Investigating association between musical features and emotion through EEG signal analysis

Jesse Rengers
University of Twente
Department of EEMCS
Enschede, The Netherlands
j.j.rengers@student.utwente.nl

This study wishes to find an association between auditory features in music and physiological reactions measured with EEG in order to develop more accurate HCI models that can recognise human emotion. The study first investigates which auditory features are most likely to influence human emotion and use these auditory features to find a relation between auditory features and physiological response. The analysis shows a relation between the amount of energy in the higher frequencies of music with activity in the right frontal lobe of the brain. Music in major mode also shows to induce more activity in the right frontal lobe.

Keywords

Brain-Computer Interface, Electroencephalography, Emotion Recognition, Music, DEAP

1. INTRODUCTION

Listening to music is an emotional experience. Detecting these music-emotions can be useful in the design of Human Computer Interaction (HCI) systems, for example in the design of a system that recommends music to induce happy feelings to people who feel sad or bored. Describing emotions encounters a subjectivity issue, rendering the limited trustworthy of responses to questionnaires about felt emotion while listening to music. In a broader view of emotion psychology, it is widely accepted that emotional experience entails three main components: a physiological reaction to a stimulus, a behavioural response and a feeling[4]. By making use of modern techniques, we are able to measure these physiological reactions to auditory stimuli. This led to the introduction of using physiological reaction to analyse emotional response to music, which is believed to provide more objective insights than questionnaire. Techniques used to measure physiological reactions inside the brain, which is the centre of emotional processing, include Electroencephalography (EEG), Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET). Among these brain-imaging techniques, EEG is the only technique that can be performed outside a laboratory, making it suitable for a lot of brain activity studies. EEG is a passive technique to measure brain activity at the scalp of the head. EEG records the electrical potentials of a population of neurons underneath a number of non-invasive electrodes [4]

In recent years, a lot of research has already been done on emotion recognition. [2] is a survey on multiple studies on emotion recognition using EEG with various stimuli including memories, images, videos and music.

A considerable amount of these emotion-recognition studies using entertaining stimuli is focused on the link between physiological reaction and felt emotion. However, the link between music and its physiological reaction, where physiological signals were presumed to be emotional-laden, is still not well studied. A better understanding between the association of musical stimuli and brain physiological signals may indicate the success of emotion elicitation and pave a way to constructing more accurate emotion recognition models.

This study focuses on the link between music stimuli and its corresponding physiological reaction recorded by EEG, concentrating at emotion-related response. Possible results could help in the research field of Affective Computing and thereby in the creation of more intelligent Human Computer Interaction systems that can understand emotions and react to it.

The study makes use of the widely used and publicly available DEAP dataset, that provide EEG's from 32 participants that all watched music videos. More information about DEAP can be found in section 4.1. To limit the scope of this study, we will only focus on the acoustic features of the music-videos and ignore its visual features. To exclude physiological reactions induced by visual features in the EEG, we will focus ourselves more to the frontal and temporal lobes as these have been shown to be affiliated with emotions and hearing respectively [2, 12]. The occipital lobe that is primarily responsible for processing visual information will be less focused [1].

This paper is structured as follows: Section 2 will state the research questions that are going to be answered. Section 3 will explain earlier work and is also going to be answering **RQ2**. Then, section 5 will explain the method in which **RQ1** is being answered, where section 4 will provide some background information about the dataset and tools being used. In section 6 we show our results, which will be discussed in section 7. The paper will be concluded in section 8, which will also raise some ideas for further research.

33th Twente Student Conference on IT July. 3nd, 2020, Enschede, The Netherlands.

2. RESEARCH QUESTION

To investigate the relation between auditory features and emotion, we have two research questions to be answered:
RQ1 Which auditory features are most prominent in influencing human emotion?
RQ2 Can we find a relation between these auditory features and their induced physiological reaction?
Section 5 will explain the method in which these research questions will be answered.

