Brain and Music: 
Music Genre Classification using Brain Signals

Pouya Ghaemmaghami*, and Nicu Sebe*
*Department of Information Engineering and Computer Science, University of Trento, Italy
Email: {p.ghaemmaghami,niculac.sebe}@unitn.it

Abstract—Genre classification can be considered as an essential part of music and movie recommender systems. So far, various automatic music genre classification methods have been proposed based on various audio features. However, such content-centric features are not capable of capturing the personal preferences of the listener. In this study, we provide preliminary experimental evidence for the possibility of the music genre classification based on the brain recorded signals of individuals. The brain decoding paradigm is employed to classify recorded brain signals into two broad genre classes: Pop and Rock. We compare the performance of our proposed paradigm on two neuroimaging datasets that contains the electroencephalographic (EEG) and the magnetoencephalographic (MEG) data of subjects who watched 40 music video clips. Our results indicate that the genre of the music clips can be retrieved significantly over the chance-level using the brain signals. Our study provides a primary step towards user-centric music content retrieval by exploiting brain signals.

Index Terms—Brain decoding, music genre classification, multimedia content retrieval, brain signal processing, EEG, MEG

I. INTRODUCTION

Nowadays, with the rapid growth of the Internet, a large amount of data has become available on-line. Among all the different sources of information, music certainly is one of the most important ones for entertaining people. This has arisen the need for organizing and managing these large digital music databases. Among many music descriptors such as song title, artist, album, etc, probably, the most widely used meta-data for indexing and retrieving music is the genre of the music [1], [2], [3]. As a result of this, music genre classification is one of the major research directions in music information retrieval (MIR) and research areas ranging from multimedia retrieval (music and video) to neuroscience.

Among all different approaches for searching for a music song, probably the most common one is the content-based approach. So far, various content-based genre classification methods have been proposed based on a variety of features. These features include MFCC, spectral centroid, spectral flux, zero crossings, energy, pitch, rhythm patterns, harmonic contents and etc [4], [5], [6], [7], [8]. However, in spite of all efforts done during the last years, content-based approaches are always dependent on the availability of multimedia contents. When such contents are not accessible, these approaches are not applicable anymore. Besides, the main drawback of such approaches is that they are not able to capture the personal preferences of the human listeners. Such preferences matter since there might be disagreement between human beings on the definition of the musical genre due to the fuzzy boundaries between the different genres. In light of this, in this paper we use an alternative method for music genre classification which is a user-centric approach that aims at capturing the perception of people. The rationale behind this is that the recommendation system that accesses the people’s perception of the music (e.g. via neurophysiological data), might be able to distinguish the music genre better.

Recent works on affective computing, suggests the possibility of decoding affects from neurophysiological data. In [9], authors captured physiological responses of participants while they were watching movie scenes. They showed that the predicted affects from physiological responses of participants are significantly correlated with their self-assessed emotional responses. Koelstra et al. [10] and Abadi et al. [11] studied emotional responses of subjects induced by excerpts of music and video clips. They show that emotional information is encoded in brain signals. Inspired by these works, in this paper, we address the specific problem of genre classification of music video clips using brain data. We show that music genre can be decoded from brain signals using a brain decoding paradigm. We tested our hypotheses on two neuroimaging datasets: (1) DECAF dataset [11] that contains magnetoencephalographic data of 30 subjects who watched 40 music video clips and (2) DEAP dataset [10] which contains electroencephalographic data of 32 subjects who watched the same 40 music clips. Figure 1 illustrates the overall framework used in our study.

To summarize, our main contributions are: (1) we study the possibility of user-centric genre classification by employing brain signals; (2) we apply our classification paradigm on two different brain datasets that contain different neuroimaging modalities (magnetoencephalographic data [11] and electroencephalographic data [10]); (3) we augment the aforementioned EEG and MEG datasets by providing music genre labels for each music clip which can be used and investigated in other studies. This study will contribute to various disciplines and research areas ranging from multimedia retrieval (music information retrieval) to neuroscience.

