Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Comparison of Functional Connectivity and Multivoxel Pattern Analysis for Classification of Musical Genres

Mina Ghoorchian[1], Mohammad-Reza Salehi[2], Hamid Soltanian-Zadeh[3], Gholamreza Zandesh[4]

[1] *Institute for Cognitive Science Studies, Tehran, Iran*
[2] *University of Tehran, Tehran, Iran*
[3] *University of Tehran, Tehran, Iran*
[4] *Iranian Orchestra for New Music, Iran*

Abstract:- Humans often tend to group auditory stimuli into discrete categories, including categories such as animal species, language, musical instruments, and musical genres. Among these classifications, music genre emerges as a common dimension in assessing human music preferences. Neuroimaging studies have reported a correlation between brain activities and musical genres but there is a little study to investigate performance of functional connectivity and multivoxel pattern analysis for the classification of the musical genres.

In this work, we used fMRI data of 5 subject in 6 sessions during listening to different musical genres to find a relation between music genres and functional connectivity and multivoxel pattern analysis (MVPA). To this end, we extracted features from the audio files and explored their relation with brain activity. We obtained promising results for MVPA with 74.3% accuracy but no significant or separable result for functional connectivity. Voxel averaging during the parcellation process may have removed some important information useful for the classification of the musical genres.

Keywords: *Brain Functional Connectivity, FMRI, Musical Genre, Multi-Voxel Pattern Analysis (MVPA)*

1. Introduction

Recently, there has been a lot of interest in the brain basis of music research. When adults listen to a piece of music, they understand it with a very high speed of information processing and most of these processes are done automatically. There is no opportunity for contemplative thinking during the handling of every single detail. The fluency of listening and understanding a segment of music pertain to perceptual learning from previous musical experiences [1].

The research conducted on the progression of music perception and recognition has revealed a demonstrated connection between the cognitive components that are initially formed during early ages and the components that are shaped in response to experience [1]. In other research, it has been shown that the mother's voice is recognized by very young infants, and this phenomenon might stem from the neonatal exposure to the mother's distinctive patterns of pitch and stress accents [2]. In another study, they have examined the features related to music, for example, loudness, in the brain [3]. In another research task, participants were repeatedly stimulated with 25 music clips from different kind genres with and without speech content. Analysis of fMRI data has revealed observable patterns of neural response across brain regions that are recognized to be engaged in auditory and speech processing [4].

In this work, we have explored the relation between musical genres or their important features and functional connectivity and multivoxel pattern analysis (MVPA) to be able to classify musical genres based on the information extracted from resting state fMRI data.

2. Musical Genre

Musical genres are categories that define a particular piece of music. Music follows a certain pattern or common sound, and this is how songs are categorized [5]. Music trends and styles are shaped by history. Various epic events have occurred in the world throughout history and music has been a direct statement or message. As mentioned, there are different genres, but in this research, a number of them have been used, which can be named as follows: Blues, classic, country, disco, Hip-Hop/Rap (Rap can be said to be completely separated from Hip-Hop but some still consider it Hip-Hop), Jazz, Metal, pope, Reggae and rock.

3. Materials

A. Subjects

The Music Genre fMRI Dataset was utilized for our study, which included participants between the ages of 23-33. These individuals had musical experience ranging from 4 to 15 years and possessed normal hearing. They took part in MRI scanning as well as behavioral tests as part of the research. The dataset can be accessed at http://openneuro.org. To evaluate the musical background, a questionnaire was used about the musical instrument that was being taught. A consent form was also taken from all the participants before the test. This experiment was accepted by the ethics and safety committee of the Institute of Information and Communication Technology, Osaka, Japan, by the author of the referenced article [6].

B. Data

The data were collected with a custom 3.0 Tesla Siemens Skyra scanner with a 32-channel coil (TIM Trio; Siemens, Erlangen, Germany) [6]. The subjects were scanned with the following specifications: For functional scanning, 68 interleaved axial slices, thickness of 2.0 mm, echo time (TE) = 30 ms, Repetition time (TR) = 1,500 ms, flip angle (FA) = 62°, voxel size = 2 × 2 × 2 mm3, field of view (FOV) = 192 × 192 mm^2, multi-band factor = 4. A total of volumes for each run = 410, For anatomical reference, TE = 3.26 ms, TR = 2,530 ms, FA = 9°, voxel size = 1 × 1 × 1 mm^3 , FOV = 256 × 256 mm^2 [6].

For volume-based analysis, we employed the default preprocessing pipeline in the SPM12 toolbox. This pipeline includes direct normalization to MNI-space. The SPM12 toolbox can be freely accessed at https://www.fil.ion.ucl.ac.uk/spm/software/spm12.