3. RELATED WORK

In modern research, emotions are often categorised by two perpendicular factors, valence (pleasant-unpleasant) and arousal (high-low)[13]. For instance, being happy is a pleasant feeling (high valence), while being angry is unpleasant (low valence). Meanwhile, Arousal represents the arousness of this felt emotion, making us capable of distinguishing angry and sad emotions as both are considered posting negative valence.
[14] found that valence of affective musical excerpts can be distinguished by hemispheric asymmetrical pattern found in the frontal EEG activity. In particular, it was found that subjects exhibited greater relative left frontal EEG activity during the presentation of positively valenced musical excerpts and greater relative right frontal EEG activity during the presentation of negatively valenced musical excerpts. It could therefore be interesting to analyse the difference in the activity between the left hemisphere and the right hemisphere
A small literature study was conducted in order to answer **RQ 1**. A lot of studies have been done to investigate the relation between characteristics of musical structure and emotional responses during music listening [10]. The effect of Tempo and Pitch seem to be important musical structures to induce emotion. [5] showed that faster tempi was associated with happiness and slow tempi was associated with sadness.
Both [15] and [11] extract 20+ musical features with some overlapping features. In [11] the four most correlated musical features were shown to be: dissonance (or roughness), mode, onset rate and loudness. [9] is another study, containing 'Music-Emotion Rules' generated from a cumulative analysis of 102 unique studies. The most significant rules are shown in 1. Some of these rules are in agreement with [5]. Considering these former studies, we decided to consider dissonance, mode, tempo and brightness for our current study.

Emotion	Structural Music-Emotion Rules
Happy	Tempo fast, Mode major, Harmony simple, Pitch high
Angry	Harmony complex, Tempo fast, Mode minor, Loudness loud
Sad	Tempo slow, Mode minor, Pitch low, Harmony complex
Tender	Mode major, Tempo slow, Loudness soft, Harmony simple

Table 1. Most significant 'Structural Music-Emotion Rules' taken from [9]

4. BACKGROUND

4.1 DEAP dataset

As already stated, DEAP is a widely used dataset. The DEAP dataset consists of the EEG recordings of 32 volunteers that watched a 40 one-minute excerpts of music videos that were all available on YouTube [6]. These 40 music videos are a subset out of a set of 120 videos that were selected based on emotional tags, each of which was labelled by 14 independent volunteers as arousal, valence and dominance scores. The scores were used as expected emotions in the experiment. For all 120 music videos, a normalised arousal and valence score was calculated, and the music videos where plotted on an arousal-valence space. The 10 most emotionally-extreme music videos out of the 4 quadrants (low arousal - low valence, low arousal - high valence, high arousal - low valence, high arousal - high valence) were selected. Next, 32 different volunteers watched all 40 one-minute excerpts while their brain activity was measured using EEG together with peripheral physiological signals, which are not considered here. The volunteers also rated each excerpt based on arousal, valence, dominance, liking and familiarity (felt emotion). Because of 15 of the 40 music videos are not available on YouTube anymore, we can only make use of 25 videos.

4.2 MIRToolbox

In order to analyse the music extracts, we will make use of MIRToolbox as proposed in [3]. MIRToolbox is an open-source MATLAB toolbox developed at the University of Jyväskylä. The software includes function to extract auditory features such as timbre, tonality, rhythm or form [8]. The reliability of the functioning of MIRToolbox is questioned in [7] where it was found that the performance for brass instruments was not satisfactory. However, performance test on other features such as beat, rhythm and melody is still needed.

