F1000Research

DATA NOTE

High-resolution 7-Tesla fMRI data on the perception of musical genres – an extension to the *studyforrest* dataset [version 1; referees: 2 approved with reservations]

Michael Hanke[1,2], Richard Dinga[1], Christian Häusler[1], J. Swaroop Guntupalli[3], Michael Casey[4], Falko R. Kaule[1,5], Jörg Stadler[6]

[1]Psychoinformatics lab, Department of Psychology II, University of Magdeburg, Magdeburg, Germany
[2]Center for Behavioral Brain Sciences, Magdeburg, Germany
[3]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire, USA
[4]Bregman Laboratory, Department of Music and Department of Computer Science, Dartmouth College, Hanover, New Hampshire, USA
[5]Visual Processing Laboratory, Ophthalmic Department, Otto-von-Guericke-University, Magdeburg, Germany
[6]Leibniz Institute for Neurobiology, Magdeburg, Germany

Abstract

Here we present an extension to the *studyforrest* dataset – a versatile resource for studying the behavior of the human brain in situations of real-life complexity (http://studyforrest.org). This release adds more high-resolution, ultra high-field (7 Tesla) functional magnetic resonance imaging (fMRI) data from the same individuals. The twenty participants were repeatedly stimulated with a total of 25 music clips, with and without speech content, from five different genres using a slow event-related paradigm. The data release includes raw fMRI data, as well as precomputed structural alignments for within-subject and group analysis. In addition to fMRI, simultaneously recorded cardiac and respiratory traces, as well the complete implementation of the stimulation paradigm, including stimuli, are provided. An initial quality control analysis reveals distinguishable patterns of response to individual genres throughout a large expanse of areas known to be involved in auditory and speech processing. The present data can be used to, for example, generate encoding models for music perception that can be validated against the previously released fMRI data from stimulation with the "Forrest Gump" audio-movie and its rich musical content. In order to facilitate replicative and derived works, only free and open-source software was utilized.

This article is included in the Real-life cognition channel.

This article is included in the Neuroinformatics channel.

Open Peer Review

Referee Status: ☑☑

	Invited Referees	
	1	**2**
version 1 published 29 Jun 2015	？ report	？ report

1 **Cristiano Micheli**, School of Medicine and Health Sciences, University of Oldenburg Germany, **Jochem Rieger**, University of Oldenburg Germany
   Kirsten Weber, Max Planck Institute for Psycholinguistics Netherlands

2 **Karsten Müller**, Max Planck Institute for Human Cognitive and Brain Sciences Germany

Discuss this article

Comments (0)

Corresponding author: Michael Hanke (michael.hanke@gmail.com)

Competing interests: No competing interests were disclosed.

Background

Previously, we have released a large, high-resolution, 7 Tesla fMRI dataset on the processing of natural auditory stimuli – a two-hour audio movie[1]. Recently, we have extended this initial release with a detailed annotation of the emotional content of the stimulus[2] to broaden the range of research questions that could be addressed with these data. Here we further amend this dataset with additional high-resolution fMRI data from the same participants on the perception of musical genres. We employed a proven paradigm and stimuli that have been previously shown to enable investigation of distributed population codes of musical timbre in bilateral superior temporal cortices[3].

The present data release enables comparative studies of the representation of musical genres (spectrum, timbre, vocal content) with ultra high-field, high resolution fMRI data from a larger sample of participants. In conjuction with the previous data releases, it will also further expand the continuum of research question that can be approached with the joint dataset. For example, the development of encoding models for cortical representations of music in complex auditory stimuli (the audio-movie contains several dozen musical excerpts from a broad range of genres). To this end, we include extracted audio features that represent the time-frequency information of each stimulus in four different views. The views are mapped to different perceptually-motivated scales (mel and decibel scales) and via a decorrelating linear transformation (DCT-II). It is hoped that providing these example features will catalyze discoveries of auditory stimulus codes in neural populations.

Lastly, these data can also serve as a public resource for benchmarking algorithms for functional alignment [e.g., 4], or other analyses, and thus, further the availability of resources for the investigation of real-life cognition[5].

Materials and methods

Participants

Acquisition of the data described herein was part of a previously published study[1], and took place in close temporal proximity (no more than a few weeks apart). The participants in this data release are identical to those previously reported. They were fully instructed about the nature of the study and were paid a total of 100 EUR for their participation, which included the previously reported data acquisitions, as well as the one described herein. All data acquisitions were jointly approved by the ethics committee of the Otto-von-Guericke-University of Magdeburg, Germany (approval reference 37/13).

Stimulus

All stimuli employed in this study are identical to those used in a previous study [for details refer to 3]. They were five natural, stereo, high-quality music stimuli (6 s duration; 44.1 kHz sampling rate) for each of five different musical genres: 1) Ambient, 2) Roots Country 3) Heavy Metal, 4) 50s Rock'n'Roll, and 5) Symphonic (see Figure 1 for details).

