Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Electroencephalography (EEG) dataset during naturalistic music listening comprising different genres with familiarity and enjoyment ratings

Krishna Prasad Miyapuram[a,*], Nashra Ahmad[a], Pankaj Pandey[a], James Derek Lomas[b]

[a] *Brain and Informatics Lab[1], Centre for Cognitive and Brain Sciences, Indian Institute of Technology Gandhinagar, India*

[b] *Department of Human Centered Design, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands*

ARTICLE INFO

Article history:
Received 11 March 2022
Revised 26 August 2022
Accepted 28 September 2022
Available online 8 October 2022

Dataset link: Music Listening- Genre EEG dataset (MUSIN-G) (Original data)

Keywords:
Music listening
EEG
Familiarity to music
Enjoyment to music
Music genre dataset

ABSTRACT

The article provides an open-source Music Listening- Genre (MUSIN-G) EEG dataset which contains 20 participants' continuous Electroencephalography responses to 12 songs of different genres (from Indian folk music to Goth Rock to western electronic), along with their familiarity and enjoyment ratings. The participants include 16 males and 4 females, with an average age of 25.3 (+/-3.38). The EEG data was collected at the Indian Institute of Technology Gandhinagar, India, using 128 channels Hydrocel Geodesic Sensor Net (HCGSN) and the Netstation 5.4 data acquiring software. We provide the raw and partially preprocessed data of each participant while they listened to 12 different songs with closed eyes. The dataset also contains the behavioural familiarity and enjoyment ratings (scale of 1 to 5) of the participants for each of the songs. In this article, we further discuss the preprocessing steps which can be used on the dataset and prepare the data for analysis, as in the paper [1].

* Corresponding author.

E-mail addresses: kprasad@iitgn.ac.in (K.P. Miyapuram), ahmad_nashra@iitgn.ac.in (N. Ahmad), pankaj.p@iitgn.ac.in (P. Pandey), J.D.Lomas@tudelft.nl (J.D. Lomas).

Social media: @nashraahmad_ (N. Ahmad)

[1] @pfuturism

https://doi.org/10.1016/j.dib.2022.108663

Specifications Table

Subject	Neuroscience: Cognitive
Specific subject area	Music Cognition, Music Information Retrieval, Cognitive Neuroscience, Musicology, Non-Invasive Brain Imaging
Type of data	Table: 1, Order of Song presentation for each participant Table: 2, Experimental Design and EEG markers information Table: 3, Song details and behavioural ratings Figure: 1, Behavioural ratings (familiairty)
How the data were acquired	EEG data was acquired on a 128 channel high-density Hydrocel Geodesic Sensor Net (HCGSN) by Magstim EGI, the data acquiring software was Netstation 5.4, the experiment was presented to the subjects using the Eprime software with an extension to Netstation, the songs were played using speakers, and the behavioural ratings were taken through the keyboard's number pad (1 to 5).
Data format	Raw (.set file) and partially preprocessed (.set) EEG Data in BIDS format, with behavioural ratings of familiarity and enjoyment (text file).
Description of data collection	The EEG responses were collected on 12 different songs, while the participants sat on a comfortable chair in a dimly lit room. The experiment included a two-minute silence, followed by song listening and rating of familiarity and enjoyment. We used the Harman Kardon Soundsticks with Bluetooth 20 W Portable Bluetooth Laptop/Desktop Speaker (Transparent, 2.1 Channel) for presentation of the audio. The participants kept their eyes closed and were indicated to open them to rate each song. The sampling rate was kept at 250 Hz for some participants. Certain participants had the initial sampling rate of 1000 Hz which was downsampled to 250 Hz prior to data analysis.
Data source location	• Institution: Indian Institute of Technology Gandhinagar (IITGN) • City/Town/Region: Gandhinagar, Gujarat • Country: India • Latitude and longitude for collected data: 23.2114° N, 72.6842° E
Data accessibility	Data is hosted on a public repository Repository name: OpenNeuro Data identification number: OpenNeuro Dataset ds003774 Direct URL to data: https://doi.org/10.18112/openneuro.ds003774.v1.0.2
Related research article	[1] P. Pandey, N. Ahmad, K. P. Miyapuram and D. Lomas, "Predicting Dominant Beat Frequency from Brain Responses While Listening to Music," 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021, pp. 3058-3064, doi: 10.1109/BIBM52615.2021.9669750.