5. METHOD

First we have to decide which EEG channels to use for our analysis. To minimize the effect of visual stimuli induced by the music videos, we will concentrate on the frontal and temporal lobe of the brain. Considering earlier studies on emotion recognition in EEG, like [16], and the available channels that were used to generate DEAP, we chose to use the following set of electrodes: Fp1, Fp2, F3, F4, F7, F8, T7, T8. These electrodes are placed according to the international 10-20 system. The placement of these electrodes is shown in figure 1

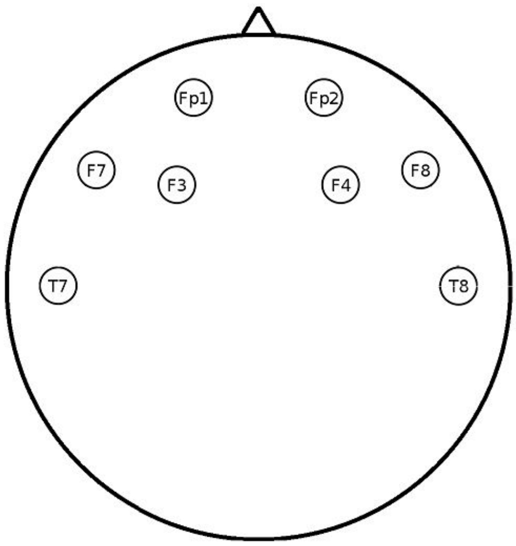

Figure 1. Placement of the 8 electrodes that are studied

Then, we divide all songs and neural signals into segments. All songs $s \in S$ and trials $t \in T$ are segmented into 56 segments with a length of 5 seconds with an overlap of 4 seconds such that $s_i, s_{i+1}, s_{i+2}, .., s_n$ correspond to $t_i, t_{i+1}, t_{i+2}, .., t_n$. Then, features from these segments are extracted as described in the following sections.

5.1 Neural signal feature extraction

Pre-processing of the Neural signal was already done by the creators of the DEAP dataset. The following steps were taken: The data was downsampled to 128 Hz, EOG artefacts were removed, a bandpass frequency filter from 4.0-45.0 Hz was applied and the data was averaged to the common reference [6].

The Power Spectral Density (PSD) is calculated over all t_i using Welch's method with a hamming-window length of 1 second and an overlap of $\frac{1}{2}$ second. The average band power in these PSDs gets calculated in 4 frequency bands: Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-20 Hz) and gamma (20-30 Hz). The average power is calculated by integrating the PSD estimate with the rectangle method.

5.2 Auditory feature extraction

The average tempo, mode, dissonance and brightness are calculated over all segments s_i using the MIRToolbox.

After the features are extracted, we perform two kind of statistics on the data: participant level statistics and group level statistics. For both statistics, we calculate a chance level ($cl(t)$) which indicates the chance that a trial t is correlated with a random song s. The procedure to calculate the chance level for trial x is the following: we calculate the correlation coefficient (cc) between trial x and song y, such as shown in figure 2, for all y where $x \neq y$. The correlation coefficient is calculated using the segments as data points.

$$cl(t^x) = mean(\{cc(t_i^x, s_i^y)|\forall y, y \neq x, i \in [1-56]\})$$

After the chance level is known, we calculate the matched correlation coefficient, which is the correlation coefficient between trial t^x and song s^x

$$ml(t^x) = cc((t_i^x, s_i^x)|i \in [1-56])$$

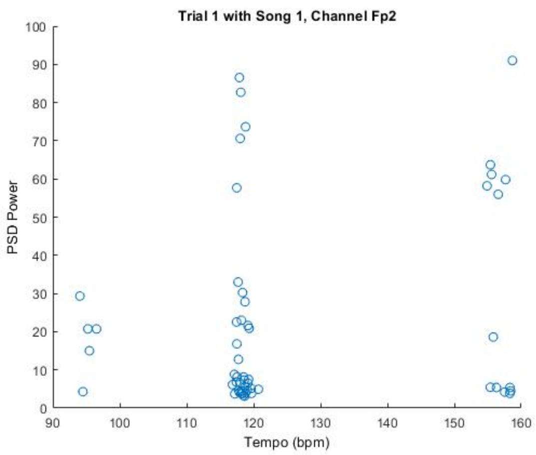

Figure 2. Scatterplot arbitrarily taken from trial 1 with song 1 in the Fp2 channel and Tempo as auditory feature. The corresponding correlation coefficients is 0.2285.