C. Stimuli and Task

Ten genres including country, reggae, rock, pop, metal, jazz, hip-hop, disco, metal, and blues, which were randomly extracted from the GTZAN music genre data set (http://marsyasweb.appspot.com/download/data_sets/), were used in this work. A total of 540 music data were selected. The total number of experiments was 18 runs: 12 runs for training and 6 runs for testing. Each run contains 40 music clips and lasts 10 minutes. Out of the 540 music clips, 480 were considered for training and 60 for testing in the original paper [6], but we used 6 runs for testing. For data reproducibility, 10 pieces of music were presented four times in each test run. Participants were told to wear headphones (Model S14, Sensimetrics) during the experiment. After the end of the ten rounds of each run, there were a request that participants explained about their physical condition, and in case of feeling tired or sleepy, the participants were given a 1-2-minute break. Following each day's scanning session, participants were queried about whether they experienced any episodes of falling asleep during the scan. Based on participants' self-reports, there were no instances where participants indicated that they fell asleep during the MRI scans. The experiment spanned three days, during which six runs were conducted per day.

4. Methods

One perspectives of the brain research involve the assessment of cognitive function performance in relation to brain activities. In this context, our primary objective is to establish a correlation between the pattern of the brain activities and the genre of music. We do this by using various features commonly extracted from audio files and their relationship with brain functional connectivity patterns. With this analysis, we aim to identify the representations of music genre in brain activities through these features.

The features we consider include a range of audio characteristics, including Mel Frequency Cepstral Coefficients (MFCCs), which is a small set of features summarizing spectral characteristics and models the characteristics of the human voice, Chroma Frequencies, which is a powerful presentation of musical notes in 12 "bins" that correspond to the 12 "semitones" or "chroma" of musical octaves across the entire spectrum, root mean square (RMS), Spectral Centroid, which indicates the center of mass for a sound, calculated as a weighted average of its frequencies, and zero-crossing rate, which is the rate at which a signal changes polarity from positive to negative or vice versa.[7]-[8]

To extract features, we used Librosa library, a Python module for audio signal analysis with a focus on music. All features were extracted in time windows of 1.5 seconds (TR). This analysis was limited to a number of regions of interest (ROI) including STG (22, 41, and 42 of Brodmann atlas) that have been used in previous studies. To extract ROI time series, we utilized SPM and Dpabi toolbox on the preprocessed fMRI data.

We used a multi-layer neural network with L2 regularization to estimate the relevant features based on brain activity. The considered error for the neural network was the mean squared error. For estimation of music genre, we used a multi-class SVM classifier.

We have also investigated the correlation between functional connectivity in the brain and music genre. To calculate functional connectivity, fMRI data was mapped on the AAL atlas after preprocessing by Dpabi toolbox, and functional connectivity was measured using linear correlation.

To explore the potential correlation between brain functional connectivity and music genre, an approach similar to the ANOVA statistical test was employed. At the first step, we calculated the Between Sum of Squares (SSB) and Within Sum of Squares (SSW) using the following formulas (1), (2):

$$SS_{between} = SSB = \sum_{j=1}^{k} n_j(\overline{x_j} - \bar{x})^2 \quad (1)$$

$$SS_{within} = SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \overline{x_j})^2 \quad (2)$$

Next, these values were normalized by dividing them by their degrees of freedom. The results are Between Mean of Squares (MSB) in equation (3) and Within Mean of Squares (MSW) in equation (4).

$$MSB = \frac{SSB}{k-1} \quad (3)$$

$$MSW = \frac{SSW}{n-k} \quad (4)$$

Here, 'k' represents the number of groups (in this case, 10), and 'n' represents the number of observations. The statistic of interest is the ratio of MSB to MSW. Under Gaussian Distribution assumption, this statistic follows an F distribution (ANOVA test) but we used permutation test since our data is functional connectivity. The significance threshold of $p = 0.05$ was set, and because of multiple comparisons (6670 unique functional relationships), we applied the Bonferroni correction.

Furthermore, for investigating the potential relationship between the extracted audio features and functional connectivity, a linear correlation analysis was employed. Similar to the previous test, due to uncertainty about the Gaussian distribution of the data, we applied a permutation test. The significance level of $p = 0.05$ was set, and again, the Bonferroni correction was applied to address multiple comparisons.

All the data processing and analyses in this section were carried out using the MATLAB R2018b.

5. Results

Table 1 contains the average of Mean Square Error (mean of all coefficients for MFCC and Chroma Frequencies) of each mentioned feature estimated by MLP.

Table 1: MSE of estimated features

MFCCs	Chroma	RMS	Spectral Centroid	zero-crossing rate
0.114	0.167	0.093	0.248	0.073

In the following, we reduced the dimensions of the data to two dimensions by PCA for better visualization of the separation of classes. Figure 1 is the scatter plot of the first two coefficients of different music genres.