Procedures and stimulation setup

The setup for audio-visual presentation was as previously reported[1]. Participants listened to the audio using custom-built in-ear headphones,
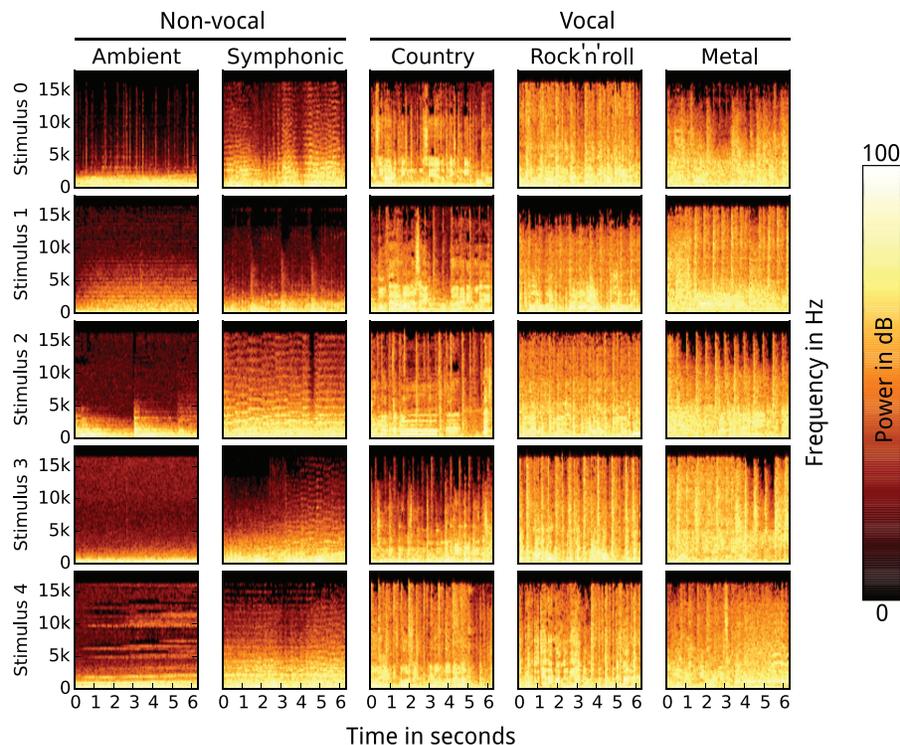
Figure 1. Spectrograms for all 25 stimuli showing structural differences in the time-frequency characteristics of the five musical genres. Each stimulus was a six second excerpt from the middle of a distinct musical piece. Excerpts were normalized so that their root-mean-square power values were equal, and a 50 ms quarter-sine ramp was applied at the start and end of each excerpt to suppress transients. Most prominent are the differences between music clips with and without vocal components.

and an LCD projector displayed visual instructions on a rear-projection screen that they saw via a mirror attached to the head coil.

At the start of each recording session, during the preparatory MR scans, participants listened to a series of longer excerpts of musical pieces and songs from the five different genres. During this phase participants were instructed to request adjustments of the stimulus volume in order to guarantee optimal perception of the stimuli against the noise pedestal emitted by the scanner. There was no overlap between the songs presented in this phase and those used as stimuli in the main experiment.

Eight scanning runs followed the initial sound calibration. Each run was started by the participant with a key-press ready signal. There were 25 trials, with five different stimuli (Figure 1) for each of the five genres per run (see Figure 2 for details on the

experiment design). At the end of each run participants were given the opportunity for a break of variable length until they indicated readiness for the next run. Most participants started the next run within a minute.

Stimulus presentation and response logging were implemented using PsychoPy[7] running on a computer with the (Neuro)Debian operating system[8].

Functional MRI data acquisition

The acquisition protocol for functional MRI was largely identical with the one previously reported[1], hence only differences and key facts are listed here.

Importantly, the same landmark-based procedure for automatic slice positioning that was used to align the scanner field-of-view between acquisition sessions, was used again to align the field-of-view
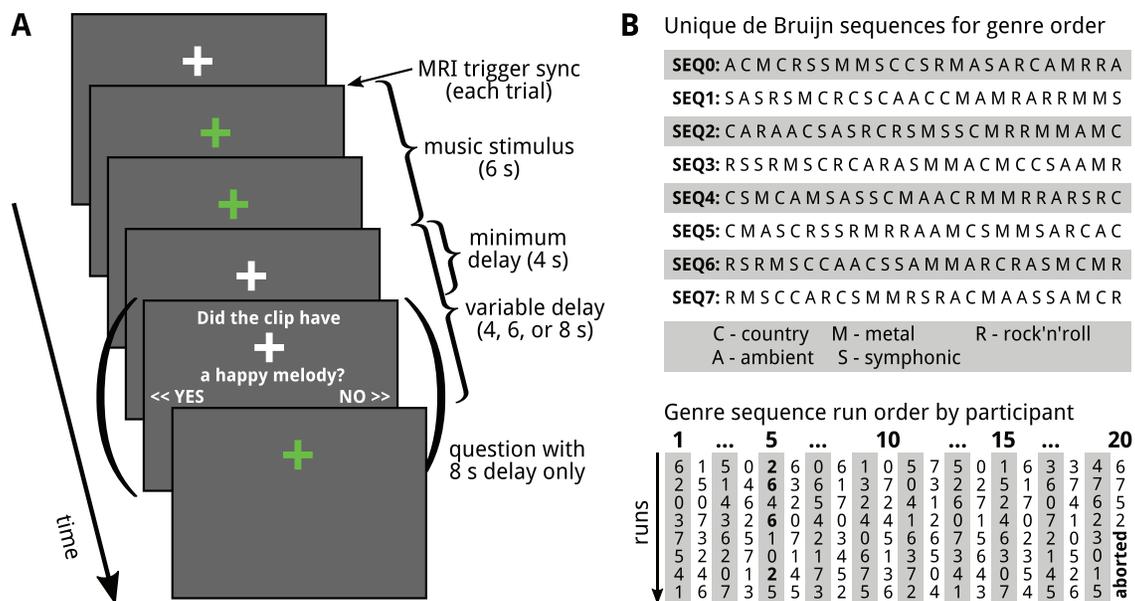

Figure 2. Experiment design. (**A**) Trial configuration. The start of each trial was synchronized with the MRI volume acquisition trigger. When the trigger was received the permanently displayed white fixation cross turned green, a 6 s music stimulus was presented, and, immediately afterwards, the fixation cross turned white again. Stimulation was followed by a variable delay (minimum delay 4 s). For the five trials of a genre, a 4 s and 8 s delay occurred once, while the remained three trials included a 6 s delay period. Thereby all trials had 4–8 s of uniform stimulation (no audio, white fixation cross) after each musical stimulus. The order of delays was randomized within a run. During trials with an 8 s second delay participants were presented with a yes/no question four seconds after the end of the music stimulus. The content of the question was randomized and asked for particular features of the stimulus that had just ended (e.g., "Was there a female singer?", "Did the song have a happy melody?"). Participants had to indicate their response by pressing one of two buttons with the index or middle finger of their right hand corresponding to the response alternative presented on the screen. "Yes" was always mapped to the left side (index finger), "No" always to the right side (middle finger). The question had the purpose of keeping the participants attentive to the stimuli and counteract the effect of increasing familiarity across multiple runs. (**B**) Run configuration. The 25 stimuli were identical across runs and presented exactly once per run. Order of stimulus genres within each run was counter-balanced using De Bruijn cycles[6] (alphabet size = 5, counter-balancing level = 2), hence each genre was followed by any other genre equally often and exactly once. Eight unique genre order sequences were generated and used for all participants, while randomizing the order of run sequences across participants. This was done in order to enable the application of the hyperalignment algorithm[4]. Data acquisition for two participants showed anomalies with respect to this procedure (see Table 2 for details).

for this acquisition with the one in the previous study[1]. As the exact same alignment target was used, this led to a very similar field-of-view configuration across acquisitions.