Value of the Data

- This dataset [2] involves EEG responses to naturalistic music listening or listening to everyday songs. The uniqueness of this open-source Music Listening - Genre (MUSIN-G) EEG dataset lies in its stimuli which include 12 songs of genres varying from Indian folk music to Goth Rock to western electronic.
- The enjoyment of musical stimuli is shaped by familiarity and preferences for melody, beat, pitch, harmonicity, etc. Hence, we considered two subjective factors: familiarity with the song and enjoyment while listening to it, as they have been highly associated with changes in the neural responses [3,4].
- As our dataset contains naturalistic music and is rich in different music features, it can further also be used by studies on Music Information Retrieval (MIR), involving extraction of specific features from brain responses to music. Studies can further analyse different features from the dataset to understand specific neural entrainment of naturalistic music. Studies on

machine learning and neural networks could also utilize this dataset [5]. This dataset would be of interest to researchers from different fields, especially from musicology, neuroscience and computer science.

- This dataset will further provide the base for music studies and the development of other similar experiments utilizing physiological and brain imaging data on naturalistic music.

1. Data Description

The Raw data, (named sourcedata) [2] contains the EEG data from each of the 20 participants for the entire experiment including a 2-minute baseline followed by EEG responses to all the 12 songs presented in randomised order as shown in Table 1. Each song was presented once to every participant. The entire timeline, procedure, and event markers (in EEG data) of the experiment can be found in Table 2.

The partially preprocessed data [2] contains EEG recordings and behavioural responses on each of the 12 songs for each of the 20 participants. The behavioural data [2] contains responses, taken while the participants pressed one key from 1 to 5 for each of the presented songs on the keyboard, to indicate familiarity and enjoyment to each song once. The keypress of 1 indicates the highest enjoyment or familiarity, and 5 indicates the least familiarity or enjoyment.

Both the raw and segmented data were converted into BIDS format using eeg2bids extension for EEGlab [6]. The dataset contains all the stimuli information in the '*stimuli*' folder, with the description of songs (also provided in Table 3), and the average behavioural data with the enjoyment and familiarity ratings of the participants (Figure 1). The segmented raw EEG data with all the other information of each of the participants is given separately for all the songs. For example, the folder *'sub-001'*, contains the segmented raw EEG data for each of the session/song ('*ses-01'* to '*ses-12*') of Participant 1. The EEG data from the 128 channels for each of the song/session is provided in the '*.set*' file. This file format is native for EEGLAB software [6] running in Matlab.

Table 1
Order of song presentation for different participants.

sub-001	11	1	10	4	5	6	12	8	9	3	2	7
sub-002	8	4	2	6	3	11	10	1	5	7	12	9
sub-003	8	7	3	12	4	1	11	6	2	10	9	5
sub-004	5	7	11	1	12	6	2	9	3	8	10	4
sub-005	6	3	2	10	1	8	7	9	4	11	12	5
sub-006	7	8	4	9	5	1	2	12	11	10	3	6
sub-007	1	7	6	8	12	9	11	5	2	3	10	4
sub-008	7	8	12	2	11	10	6	1	3	5	9	4
sub-009	9	3	8	7	10	5	12	1	6	2	4	11
sub-010	1	7	12	8	5	9	2	6	10	3	11	4
sub-011	7	8	3	12	5	9	11	10	4	1	2	6
sub-012	7	9	3	8	4	11	1	12	2	5	10	6
sub-013	1	9	11	7	2	10	6	12	4	8	5	3
sub-014	1	12	6	7	10	11	5	4	3	9	2	8
sub-015	10	12	2	7	11	5	8	3	9	4	1	6
sub-016	3	1	8	7	9	5	10	4	12	11	2	6
sub-017	2	8	1	5	12	7	9	3	6	11	10	4
sub-018	7	4	6	9	2	11	10	1	5	8	3	12
sub-019	10	3	2	11	1	7	8	9	4	12	6	5
sub-020	12	9	4	1	8	10	6	7	11	3	5	2