The rest of this paper is organized as follows. In section II we briefly review the literatures on genre classification. Afterwards, in section III we explain the employed datasets, data preprocessing, feature extraction and classification procedure. Furthermore, the method used regarding music
Fig. 1. The framework used in this study for music genre classification by exploiting brain signals.

II. RELATED WORKS

A. Content-Centric Music Genre Classification

There is a large body of works on content-based genre classification approaches. One of the earlier works is introduced by Tzanetakis and Cook [4] where the authors represent a music piece using timbral texture, rhythmic features, and pitch-related features. Their proposed features set has been widely used for music genre classification [5], [6], [7], [8]. Other characteristics such as contextual informational [12], temporal information [13], and semantic information [14] have been investigated in the literatures to improve the accuracy of genre classification. Recently, “sparse feature learning” methods have also been investigated for constructing a codebook for music songs [15], [16], [17], [18]. Elsewhere, Costa, et al. [3] proposed a robust music genre classification approach by converting the audio signal into a spectrogram and extracting features from this visual representation by treating the time-frequency representation as a texture image.

B. Affective Content and Brain Decoding

Brain decoding has recently received considerable attention across many domains particularly in brain computer interfacing and rehabilitation communities due to its potential for helping people who suffer from brain injuries [19], [20]. Nevertheless, in most cases, the performance of such brain decoding algorithm is not very high because of low signal-to-noise ratio and non-stationarity of brain signals.

However, in spite of these limitations, recent works on affective computing, shows the possibility of decoding affects from brain data. In [10], authors studied emotional responses of experimental subjects induced by music video excerpts. They employed electroencephalography (EEG) to record brain activity of participants while they were watching music video clips. A similar study was done in [11] with magnetoencephalography. These two studies indicate that emotional information is encoded in brain signals (MEG and EEG). In a recent work by Ghaemmaghami et al. [21], authors investigated the possibility of movie genre classification by employing MEG brain signals. They adopted a brain decoding paradigm to classify (MEG) data of experimental participants, who watched excerpts of movie clips, into four broad genre classes (Comedy, Romantic, Drama, and Horror). Moreover, they showed that there is a significant correlation between audio-visual features of the movie excerpts and MEG features extracted from visual and temporal areas of the brain.

Our brief review of literatures reveals that music genre classification has been achieved so far with content-based approaches. On the other hand, brain decoding algorithms were successfully employed on many tasks using various neuroimaging techniques. However, the efficacy of the brain decoding approaches on music genre classification has not been explored. Therefore, this study aims at investigating the possibility of classifying musical genres using brain data. As far as we know, we are the first showing that the extracted features from brain signals can be used for the music genre classification task.

III. EXPERIMENTAL SETUP

In this section, we describe the employed datasets, annotation process, feature extraction and classification procedure.
A. Datasets

In our experiments, we used two publicly available datasets. These datasets contain the electroencephalographic (EEG) and the magnetoencephalographic (MEG) data of volunteers who watched 40 music video clips. The advantage of using these two datasets is that they contain the same music clips (the duration of each clip is 60 seconds) so that the results can be compared. The details of these datasets are described below.

MEG dataset: The MEG dataset, we employed in this study is the DECAF dataset [11]. This dataset contains the MEG brain signals of 30 volunteers while they were watching 40 music video clips. These music clips were projected onto a screen placed in front of the subject inside the MEG acquisition room with 20 frames/second and at a screen refresh rate of 60 Hz. The magnetoencephalographic data were recorded in a magnetically shielded room with 1KHz sampling rate and in a controlled illumination using a Electa Neuromag device that outputs 306 channels (102 magnetometers and 204 gradiometers).

EEG dataset: The EEG dataset, we employed in this study is the DEAP dataset [10]. This dataset contains the EEG brain signals of 32 participants while they were watching 40 music video clips. These music clips were projected onto a screen placed about a meter in front of the subject at a screen refresh rate of 60 Hz. The electroencephalographic data were recorded in controlled illumination, at a sampling rate of 512 Hz, using a Biosemi ActiveTwo system that outputs 32 channels.