Each acquisition run consisted of 153 volumes (repetition time of 2.0 seconds with no inter-volume gaps).

Physiological recordings
The cardiac and respiratory traces were recorded for the full duration of all eight runs. The acquisition setup for physiological was identical with the one previously reported[1].

Dataset content
The released data comprises raw and pre-processed fMRI data, physiological recordings, behavioral log files, and auditory stimuli (total ≈95 GB). Table 1 provides an overview of the location of individual data components. The following sections briefly describe important properties.

Behavioral log files
Log files are available as plain text files with comma-separated value markup. All enumerations are zero-based. Each lines represents a trial. Columns for the following information are present: order of run in sequence (run), ID of trial sequence for this run (run_id; see Figure 2B), fMRI volume corresponding to stimulation start (total: volume, in the current run: run_volume), stimulus file name (stim), music genre label (genre), inter-stimulus interval in seconds (delay), flag whether a control question with presented (catch), measured asynchrony between MRI trigger and sound onset in seconds (sound_soa), and time stamp of the corresponding MRI trigger with respect to the start of the experiment in seconds (trigger_ts).

Information on the stimulus timing is also available in per-subject, per-run, per-condition plain-text files in FSL's EV3 format: one line per stimulation event, three columns with stimulus onset and duration (both in seconds relative to the start of a scan), as well as a third column with an arbitrary intensity weight that is always set to 1.

fMRI data
All functional MRI data were converted from the DICOM format into the NIfTI format for publication using the same procedure as in 1.

fMRI data are available in three different flavours, each stored in an individual 4D image for each run separately. Raw BOLD data are stored in bold.nii.gz. While raw BOLD data are suitable for further analysis, they suffer from severe geometric distortions. BOLD data that have been distortion-corrected[9] at the scanner console are provided in bold_dico.nii.gz. In addition, distortion-corrected data that have been anatomically aligned to a per-subject BOLD template image are available: bold_bold7Tp1_to_subjbold7Tp1.nii.gz.

Participant motion estimates
Head movement correction was performed with respect to a dedicated reference scan at the start of the recording session within scanner online reconstruction as part of the distortion correction procedure. The associated motion estimates are provided in a whitespace-delimited 6-column text file (translation X, Y, Z in mm, rotation around X, Y, Z in deg) with one row per fMRI volume for each run separately.

Physiological recordings
Physiological data were truncated to start with the first MRI trigger pulse and to end one volume acquisition duration after the last trigger pulse. Data are provided in a four-column (MRI trigger, respiratory trace, cardiac trace and oxygen saturation), space-delimited text file for each run. A log file of the automated conversion procedure is provided in the same directory (conversion.log). Sampling rate for the majority of all participants is 200 Hz (see Table 2 for exceptions).

Audio features
Recent experiments have shown that audio features can be predicted via regression models from fMRI signals to test stimulus coding hypotheses[3,10]. To facilitate this activity with the current data we extracted four audio features from down-mixed mono stimuli. Feature extraction used a front-end windowed short-time Fourier transform, with window size 16384 samples (371.52 ms) and hop size 4410 samples (100 ms) yielding 63 overlapping feature vectors per stimulus file. Window parameters were chosen to trade temporal for spectral acuity, yielding frequency samples spaced linearly at 2.69 Hz intervals from 0–22.05 kHz. The four features extracted from this representation are described below.

Mel-Frequency Spectrum (mfs) – 48 dimensions. Motivated by human auditory perception the mel scale organizes frequency by equidistant pitch locations as determined by psychophysical experiments. We used the essentia open source audio processing library[11] to extract the mel-frequency spectrum, which yielded energy in mel bands by applying a frequency-domain filterbank[12] to the short-time Fourier spectrum. Frequency-domain filtering consisted of applying equal area overlapping triangular filters to the Fourier spectrum spaced according to the mel scale and normalized such that the sum of coefficients for every filter equals one.

Mel-Frequency Cepstral Coefficients (mfcc) – 48 dimensions. Cepstral features have been widely reported to perform well in speech recognition and music classification systems[13], where the task is required to be sensitive to timbre. Typically, only the lower 10–20 cepstral coefficients (low quefrency) are retained; these encode the shape of the broad spectral envelope – an acoustic correlate of timbre. However, when sensitivity to timbre is not required, utilizing the upper coefficients (high quefrency), that encode fine spectral structure such as pitch, makes the feature robust to timbral changes[14]. We extracted the full set of 48 cepstral coefficients from the mel-frequency spectrum, by mapping the mel spectrum to a decibel amplitude scale and multiplying by the discrete cosine transform (DCT-II) matrix. It is expected that any application would first remove the constant first column and retain either the subsequent 13–20 coefficients or the remaining upper coefficients after those, depending on whether sensitivity or robustness to timbral difference is required. The remaining two features yield such a separation into low and high quefrency spectral components.

Table 1. Data set layout. File paths and descriptions for all available content.

File path	Description
`/`	**Meta data** OpenFMRI study meta data

`[model,scan,study,task]_key.txt`
 `models/model001/[condition_key,task_contrasts].txt`
 `acquisition_protocols/task002_fmri_session.pdf`
	Siemens 7T Magnetom acquisition protocol settings

`stimulus/task002/`
 `[intro,runs].csv`
 `mg_7T.psyexpi, make_design.py`
 `stimuli/[ambient,country,metal,rocknroll,symphonic]_00[0-4].wav`
 `make_audiofeature.py, features/*`
	Stimulus file and paradigm implementation
	 Helper CSV tables for PsychoPy experiment
	 PsychoPy (v1.82) experiment, sequence generation script
	 Stimulus files (PCM WAVE, 16 bit, 44100 Hz stereo)
	 Extracted audio features and script