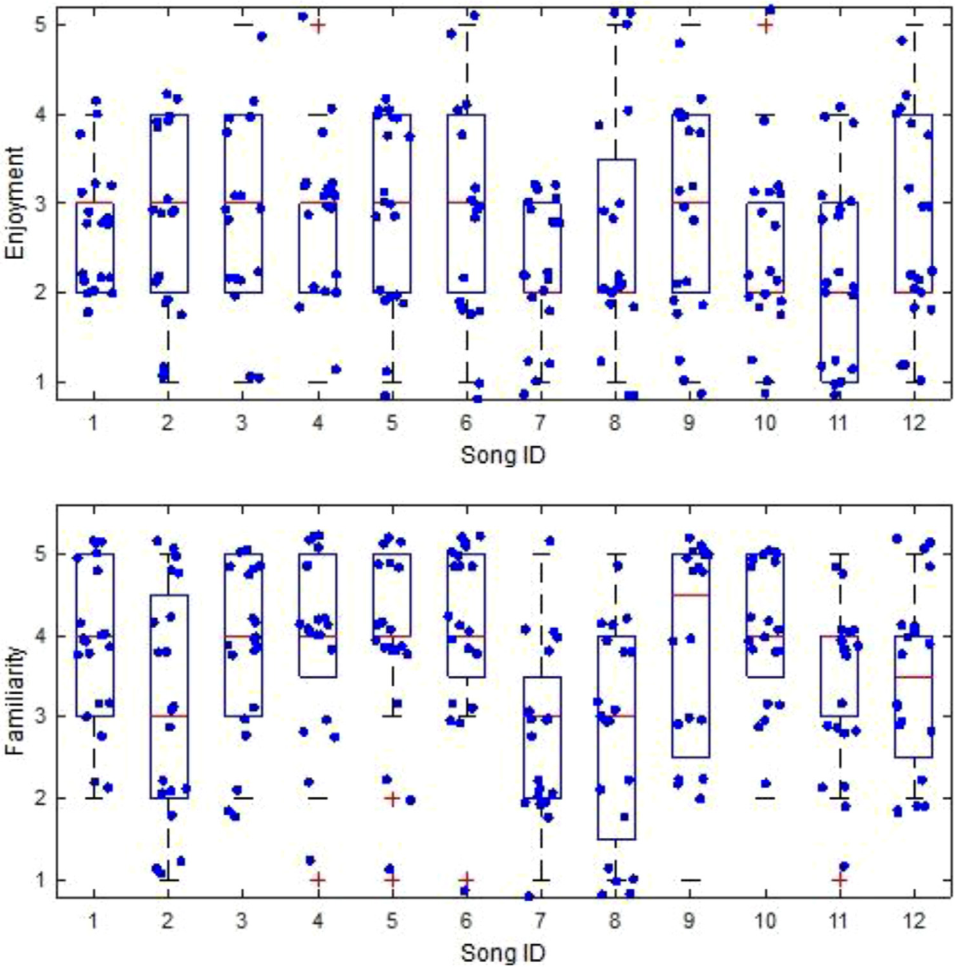

Fig. 1. Behavioural ratings of Enjoyment and Familiarity for each of the songs. The dots represent individual participants responses, which are jittered horizontally and vertically for visualization purposes.

Table 2
Experimental Set up Illustration a) the baseline acquisition and b) the entire experiment. Both the baseline and the experiment were collected one after the other in a single session.

	A) Baseline	
Duration	Object	Marker
1000 ms	Countdown 3	fix3p
1000 ms	Countdown 2	fix2p
1000 ms	Countdown 1	fix1p
500 ms	Close Eyes Beep	clyp
12000 ms	Baseline (silence)	stim
500 ms	Open Eyes Beep	opyp
	B) Music Listening	
1000 ms	Countdown 3	fix3
1000 ms	Countdown 2	fix2
1000 ms	Countdown 1	fix1
500 ms	Close Eyes Beep	clys
10000 ms	Silence	fxcl
Song Duration	Music	stm+
10000 ms	Silence	fxnd
500 ms	Open Eyes Beep	opys
Response Time	Familiarity rating	fam+
Response Time	Enjoyment rating	enjy

Table 3
Representing Song Details (adopted from [1], which can be referred for complete details).