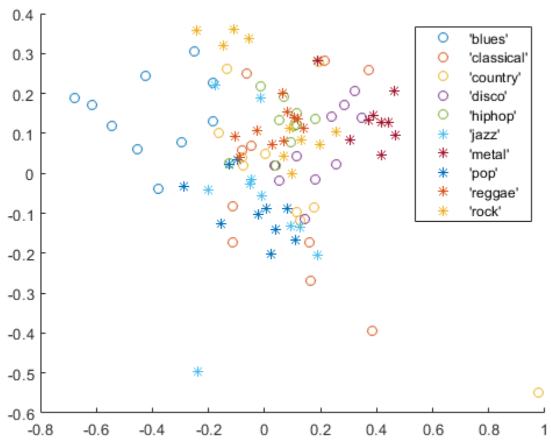

Figure 1: Class separability visualization for MVPA

The mean accuracy obtained with the help of the k fold cross validation for multi-class SVM classifier is 74.3 ∓ 5.9.

In the investigation of the relationship between functional connectivity and music genres, unfortunately, our analysis did not reveal any statistically significant associations.

For MVPA we used SVM algorithm on the voxels of our ROIs. The final estimated class is obtained by voting among all 10 frames (15 s) of a fMRI block. The classification accuracy is 68.2 which is less accueate than when we estimate audio features and then use SVM for classification. Figure 2 shows the confusion matrix of classification by MVPA.

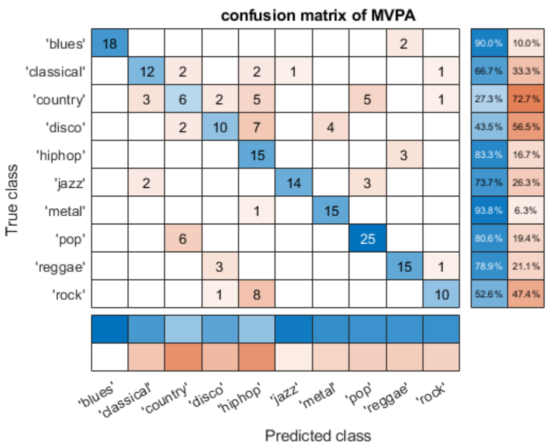

Figure 2: Confusion matrix for MVPA

In the investigation of the relationship between functional connectivity and music genres, unfortunately, our analysis did not reveal any statistically significant associations. Furthermore, even after employing Principal Component Analysis (PCA) on 100 selected edges with the lowest p-values to reduce the dimensionality to 2 dimensions for visualization, it is clear that the that there is no separability in the data, as depicted in Figure 3.

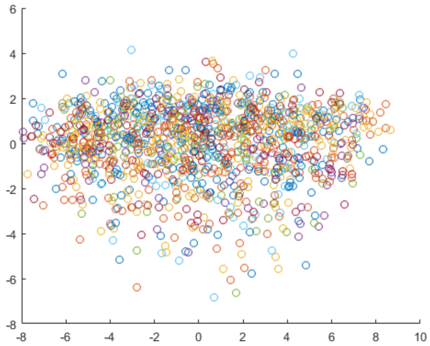

Figure 3: Class separability visualization for connectivity

In the exploration of potential correlations between functional connectivity and selected features, encompassing a total of 58 features (mean and variance), the highest correlation coefficient was 0.13. Also, the maximum correlation in each feature group was 0.08. These correlation coefficients are small and unfortunately no significant correlation was found.

Additionally, by selecting the top 100 (58*100) edges with the highest correlation for each feature (totally 385 unique edges) and applying PCA, 100 features were selected, which were employed in a multiclass support vector machine (SVM) classification task. The achieved accuracy of 56.9% was close to the chance-level, indicating that meaningful distinctions within the data were challenging to establish.

6. Conclusion

Based on a pervious study [6] and our work, it can be concluded that the separability between genre classes is larger in the voxel-based analysis than the functional connectivity analysis.

The poor results of functional connectivity may be due to brain parcellation, where the brain voxels are averaged, potentially leading to the unwanted omission of crucial information. This problem may be solved by using Networks of Multivariate Representations [4].

Refrences

[1] S. Chen, B. Mulgrew, and P. M. Grant, “A clustering technique for digital communications channel equalization using radial basis function networks,” IEEE Trans. on Neural Networks, vol. 4, pp. 570-578, July 1993.
[2] Levitin, D.J. Foundations of Cognitive Psychology: Core Readings, (2002).
[3] Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage, 59(4), 3677-3689.
[4] Hanke, M., Dinga, R., Häusler, C., Guntupalli, J. S., Casey, M., Kaule, F. R., & Stadler, J. (2015). High-resolution 7-tesla fmri data on the perception of musical genres–an extension to the studyforrest dataset. F1000Research, 4, 174.
[5] Sarrazin, N. (2016). Music and the Child. Open SUNY Textbooks
[6] Nakai, T., Koide-Majima, N., & Nishimoto, S. (2021). Correspondence of categorical and feature-based representations of music in the human brain. Brain and Behavior, 11(1), e01936.
[7] Li, T., & Ogihara, M. (2005, March). Music genre classification with taxonomy. In Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. (Vol. 5, pp. v-197). IEEE.
[8] Alías, F., Socoró, J. C., & Sevillano, X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. Applied Sciences, 6(5), 143.