`SUBJECT/behav/task002_run00[1-8]/behavdata.txt` `SUBJECT/model/model001/onsets/task002_run00[1-8]/cond00[1-5].txt`	**Behavioral log files** (CSV format) Stimulation timing specifications (FSL EV3 format)
`SUBJECT/physio/task002_run00[1-8]/[physio.txt.gz,conversion.log]`	Physiological data and conversion log file

`SUBJECT/BOLD/task002_run00[1-8]/`
 `bold[_dico].nii.gz`
 `moco_ref.nii.gz`
 `bold_dico_moco.txt`
 `*dicominfo.txt`
 `bold_dico_bold7Tp1_to_subjbold7Tp1.nii.gz`
 `qa/bold_dico_fgmask_bold7Tp1_to_subjbold7Tp1.nii.gz`
 `qa/bold_dico_moest_bold7Tp1_to_subjbold7Tp1.txt`
	BOLD fMRI data
	 Raw and distortion-corrected BOLD fMRI
	 Motion correction reference scan
	 Motion estimates (3x translation in mm, 3x rotation in deg)
	 DICOM meta data dump from MRIConvert
	 BOLD fMRI aligned to per-subject template (6 DoF; FLIRT)
	 Foreground voxel mask (across time series)
	 Motion estimates with respect to per-subject template (MCFLIRT)

`SUBJECT/templates/bold7Tp1/`
 `[head,brain,brain_mask].nii.gz`
 `in_grpbold7Tp1/[head,brain_mask].nii.gz`
 `in_grpbold7Tp1/[subj2tmpl,tmpl2subj]_warp.nii.gz`
 `qa/fgmasks_bold7Tp1_to_subjbold7Tp1.nii.gz`
 `qa/jointfgmask_bold7Tp1_to_subjbold7Tp1.nii.gz`
 `qa/jointfgbrainmask_bold7Tp1_to_subjbold7Tp1.nii.gz`
 `qa/lvl2/[head,head_avgstats,aligned_head_samples].nii.gz`
	7T BOLD fMRI per-subject template (all phase 1 data)
	 Average head and skull-stripped image, plus mask
	 Per-subject template in group template space
	 FNIRT warps from/to group BOLD template space
	 4D image with foreground masks across all 7T BOLD fMRI acq.
	 Intersection of above image across volumes
	 Intersection of above image with template brain mask
	 Last iteration template image, overlap stats, and aligned samples prior cropping

`templates/grpbold7Tp1/`
 `[head,brain].nii.gz`
 `xfm/[mni2tmpl,tmpl2mni]_12dof.mat`
 `in_mni/brain_12dof.nii.gz`
 `from_mni/MNI152_T1_1mm.nii.gz`
 `from_mni/avg152T1_[brain,csf,gray,white].nii.gz`
 `from_mni/MNI152_T1_1mm_brain_mask[_dil].nii.gz`
 `qa/lvl4/[head,brain].nii.gz`
 `qa/lvl4/[head_avgstats,aligned_head_samples].nii.gz`
 `qa/subjbold7Tp1_to_grpbold7Tp1/brain_mask_[intersection,stats].nii.gz`
 `qa/subjbold7Tp1_to_grpbold7Tp1/aligned_brain_samples.nii.gz`
	7T BOLD fMRI group template (all phase 1 data)
	 Average head and skull-stripped image
	 Affine transformation to/from MNI152 (FLIRT)
	 Template transformed and re-sliced into MNI152 (1mm)
	 MNI152 template from FSL in group template space
	 Template generation overlap stats, aligned samples
	 Per-subject template alignment quality control files

Table 2. Overview of known data anomalies (F: functional data, P: physiological recordings during fMRI session).

Modality	Participant	Run	Description
P	1,2	1–8	sampling rate is 100 Hz
F	2	2	significant movement (translation) during scan
F	5	3,7	experiment had to be restarted after the second run; due to technical limitations sequences for the first two runs were repeated as run 4 and 7 (see Figure 2B)
F	5	4–7	significant movement (rotation) during scan
F	8	6–8	significant movement (rotation) during scan
F	10	5–7	significant movement during scan
F	11	4–8	significant movement (translation) during scan
F	13	7–8	significant movement (translation) during scan
P	18	6	accidental data acquisition stop during the run
F	19	3–4	significant movement (translation) during scan
F	20	2	significant movement (translation) during scan
F & P	20	5–8	no data; participant aborted experiment after four runs

Low-Quefrency and High-Quefrency Mel-Frequency Spectrum (lq_mfs, hq_mfs). Although proven to be useful in machine classification tasks, cepstral coefficients are in a different domain than the spectrum. The last two features map selected cepstral coefficients back to the spectrum domain by reconstructing the 48 mel-frequency spectrum bands using the low-quefrency and high-quefrency mfcc coefficients respectively. In each case, the non-selected coefficients were zeroed and the resulting feature mapped back to the spectral domain using the inverse (transposed) DCT-II matrix and then inverting the decibel amplitude scale. These two sets of features represent broad-spectrum information (timbre) and fine-scale spectral structure (pitch) respectively. The product of these two spectra yields the mel-frequency spectrum.

Source code
The source code for descriptive statistics in Figure 1 and Figure 3, as well as the implementation for the analysis presented in Figure 4 is available in a Git repository at https://github.com/psychoinformatics-de/paper-f1000_pandora_data. Source code for the implementation of the stimulation paradigm and audio feature extraction are included in the data release. Additional scripts for data conversion and quality control are available at: https://github.com/hanke/gumpdata.

Dataset validation
In order to assess data quality, we investigated whether different BOLD response patterns associated with the five musical genres could be discriminated, using either univariate statistical parametric mapping or multivariate pattern (MVP) classification accuracy (searchlight-based analysis, radius of two voxels, sparse spatial sampling with sphere-centers spaced by two voxels, leave-one-run-out cross-validated classification analysis with a support vector

machine, accuracy mapped on a voxel reflects the average across all sphere-analysis a voxel participated in). Inspection of the participant motion estimates revealed a median translation of less than a voxel size, and a maximum rotation of about 1 deg (see Figure 3 for outliers).

Despite the variable magnitude of motion, no participant was excluded from the subsequent analysis.