Song No.	Genre	Average Familiarity Response	Average Enjoyment Response
1	Deep House	3.9	2.7
2	Indie	3.2	2.7
3	Electronics	3.9	3
4	New Age	3.9	2.9
5	Electronic Dance	3.9	3
6	Ambient	4.1	3
7	Hindustani Classical	2.8	2.3
8	Indian Semi-Classical	2.8	2.6
9	Indian Folk	3.8	3
10	Soft Jazz	4.1	2.5
11	Goth Rock	3.4	2.2
12	Progressive Instrumental Rock	3.5	2.7

2. Experimental design, materials and methods

2.1. Stimuli

The stimuli consist of naturalistic music, with 12 songs from different genres, selected from across the world. Song information has been provided in Table 3. The songs were very different from each other as they were from varying genres (from Indian folk music to Goth Rock to western electronic), which made the participants experience unrelated songs. Most of the songs did not contain any lyrics, except Song 2 (English), Song 8 (Hindi), and Song 9 (Gujarati). The raw wave files of the songs are provided for research purposes in the code section of the dataset [2].

2.2. Participants

The data was collected on 20 undergraduate and postgraduate Indian students of the Indian Institute of Technology Gandhinagar (IITGN), India. Participants included 16 males and 4 females, with an average age of 25.3 (+/-3.38). Participants had little to no musical training and hence were considered non-musicians. All participants gave their consent to participate in the experiment prior to the start of the experiment. The study was approved by the institutional ethics committee at the Indian Institute of Technology Gandhinagar, India.

2.3. Procedure

The participants sat comfortably in a dark room and were made to wear the 128-channel high-density Electroencephalography (EEG) cap (manufactured by Electrical Geodesics Inc). The participants sat in front of a computer screen with 56cms from the speakers. The speakers used were *Harman Kardon Soundsticks with Bluetooth 20 W Portable Bluetooth Laptop/Desktop Speaker (Transparent, 2.1 Channel).*

The experiment included two parts - the baseline condition included a two-minute silence, followed by a music listening session. In the baseline condition, the participants first heard a single beep, indicating to close their eyes and after a silence of 2 minutes, they were indicated to open their eyes with a double beep. The silence was provided in order to obtain a baseline response, and also to make the participants relaxed before the start of the experiment.

For the music listening session or the main experiment, the participants were again instructed to close their eyes with a single beep. After a gap of 10 seconds, the song was presented to them through a pair of speakers connected to the computer. The order of the song presentation was randomised across participants. After listening to each song, and silence for 10 more seconds, a double beep was presented suggesting they open their eyes followed by answering two questions for each song. The first question asked the participants to rate their familiarity with the song on a scale of 1 to 5. Instructions for rating were: “Press 1’ for extremely familiar, Press ‘2’ for familiar, Press ‘3’ for neutral, Press ‘4’ for unfamiliar, Press ‘5’ for extremely unfamiliar”. The second question asked them to rate their enjoyment of the song on a scale of 1 to 5, where 1 indicated extremely enjoyable and 5 indicated extremely unenjoyable song. The instructions for rating were: “Press ‘1’ for extremely enjoyable, Press ‘2’ for enjoyable, Press ‘3’ for neutral, Press ‘4’ for unenjoyable, Press ‘5’ for extremely unenjoyable.”

Participants listened to each song only once and provided their behavioural rating. Table 3 provides the average behavioural ratings from the participants for both familiarity and enjoyment for each song (Figure 1). EEG responses were collected throughout the experiment using Netstation recording software. The behavioural ratings for each participant and song are provided in the Stimuli directory of the dataset [2].

Stimulus presentation and behavioural ratings of the participants were controlled through E-prime presentation software. The markers for various events as shown in Table 2 were recorded along with the EEG data using E-prime extensions for Netstation.