5.3 Participant level statistics

For all frequency bands and all channels we calculate the chance level and the matched level as described. Then we check if there is a significant difference between the chance level and the matched level using a one-sampled t-test. Only the significant ($\alpha = 0.05$) correlation coefficients are considered.

5.4 Group level statistics

For all participants $p \in P$ we calculate the chance level and matched level for all trials $t \in T$ as described. Then for all trials we average the chance level and matched level over all participants:

$$cl(t^x) = mean(cl(t_{p1}^x) + cl(t_{p2}^x) + .. + cl(t_{p32}^x))$$

$$ml(t^x) = mean(ml(t_{p1}^x) + ml(t_{p2}^x) + .. + ml(t_{p32}^x))$$

And we perform a t-test to detect significant differences between the chance level and the matched level. Only the significant ($\alpha = 0.05$) correlation coefficients are considered.

6. RESULTS

6.1 Participant level results

A total of 378 significant correlation coefficients were found using the method described above. These values are grouped by feature and are shown in table 3 for the tempo feature, table 4 for the brightness feature, table 5 for the mode feature and table 6 for the dissonance feature. All tables show the absolute value of the amount of participants showing positive correlations minus the amount of participants showing negative correlations. Only correlation coefficients that show a significant difference in the matched level over the chance level as described earlier are considered. The Channels are grouped by hemisphere and lobe for easier analysis.

6.2 Group level statistics

At the group level, there are three significant differences in the correlation coefficient in the Beta and Gamma bands for electrode Fp2. These are shown in table 2 together with their chance level, matched level and standard deviation.

Brightness induced a significant difference in the correlation coefficient in the Beta and Gamma bands in channel Fp2. Mode induced a significant difference in the Gamma band in channel Fp2. The other auditory features (tempo and dissonance) show no significant difference in any of the frequency bands and channels.

7. DISCUSSION

To answer our second research question (**RQ2**) we have to find a relation between auditory features and their induced physiological reaction.

From visual inspection of the data on participant level, we can conclude that the feature **tempo** on average is positively correlated with PSD power in the Theta, Alpha and Gamma band in all studied electrodes. Only the Beta band does shows a more neutral diversity, containing some evidence that Tempo is negatively correlated in the left hemisphere of the frontal lobe.

Brightness is also positively correlated with PSD power in almost all frequency bands and channels, meaning that the more energy in high frequencies of the music, the more PSD power is measured.

Mode is a feature that shows some diversity, making it

Group level resuls						
Feature	Channel	Band	Chance level	Matched level	SD	Sign
Brightness	Fp2	Beta	-0.0199	0.0128	0.0947	Positive
Brightness	Fp2	Gamma	0.0014	0.0262	0.0712	Positive
Mode	Fp2	Gamma	-0.0022	0.0158	0.0503	Positive

Table 2. Group level analysis results

Tempo				
Channel	Frequency Band			
	Theta	Alpha	Beta	Gamma
Fp1	3	2	1	3
F3	1	0	1	2
F7	3	1	1	1
T7	3	2	0	0
Fp2	2	2	2	3
F4	3	3	2	1
F8	3	1	1	2
T8	3	1	0	0

Table 3. Amount of positive (green) or negative (red) correlation coefficients per channel and frequency band induced by tempo

Brightness				
Channel	Frequency Band			
	Theta	Alpha	Beta	Gamma
Fp1	1	4	5	1
F3	2	4	4	1
F7	0	4	2	3
T7	0	4	5	0
Fp2	3	7	6	4
F4	4	2	0	3
F8	3	5	0	0
T8	3	3	2	2

Table 4. Amount of positive (green) or negative (red) correlation coefficients per channel and frequency band induced by brightness

hard to make conclusions. However, we see some indication that Mode can be positively correlated in the right hemisphere.
The feature **Dissonance** shows a low amount of significant differences, making it impossible to draw a general conclusion about this feature.
All participant analysis results considered, it is hard to draw conclusions with such a limited amount of data that shows no consentient about possible relations.