B. Annotating Music Clips

The definition of the musical genre is very subjective so that one song may belong to different genres according to different individuals. As a result of such arbitrariness in the definition of the musical genre, many researchers have shown that even major music taxonomies are very inconsistent [22], [1], [23], [3].

To deal with such a difficulty and given the few number of total samples in our employed datasets (40 excerpts of music clips), in this study, three human annotators were asked to watch the music video clips and classify each music clip into one of the two categories; The first category include the following genres: Pop, Dance, Disco and Techno. We refer to this category as the POP category. The second category include the following genres: Rock and Metal. We refer to this category as the ROCK category. The music genre of each clip was picked based on the majority voting between the annotators. To evaluate the consistency of the annotation across subjects, we measured the Cohen’s Kappa agreement between annotators’ labeling. The obtained average $\kappa$ across observers (69.8% ± 5%, $p-value < 0.001$) indicates a substantial agreement [24] between the annotators. We refer to the majority voting labels as the ground-truth labels. Table II presents the name of the music clips together with their ground-truth labels.

C. Feature Extraction

MEG features: Using the MATLAB Fieldtrip toolbox [25] and following [11], the MEG trials are extracted and pre-processed as follows: 1) Upon down-sampling the MEG signal to 300 Hz, High-pass and Low-pass filtering with cut-off frequencies of 1 Hz and 95 Hz are performed respectively. 2) Then, the spectral power of the 102 combined-gradiometer sensors of the MEG trials is estimated with a window size of 300 samples. 3) MEG features are calculated by averaging the signal power over four major frequency bands: theta (3:7 Hz), alpha (8:15 Hz), beta (16:31 Hz) and gamma (32:45 Hz). The output of this procedure for each trial is a matrix with the following dimensions: 102 (number of the MEG combined-gradiometer sensors) x 4 (major frequency bands) x 60 (length of a music clip in seconds).

EEG features: We used the publicly available pre-processed EEG data [10]. These pre-processing steps include: EEG signal down-sampling to 128 Hz, EOG artifacts removal and bandpass frequency filtering (4 - 45 Hz). Then, for every trial, the spectral power of each channel is estimated with a window size of 128 samples. EEG features are calculated by averaging the signal power over four major frequency bands: theta (3.7 Hz), alpha (8.15 Hz), beta (16.31 Hz) and gamma (32.45 Hz). The output of this procedure for each trial is a matrix with the following dimensions: 32 (number of the EEG sensors) x 4 (major frequency bands) x 60 (length of a music clip in seconds).

MCA features: For each second of the music video clips, low-level audio-visual features are extracted. These low-level Multimedia Content Analysis (MCA) features include: Mel-frequency cepstral coefficients (MFCC), spectral flux, zero crossing rate, pitch, energy, formants, silence ratio, lightning key, shadow proportion, visual details, colour variance, motion, and etc (see [11] for more details).

D. Classification Procedure

In the classification experiments we adopted an svm classifier under the leave-one-clip-out cross-validation schema to decode the brain/multimedia feature descriptors into our target genre classes. The feature descriptors are calculated as follows:

MEG-based, EEG-based and MCA-based descriptors: MEG/EEG/MCA descriptor of each trial is calculated by averaging the computed MEG/EEG/MCA features over the length of the music video clip (60 seconds). Hence, the length of each MEG descriptor is 408 (4 bands x 102 combined-gradiometer sensors), the length of each EEG descriptor is 128 (4 bands x 32 EEG sensors) and the length of each MCA
MEG+MCA fusion: For each subject, the MEG descriptors and MCA descriptors are concatenated and a feature vector of the length 408+166=574 is created as a result of such fusion.

EEG+MCA fusion: For each subject, the EEG descriptors and MCA descriptors are concatenated and a feature vector of the length 128+166=294 is created as a result of such fusion.

Note that the fusion MEG and EEG descriptors is not feasible, since the subjects in these two datasets are not the same (DEAP contains 32 subjects whereas DECAF contains 30 subjects).