The results of the univariate analysis (Figure 4A) and the MVP analysis (Figure 4B and Table 3) identify largely congruent areas. MVP analysis generally detects larger and more numerous areas, either due to higher sensitivity or a comparably more liberal statistical threshold. Notably, clusters of above-chance classification accuracy not only contain auditory cortex and other cortical fields related to speech and music processing, but also the subcortical bilateral medial geniculate bodies, a neural relay station immediately prior to the primary auditory cortex in the auditory pathway[15].

Given the confirmed wide-spread availability of genre-discriminating signal we conclude that these data are suitable for studying the representation of music and auditory features. Table 2 contains a list of all known data anomalies that may help potential data consumers to select appropriate subsets of this dataset.

Usage notes
These data are part of a larger public dataset available at http://www.studyforrest.org. The website includes information on all available resources, data access options, publications that employ this dataset, as well as source code for data conversion and data processing.
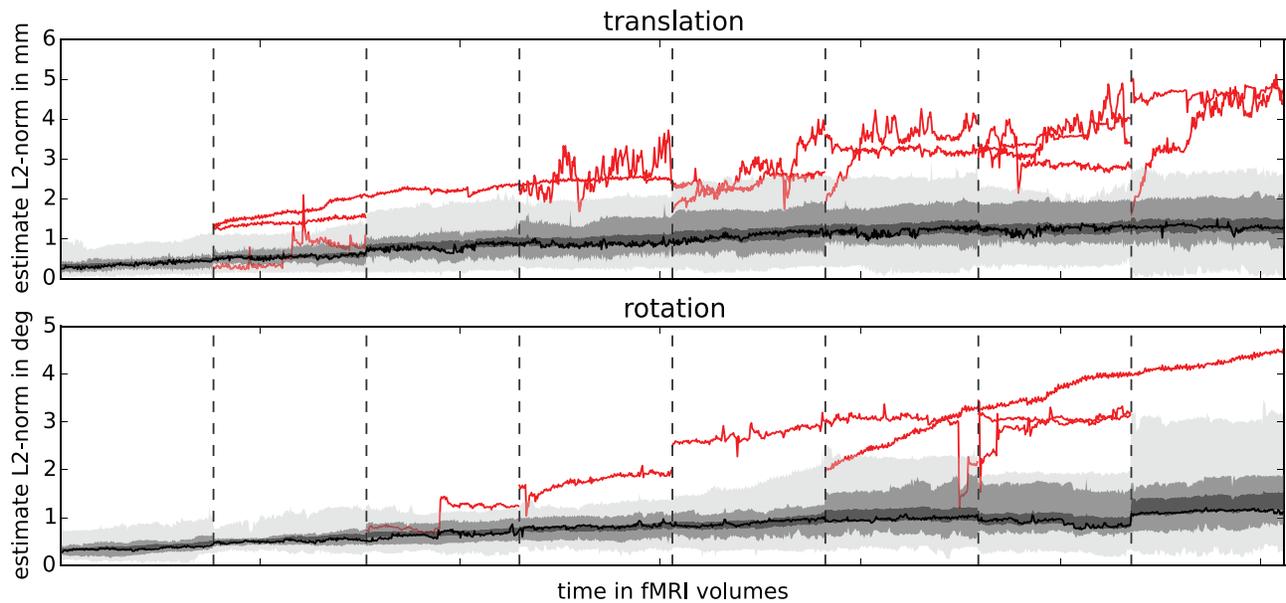
Figure 3. Summary statistics for head movement estimates across runs and participants. These estimates indicate relative motion with respect to a dedicated reference scan at the beginning of each scan session. The area shaded in light gray depicts the range across participants, while the medium gray area indicates the 50% percentile around the mean, and the dark gray area shows ± one standard error of the mean. The black line indicates the median estimate. Dashed vertical lines indicate run boundaries where participants had a brief break. The red lines indicate motion estimate time series of outlier participants. An outlier was defined as a participant whose motion estimate exceeded a distance of two standard deviations from the mean across participants for at least one fMRI volume in a run. For a breakdown of detected outliers see Table 2.
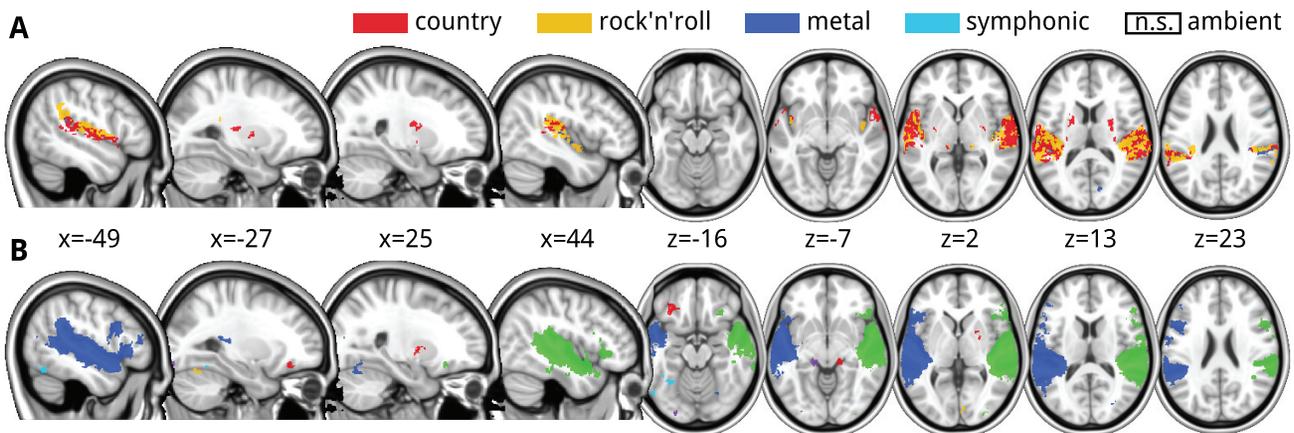

Figure 4. Localization of genre-discriminating signals in the brain. (**A**) Voxel-wise genre-selectivity label. Random-effects GLM group analysis (n=20) were computed using the FEAT component of FSL[16]. Individual contrasts were evaluated for each genre to identify voxels showing a BOLD response to this particular genre that is larger than the average response to all other genres. For all voxel clusters that show a significant difference at the group-level (cluster forming threshold $Z=3.1$, cluster probability threshold $p<0.05$) for any genre, the selectivity label was determined by the maximum Z statistic across all genres. No significant selective activation was found for the *ambient* genre. The majority of all voxels were labeled selective for one of the musical genres where stimuli contained vocals (country, rock'n'roll, heavy metal). Only a small cluster in BA44 R (Broca's area) was labeled selective for *symphonic* music, despite the lack of speech content in these stimuli. (**B**) For comparison, the location of voxel clusters with above-chance classification accuracy for predicting the genre of a music stimulus (colors only indicate individual clusters, not association with particular genres). The associated areas are largely overlapping with the results of the GLM analysis. However, genre-discriminating signals were identified in a number of additional areas. For details on the MVP analysis and cluster statistics see Table 3. Unthresholded maps for GLM and MVP analyses are available at NeuroVault.org[17] collection 308.

