, M., 2011. Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities. *NeuroImage* 58, 250–258.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default system: anatomy, function, and relevance to disease. In the year in cognitive neuroscience 2008. *Ann. N. Y. Acad. Sci.* 1124, 1124, pp. 1–38.
- Chapin, H., Jantzen, K., Scott Kelso, J.A., Steinberg, F., Large, E., 2010. Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE* 5 (12), e13812. <http://dx.doi.org/10.1371/journal.pone.0013812>.
- Friston, K., Holmes, A., Price, C., Bqchel, C., Worsley, K., 1999. Multisubject fMRI studies and conjunction analyses. *NeuroImage* 10, 385–396.
- 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.
- Garza-Villarreal, E.A., Brattico, E., Leino, S., Ostergaard, L., Vuust, P., 2011. Distinct neural responses to chord violations: a multiple source analysis study. *Brain Res.* 1389, 103–114.
- Harte, C.A., Sandler, M.B., Gasser, M., 2006. Detecting harmonic change in musical audio. *Proc. Audio and Music Computing for Multimedia Workshop*, pp. 21–26.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640.
- Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138.
- Ishizu, T., Zeki, S., 2011. Toward a brain-based theory of beauty. *PLoS One* 6, e21852. <http://dx.doi.org/10.1371/journal.pone.0021852>.
- Janata, P., 2009. The neural architecture of music-evoked autobiographical memories. *Cereb. Cortex* 19, 2579–2594.
- Janata, P., Birk, J.L., Van Horn, J.D., Leman, M., Tillmann, B., Bharucha, J.J., 2002. The cortical topography of tonal structures underlying Western music. *Science* 298, 2167–2170.
- Janata, P., Tomic, S.T., Rakowski, S.K., 2007. Characterisation of music-evoked autobiographical memories. *Memory* 15, 845–860.
- Jeffries, K., Fritz, J., Braun, A., 2003. Words in melody: an H<sup>15</sup>O PET study of brain activation during singing and speaking. *NeuroReport* 14, 749–754.

- Jordan, S.M., Krishnamoorthy, K., 1996. Exact confidence intervals for the common mean of several normal populations. *Biometrics* 52, 77–86.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. *Nature* 452 (7185), 352–355.
- Koelsch, S., 2005. Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212.
- Kringelbach, M.L., 2005. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* 6, 691–702.
- Lartillot, O., Toivainen, 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.
- Leaver, A.M., Rauschecker, J., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.
- Ledberg, A., Akerman, S., Roland, P.E., 1998. Estimation of the probabilities of 3D clusters in functional brain images. *NeuroImage* 8, 113–128.
- Liu, Hairston, J., Schier, M., Fan, J., 2011. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 35, 1219–1236.
- Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L., 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63 (6), 902–915.
- Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L., 2011. Encoding and decoding in fMRI. *NeuroImage* 56, 400–410.
- Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., et al., 2012. Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci. U. S. A.* 109, 9599–9604.
- Pereira, C.S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S.L., Brattico, E., 2011. Music and emotions in the brain: familiarity matters. *PLoS One* 6 (11), e27241.
- Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D. A., and Shulman, G. L., 2001. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–682.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: non-human primates illuminate human speech processing. *Nat. Rev. Neurosci.* 12, 718–724.
- 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.
- Schirmer, A., Fox, M.P., Grandjean, D., 2012. On the spatial organization of sound processing in the human temporal lobe: a meta-analysis. *NeuroImage* 63, 137–147.
- 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.
- Tervaniemi, M., Hugdahl, K., 2003. Lateralization of auditory-cortex functions. *Brain Res. Rev.* 43, 231–246.
- Turnbull, D., Barrington, L., Lanckriet, G., 2006. Modeling music and words using a multi-class naive Bayes approach. *Proc. of the International Symposium on Music Information Retrieval*. University of Victoria, Victoria, Canada, pp. 254–259.
- Turnbull, D., Barrington, L., Torres, D., Lanckriet, G., 2007. Towards musical query-by-semantic-description using the cal500 data set. *Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM SIGIR, Amsterdam, pp. 439–446.
- Wu, M.C., David, S.V., Gallant, J.L., 2006. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* 29, 477–505.
- Zatorre, R., Gandour, J.T., 2009. Neural specializations for speech and pitch: moving beyond the dichotomies. In: Moore, B.C.J., Tyler, L.K., Marslen-Wilson, W. (Eds.), *The Perception of Speech: From Sound to Meaning*. Oxford University Press, pp. 275–304 (Reprinted from *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363(1493), 1087–1104 (2008)).
- Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.