2.4. Example usage of EEG data

The dataset [2] provides raw EEG data and segmented EEG data for each song. Data can be preprocessed using the following functions and plugins of EEGLAB toolbox [6] in MATLAB. Pandey et al [1] used the following preprocessing. A high pass linear FIR filter of 0.2 Hz was applied to the data, followed by the CleanLine [7] method for removing 50 Hz of line noise. Filtered data were down-sampled to 250 Hz and subsequently segmented into 12 songs. Bad channels were removed by using spectrum criteria with a standard deviation of 3 for the outlier threshold after extracting trials. Multiple Artifact Rejection Algorithm (MARA) [8], a machine learning-based method to detect and remove artifacts, was used to compute independent components and remove noisy ones. The spherical function of EEGLAB was used to interpolate the removed channels. Lastly, the average reference was applied to the dataset. Code for preprocessing is available from https://github.com/BrainLab-IITGN/Music-Project/tree/master/code. After the preprocessing, the data can be analysed further to correlate with temporal and spectral features of the songs (refer [1] for analysis of Tempo as an example). Sonawane et al. [5] used the time-domain and frequency domain data obtained using spectopo function of EEGLAB [6] for song identification with machine learning algorithms applied with deep learning architectures. Pandey et al. [9] extended this work by demonstrating that music identification is possible with EEG data corresponding to initial snippets during the song listening task. Future work can look at neural entrainment of EEG signals with different music features and participant ratings [10].

Ethics statements

The project was approved by the Institutional Ethics committee of the Indian Institute of Technology Gandhinagar, India. The research has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Music Listening- Genre EEG dataset (MUSIN-G) (Original data) (OpenNeuro).

CRediT Author Statement

Krishna Prasad Miyapuram: Data curation, Conceptualization, Writing – review & editing; **Nashra Ahmad:** Data curation, Software, Writing – original draft, Visualization; **Pankaj Pandey:** Methodology, Software, Writing – review & editing, Visualization; **James Derek Lomas:** Conceptualization, Writing – review & editing.

Acknowledgments

We thank Science and Engineering Research Board (SERB) and PlayPower Labs for supporting the Prime Minister's Research Fellowship (PMRF) awarded to Pankaj Pandey. We thank Federation of Indian Chambers of Commerce & Industry (FICCI) for facilitating this PMRF. We also thank IIT Gandhinagar for providing the Sabarmati Bridge Fellowship to Nashra Ahmad.

References

[1] P. Pandey, N. Ahmad, K.P. Miyapuram, D. Lomas, Predicting dominant beat frequency from brain responses while listening to music, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021, pp. 3058–3064, doi:10.1109/BIBM52615.2021.9669750.

[2] K.P. Miyapuram, P. Pandey, N. Ahmad, B.R. Shiraguppi, E. Sharma, P. Lawhatre, D. Sonawane, D. Lomas, Music listening- genre EEG dataset (MUSIN-G), OpenNeuro (2021), doi:10.18112/openneuro.ds003774.v1.0.0.

[3] M.J. Tramo, Music of the hemispheres, Science 291 (5501) (2001) 54–56, doi:10.1126/science.10.1126/SCIENCE.1056899.

[4] Y. Kumagai, M. Arvaneh, T. Tanaka, Familiarity affects entrainment of EEG in music listening, Front. Hum. Neurosci. 11 (2017) 384, doi:10.3389/fnhum.2017.00384.

[5] D. Sonawane, K.P. Miyapuram, B. Rs, D.J. Lomas, GuessTheMusic: song identification from electroencephalography response, in: 8th ACM IKDD CODS and 26th COMAD, 2021, pp. 154–162, doi:10.1145/3430984.3431023.

[6] A. Delorme, S. Makeig, Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, J. Neuroscie. Methods 134 (1) (2004) 9–21, doi:10.1016/j.jneumeth.2003.10.009.

[7] T. Mullen, CleanLine EEGLAB plugin, Neuroimaging Informatics Toolsand Resources Clearinghouse (NITRC), San Diego, CA, 2012.
[8] S.H Winkler, M. Tangermann, Automatic classification of artifactual ica-components for artifact removal in EEG signals, Behav. Brain Funct. 7 (1) (2011) 1–15.
[9] P. Pandey, G. Sharma, K.P. Miyapuram, R. Subramanian, D. Lomas, Music identification using brain responses to initial snippets, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, May, pp. 1246–1250.
[10] P. Bedmutha, P. Pandey, N. Ahmad, K.P. Miyapuram, D. Lomas. Canonical correlation analysis reveal neural entrainment for each song and similarity among genres. Society for Music Perception and Cognition 2022 Conference, Portland, USA.