Table 3. Average group results of a searchlight-based (radius ≈2.5 mm) cross-validated within-subject musical genre classification analysis (n=20; SVM classifier; C parameter scaled according to the norm of the data). The table lists statistics (size, mean/max/std accuracy) as well as localization information (coordinates in mm MNI152) for clusters with above-chance classification performance in the group (cluster-level probability $p<0.05$; FWE-corrected). Clusters are depicted in Figure 4B. Statistical evaluation was implemented using a bootstrapped permutation analysis, as described by Stelzer and colleagues[18] and implemented in PyMVPA[19], using 50 permutation searchlight accuracy maps per subject, 10000 bootstrap samples, voxel-wise cluster forming threshold of $p<0.001$). Apart from two large clusters covering the majority of bilateral area for auditory perception and speech processing, additional clusters with genre-discriminating signals were identified. These include the bilateral medial geniculate bodys, as well as smaller regions on the ventral visual pathway, frontal orbital cortex, and the cerebellum. For these regions the NeuroSynth database[20] reports high posterior probabilities for the topics: *counting, motor, naming, phonology, prosody, visual, and vocal* (as determined with the Neurosynth term atlas shipped with NeuroDebian[8]).

#	voxels	max	max location (MNI)			mean	std	center of mass (MNI)			$p_{corr.}$	structure
			X	Y	Z			X	Y	Z		
1	36099	0.53	-58.0	-3.9	-0.1	0.31	0.071	-52.1	-20.2	4.7	0.0006	L sup. temporal, L Broca's area, L front. operculum
2	34451	0.52	59.5	0.0	-5.4	0.31	0.071	53.9	-17.8	1.7	0.0006	R sup. temporal, R Broca's area, R front. operculum
3	320	0.26	-26.5	32.5	-14.5	0.24	0.008	-29.9	32.3	-15.7	0.0142	L front. orbital
4	259	0.25	-24.8	-66.3	-19.4	0.24	0.007	-22.7	-65.5	-20.6	0.0142	L cerebellum
5	240	0.25	-40.7	-45.0	-16.8	0.24	0.005	-35.3	-51.7	-16.6	0.0142	L temporal occ. fusiform
6	227	0.27	28.0	-63.2	-22.7	0.24	0.007	25.3	-64.2	-18.1	0.0142	R cerebullum, R temp. occ. fusiform
7	227	0.26	33.5	-86.9	-3.1	0.24	0.005	33.6	-88.5	-0.3	0.0142	R lat. occipital
8	215	0.28	14.4	-30.3	-4.2	0.25	0.011	14.8	-29.9	-6.4	0.0145	R medial geniculate body
9	200	0.28	-13.2	-31.0	-7.8	0.24	0.011	-14.1	-30.5	-6.6	0.0159	L medial geniculate body
10	178	0.26	-50.0	-63.1	-19.4	0.24	0.011	-51.0	-64.4	-18.3	0.0194	L temporooccipital
11	152	0.25	31.1	-79.1	9.3	0.24	0.004	32.6	-79.0	9.6	0.0268	R lat. occipital
12	144	0.27	28.8	28.7	-14.8	0.25	0.007	27.4	27.7	-14.5	0.0280	R front. orbital
13	124	0.25	24.8	-1.2	4.0	0.24	0.003	23.2	0.4	3.2	0.0387	R putamen
14	111	0.25	-25.5	-89.8	-15.2	0.24	0.006	-26.2	-90.0	-13.2	0.0477	L V4
15	107	0.25	7.1	-83.4	3.3	0.24	0.005	6.4	-86.1	1.8	0.0488	R V1

Data availability

OpenFMRI.org: High-resolution 7-Tesla fMRI data on the perception of musical genres: ds000113b[21]

ZENODO: Article sources for 7-Tesla fMRI data on the perception of musical genres, doi: 10.5281/zenodo.18767[22]

ZENODO: "Forrest Gump" data release source code, doi: 10.5281/zenodo.18770[23]

Consent

Written informed consent for publication of acquired data in a de-identified form was obtained from all participants.

Author contributions

Competing interests

Grant information

Acknowledgements

References

1. Hanke M, Baumgartner FJ, Ibe P, *et al.*: **A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie.** *Sci Data.* 2014; **1**: 140003.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Labs A, Reich T, Schulenburg H, *et al.*: **Portrayed emotions in the movie "Forrest Gump" [v1; ref status: indexed, http://f1000r.es/55o].** *F1000Res.* 2015; **4**: 92.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Casey M, Thompson J, Kang O, *et al.*: **Population Codes Representing Musical Timbre for High-Level fMRI Categorization of Music Genres.** In *Mach Learn Interpret Neuroimaging.* Springer Science and Business Media. 2012; **7263**: 34–41.
   Publisher Full Text

4. Haxby JV, Guntupalli JS, Connolly AC, *et al.*: **A common, high-dimensional model of the representational space in human ventral temporal cortex.** *Neuron.* 2011; **72**(2): 404–416.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Hanke M, Halchenko YO: **A communication hub for a decentralized collaboration on studying real-life cognition [v1; ref status: not peer reviewed, http://f1000r.es/55n].** *F1000Res.* 2015; **4**: 62.
   PubMed Abstract | **Publisher Full Text**