IV. EXPERIMENTAL RESULTS

As explained in section III-D, for the sake of compatibility, the same svm classifier under the leave-one-clip-out cross-validation schema was employed to classify the extracted feature descriptors (MEG, EEG and MCA) into our genre classes (i.e. Pop and Rock). The ground-truth labels are used as the target labels in the classification procedure (see section III-B). Such evaluation, provides us with a fair comparison of our features descriptors (MEG, EEG and MCA). This procedure is performed in the two following scenarios:

Subject-level analysis
At subject level, the classification procedure was employed on the brain data of each subject separately. Thus, the classification of the MEG-descriptors and EEG descriptors are repeated 30 times and 32 times respectively (corresponding to the number of subjects in each dataset). For each subject, the 40 MEG/EEG descriptors (corresponding to the 40 music clips) are used as samples. Given the unbalanced number of samples for each genre, both accuracy and F-measure are reported as the metrics to compare the classification performance. These metrics are averaged over all subjects.

Table I compares the results of music genre classification using MEG, EEG and MCA descriptors with chance level (0.51). The chance level is computed by feeding random numbers with normal distribution into the classification procedure for 100 times. In both MEG and EEG case, the distribution of the obtained classification accuracies is better than chance level. This difference implies the existence of genre related information in the recorded brain activity. In the case of EEG descriptors, this difference is significant ($p-value < 0.001$). Furthermore, combining brain features (EEG descriptors and MEG descriptors) of each subject with MCA descriptors provides higher accuracy (0.75 and 0.82) than employing only EEG/MEG descriptors. Such brain-multimedia features fusion also outperforms the result of MCA descriptors suggesting the existence of complementary music genre related information in the brain signals.

Population-level analysis
To evaluate the efficacy of MEG/EEG descriptors at the population level, for each video clip, we computed the majority vote over predictions of the single-subject classification across all subjects. The results are summarized in table II. In case of EEG descriptors, the population level accuracy (75%) is higher than the single subject-level accuracy (60%) and it is also higher than the classification accuracy of only MCA descriptors (70%). In case of MEG descriptors, the population-level analysis does not perform well. Nevertheless, in single-subject level analysis, as explained in previous section, the average obtained results are better than chance level. However, the fusion of brain features and multimedia features (MEG+MCA and EEG+MCA) outperforms the results of only MCA features in population-level scenario. This, confirms the existence of complementary genre related information in brain signals and multimedia contents.

V. CONCLUSION

In this paper, we presented an approach for classification of music video clips into two broad genres (pop and rock) using brain signals. For the sake of simplicity, a simple SVM classifier has been employed to perform the music genre prediction using the extracted brain features. We evaluated our approach on two neuroimaging datasets (EEG and MEG). Regardless of the limitation of such datasets that is few and noisy samples, our results shows the possibility of user-centric classification of music clips into two broad genres only based on brain signals. This is one of the first studies that employs brain signals for music genre classification. As our future plan, we will replicate the experiments using portable brain recording devices (e.g. Emotiv sensors). Furthermore, we plan to extend this work by employing other classification algorithms with different features extraction strategies in order to improve our classification results.

REFERENCES


TABLE II

<table>
<thead>
<tr>
<th>MOVIE CLIP TITLES</th>
<th>GROUND-TRUTH</th>
<th>MCA</th>
<th>EEG</th>
<th>MEG</th>
<th>MEG+MCA</th>
<th>EEG+MEG</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edward Scissorhands</td>
<td>ROCK</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>57.5%</td>
</tr>
<tr>
<td>La La Land</td>
<td>MOVIE</td>
<td>GIF</td>
<td>GIF</td>
<td>GIF</td>
<td>GIF</td>
<td>GIF</td>
<td>75%</td>
</tr>
<tr>
<td>The Big Lebowski</td>
<td>ROCK</td>
<td>ROCK</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>75%</td>
</tr>
<tr>
<td>Toy Story</td>
<td>ROCK</td>
<td>ROCK</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>75%</td>
</tr>
<tr>
<td>The Princess Bride</td>
<td>ROCK</td>
<td>ROCK</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>POP</td>
<td>75%</td>
</tr>
</tbody>
</table>