6. Aguirre GK, Mattar MG, Magis-Weinberg L: **de Bruijn cycles for neural decoding.** *Neuroimage.* 2011; **56**(3): 1293–1300.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Peirce JW: **PsychoPy--Psychophysics software in Python.** *J Neurosci Methods.* 2007; **162**(1–2): 8–13.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Halchenko YO, Hanke M: **Open is Not Enough. Let's Take the Next Step: An Integrated, Community-Driven Computing Platform for Neuroscience.** *Front Neuroinform.* 2012; **6**: 22.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. In MH, Speck O: **Highly accelerated PSF-mapping for EPI distortion correction with improved fidelity.** *MAGMA.* 2012; **25**(3): 183–192.
   PubMed Abstract | **Publisher Full Text**

10. Toiviainen P, Alluri V, Brattico E, *et al.*: **Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data.** *Neuroimage.* 2014; **88C**: 170–180.
    PubMed Abstract | **Publisher Full Text**

11. Bogdanov D, Wack N, Gómez E, *et al.*: **Essentia: an open-source library for sound and music analysis.** In *Proc of the 21st ACM Int Conf Multimed.* ACM, 2013; 855–858.
    Publisher Full Text

12. Ganchev T, Fakotakis N, Kokkinakis G: **Comparative evaluation of various MFCC implementations on the speaker verification task.** In *Proceedings of the SPECOM.* 2005; **1**: 191–194.
    Reference Source

13. Logan B, Salomon A: **A music similarity function based on signal analysis.** In *Proc IEEE Int Conf Multimed Expo.* IEEE, 2001; 745–748 .
    Publisher Full Text

14. Müller M, Ewert S, Kreuzer S: **Making chroma features more robust to timbre changes.** In *Acoust Speech Signal Proc, 2009. ICASSP 2009. IEEE Int Conf on.* IEEE, 2009; 1877–1880.
    Publisher Full Text

15. Krumbholz K, Schönwiesner M, Rübsamen R, *et al.*: **Hierarchical processing of sound location and motion in the human brainstem and planum temporale.** *Eur J Neurosci.* 2005; **21**(1): 230–238.
    PubMed Abstract | **Publisher Full Text**

16. Smith SM, Jenkinson M, Woolrich MW, *et al.*: **Advances in functional and structural MR image analysis and implementation as FSL.** *Neuroimage.* 2004; **23**(Suppl 1): S208–19.
    PubMed Abstract | **Publisher Full Text**

17. Gorgolewski KJ, Varoquaux G, Rivera G, *et al.*: **NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain.** *Front Neuroinform.* 2015; **9**: 8.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Stelzer J, Chen Y, Turner R: **Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control.** *Neuroimage.* 2013; **65**: 69–82.
    PubMed Abstract | **Publisher Full Text**

19. Hanke M, Halchenko YO, Sederberg PB, *et al.*: **PyMVPA: A Unifying Approach to the Analysis of Neuroscientific Data.** *Front Neuroinform.* 2009; **3**: 3.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Yarkoni T, Poldrack RA, Nichols TE, *et al.*: **Large-scale automated synthesis of human functional neuroimaging data.** *Nat Methods.* 2011; **8**(8): 665–670.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Hanke M, Dinga R, Häusler C, *et al.*: **High-resolution 7-Tesla fMRI data on the perception of musical genres – an extension to the *studyforrest* dataset.** *OpenfMRI.* 2015.
    Data Source

22. Hanke M, Dinga R, Häusler C, *et al.*: **Article sources for 7-Tesla fMRI data on the perception of musical genres.** *Zenodo.* 2015.
    Data Source

23. Hanke M: **"Forrest Gump" data release source code.** *Zenodo.* 2015.
    Data Source

Open Peer Review

Current Referee Status: [?]   [?]

[?]   **Karsten Müller**
Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

The paper presents a very valuable functional MRI (fMRI) data set, investigating brain function on the perception of musical genres. It is a very interesting study with music perception of the human brain. The paper is well written and the data presentation is in an excellent fashion. This presentation will allow a further usage of the data even by other scientists.

1. Although it's a 'data paper', I would suggest including at least a few sentences about the motivation of the study. Why is it interesting to study the presented type of stimuli? In which context would it fit into the current state of research?

2. MRI acquisition protocol and subject selection is referred to in a previous paper. As the other reviewer suggests, I would also prefer to read acquisition parameters and subject data in this paper because it plays a central role.

3. As the current manuscript is a data paper, I would suggest to include a characterization of the MRI data quality. I would be happy to see noise parameters of the BOLD images over time. Is the noise stable? It should be possible to present signal-to-noise and contrast-to-noise values within a subject over time but also between subjects. This would be very helpful to characterize the data set. I assume that the numbers are excellent compared with 3T data.

4. There's nothing written about normalization quality. Would it be possible to demonstrate that the normalization was successful for all subjects? It would be great to read some numbers reflecting a normalization quality assessment.

5. I further suggest to present the individual foreground voxel mask images. When they are normalized, the masks can be shown in a single image with different colors showing the number of subjects having foreground in a selected voxel.

6. Are the audio files available? It would be a good idea to offer the stimulus material.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 28 July 2015

Cristiano Micheli[1], **Jochem Rieger**[2], **Kirsten Weber**[3]

[1] Applied Neurocognitive Psychology Lab, Department of Psychology, School of Medicine and Health Sciences, University of Oldenburg, Oldenburg, Germany
[2] Neurocognitive Psychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany
[3] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

General comments
I am not an fMRI expert myself; therefore I deemed it necessary to invite two co-reviewers to add comments on the manuscript. Please find a set of suggested changes hereafter.

The manuscript describes a recent rich amendment of an existing fMRI open dataset (documented in studyforrest.org), aimed at increasing the amount of shared data from the same subjects, stimuli and procedures for 7T fMRI slab recordings, and 3T whole-head anatomical scans, but with a different cognitive task consisting of listening to music in 5 different genres (vocalized or not) and answering a dual-choice question ('did the clip have a happy melody?' [yes/no]). In addition, the dataset includes recordings of several physiological parameters.

The dataset is extremely rich and very well documented and as such has the potential to be used as a reference dataset for various research questions from auditory to music and emotional processing. The choice of freely available software packages for data processing and the inclusion of the processing scripts greatly enhance reproducibility and augment the documentation of the processing steps and analyses provided in the paper. The idea of an open-source sharing framework for replication of results and follow up projects is greatly appreciated, together with the suggestion of having fMRI data available for benchmarking.
The authors describe the experimental procedures and the data processing techniques that led to a pre-processing of the data, with the goal to run a preliminary quality check and validation on the material that will afterwards be available for future studies (the authors name this part 'quality control'). Secondarily, the authors provide a pre-processing of the stimuli set, too, with the aim to provide elementary descriptors which could be used to relate fMRI brain responses to delivered music snippets for encoding models estimation. Thirdly, the authors mention the goal to validate the encoding models generated by the previous dataset (studyforrest.org). Alongside the main goals, the authors mention the possibility to use the latest fMRI corpus as a resource for benchmarking algorithms for functional alignment. Could the authors make examples of such algorithms, ideally with references?

Firstly, in my opinion the stimuli and the experimental procedures are generally clearly described, but the pre-processing of fMRI analyses could be further commented in order to allow appropriate replication. In fact, some procedures are mentioned in other papers and that would require the knowledge of previous

work done by the group. To ease the readers into the topic could the authors add one/two sentences for each procedure or method that is relevant for the points made in the present manuscript? For example at page 3 a landmark-based procedure is mentioned. Could the authors briefly mention what was used as a landmark and which software or metric allowed this? Additionally, one shortcoming of the dataset is the lack of a characterization of the audio presentation. The transforms were computed on the source waveforms. However, it is not so clear what waveform arrives at the ear and how the gain adjustments scaled them. This information would be important for the investigations of neural sound encoding with this dataset. In addition, this information would be of great use for users who intend to calculate other transforms on the source waveforms.

Secondarily, while the authors' premises are valid in only providing a quick and preliminary analysis of the data as a quality check, it would be helpful to briefly introduce the research context behind it. A couple of sentences in the introduction would be sufficient to explain the expected activated networks for those readers that are unfamiliar with neuroimaging studies on music.

Finally, it would be helpful to briefly outline which ones of the many tools are made available in the paper directly, and which ones are preparatory for future studies. For example the computation of cepstral coefficients is often used in speech recognition, to separate vocal tract from speech descriptors and to create features for recognition, but is not clear if and how they are used in the current manuscript. We are aware of the amount of work carried out by the authors, in documenting and making the material available, however the achieved goals should not be mixed with future ones. Please amend in order to outline the current state of the research framework to the general reader.

Detailed comments:
- Abstract. What is the advantage of a slow event-related paradigm (long ISI) for this experiment? Please comment on it in the text.

- p.4. Some more information about the acquisition procedures and the procedures applied to the preprocessed dataset should be provided, or at least stated where they can be found (e.g. scripts in the dataset). For example a short description of the FOV, (e.g. covering temporal and inferior frontal areas of the brain) may be added in the article itself.

- p.5. See previous comment. A better description of the physiological recordings is needed. Are they acquired with standard Siemens sensors?

- p.5. Preprocessing of fMRI data. Similarly to the previous comments, it is our opinion that the authors should thoroughly describe the set of steps that lead to the subsequent analyses, despite them being already documented in other papers. Critically, having missed on this point will not facilitate the reader in the choice among three different flavors of BOLD datasets at disposal. Please add more operative details, a flow chart figure or indicate where to find them.

- For example, a distortion correction technique is mentioned, along with a realignment of the fMRI volumes. A template image is also mentioned. How is the template generated? Is there a slice-time correction step? Is the distortion correction applied before, after or during the realignment?

- p.5. In the description of the mel coefficients could the authors briefly explain the difference between DCT_II and the other flavors of the DCT algorithm (or cite a representative reference)?

- p.7. What is the contrast of the univariate analysis? I understand that this is a standard analysis, but in our opinion its description is missing some crucial information on how it was performed. For example:

  a) I'd like some basic information on how the validation analyses were done: is the source code for the classification provided too?

  b) Given this statement "Given the confirmed wide-spread availability of genre-discriminating signal we conclude that these data are suitable for studying the representation of music and auditory features." it would be good to have a brief overview of previous studies (with references) in this field and the brain areas they usually associate with music processing + processing of different genres to facilitate understanding for readers without a background in the neurobiology of music.

p.8. Figure 4 is a bit confusing. As I understand it the color code on top only refers to Figure 4a. Maybe label a) statistical analysis and b) classification analysis and use different colors in b. The legend refers to a small symphonic music cluster in RIFG which I cannot identify on the actual brain images.

Minor comments
- Abstract and p.2. The term 'quality control' might be slightly misleading. One could just state that basic GLM and MVPA analyses were conducted as a 'proof of principle' of effective classification on the data.

- p.3 Correct the sentence 'continuum of research question' into '… questions'

- p.3 Participants: which of the functional scans was done first?

- p.4 Were the sound adjustments done only once at the beginning of the experiment or for every functional run? Is the level of the sound adjustment documented in the material?

- p.5 Correct the sentence 'flag whether a control question with presented' into '…was presented'

- p.5 "As the exact same alignment target was used, this led to a very similar field-of view configuration across acquisitions." Somewhat unclear. Does this mean that the subjects' head was in the same position in both experiments (i.e. voxels match functionally) or does it mean that the FOV was in the same scanner's coordinates? It is unlikely that both conditions are met simultaneously.

- p.5 "While raw BOLD data are suitable for further analysis, they suffer from severe geometric distortions. BOLD data that have been distortion-corrected9 at the scanner console are provided in bold_dico.nii.gz."

- Please provide more information about the correction (e.g. k-space or voxel space correction). Is it documented? Is all information that was used available in the dataset? Users may want to apply their own correction

- p.5 "A log file of the automated conversion procedure is provided in the same directory (conversion.log)". Conversion of what?

- p.5 "Audio features" Are the sound files provided with the dataset? You might want to state that here. Otherwise state how the user could obtain them to apply their own transforms. It would be

important to characterize the sound presentation system. The transforms were calculated on the sound files but it is not so clear what actually reached the ear and how the genres might have differed in energy etc. This is information could greatly extend the usefulness of this rich dataset.

- p.5 "frequency domain filter bank" Are the parameters provided? If yes, mention that here.

- p.5 The description of the audio features reports a highest signal frequency of ~22KHz but the signals in figure 1 show a clear decrease in gain for all stimuli after ~16KHz. Could the authors document what the sampling frequency of the audio waves was and if a low pass filter was applied?

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.