The group level analysis shows a limited amount of results. The result with the brightness feature is somewhat in line with the result of the participant analysis, since table 4 also show a high number of positive correlations for Fp2 in the Beta and Gamma band. This would mean that a large amount of energy in high frequencies of the music can induce activity in the right frontal lobe at frequencies 12-30 Hz. However, what also can be seen in table 4 is that there is a relatively high number of positive correlations for Fp2 in the Alpha band, which is not reflected in the group level statistics, so caution is needed when such conclusions are made.

Mode				
Channel	Frequency Band			
	Theta	Alpha	Beta	Gamma
Fp1	1	4	0	3
F3	2	1	2	1
F7	0	1	1	2
T7	0	0	0	3
Fp2	2	2	1	5
F4	2	0	0	1
F8	2	2	0	2
T8	1	0	1	1

Table 5. Amount of positive (green) or negative (red) correlation coefficients per channel and frequency band induced by mode (major/minor)

Dissonance				
Channel	Frequency Band			
	Theta	Alpha	Beta	Gamma
Fp1	1	0	1	1
F3	0	1	1	0
F7	0	1	1	0
T7	0	0	0	1
Fp2	0	0	0	1
F4	0	2	1	1
F8	0	1	1	1
T8	1	1	0	0

Table 6. Amount of positive (green) or negative (red) correlation coefficients per channel and frequency band induced by dissonance (or roughness)

8. CONCLUSION AND FUTURE WORK

This research studied the association between auditory features in music and its induced physiological reaction in the brain to detect emotions. The EEG recordings of 32 participants that watched 25 music videos were studied where we focused on 8 electrodes in the frontal and temporal lobe of the brain. These neural signals were compared with 4 auditory features in the music (Tempo, Brightness, Mode, Dissonance) and we tried to find significant relations. On a participant level no consentient relations were found but group level statistics found that brightness in the music induced activity in the right frontal lobe at frequencies 12-30Hz and music in major mode induced activity in the right frontal lobe at frequencies 20-30Hz.
Further research should be done to find more evidence for these relations. One can for example specify on the brightness feature using custom music excerpts that are made to highlight the difference in brightness. This can be valuable to strengthen the evidence in this relation.
In the existing literature, we found some clues that led to believe that tempo has an effect on the activity in the frontal lobe of the brain. However, this was not one of the conclusions that we could draw from our group level analysis. Therefore, it could also be interesting to study this association further. Another dataset, dedicated to auditory feature extraction can also lead to other insights.

9. REFERENCES

[1] P. Abhang, B. Gawali, and S. Mehrotra. *Introduction to EEG- and Speech-Based Emotion Recognition.* 2016. cited By 17.

[2] S. M. Alarcao and M. J. Fonseca. Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*, 10(3):374–393, 2017.

[3] D. A. Effects. AN EVALUATION OF AUDIO FEATURE EXTRACTION TOOLBOXES David Moffat , David Ronan , Joshua D . Reiss Center for Digital Music Queen Mary University of London Mile End Road. pages 1–7, 2015.

[4] M. S. Gazzaniga, R. B. Ivry, and G. Mangun. *Cognitive Neuroscience. The biology of the mind, (2014).* Norton: New York, 2006.

[5] K. Hevner. The affective value of pitch and tempo in music. *The American Journal of Psychology*, 49:621–630, 1937.

[6] S. Koelstra, C. Mühl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. DEAP: A database for emotion analysis; Using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, jan 2012.

[7] N. Kumar, R. Kumar, and S. Bhattacharya. Testing reliability of Mirtoolbox. *2nd International Conference on Electronics and Communication Systems, ICECS 2015*, (Icecs):710–717, 2015.

[8] O. Lartillot and P. Toiviainen. A matlab toolbox for musical feature extraction from audio. pages 237–244, 2007. cited By 447.

[9] S. R. Livingstone, R. Muhlberger, A. R. Brown, and W. F. Thompson. Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, 34(1):41–64, 2010.

[10] E. Miranda and J. Castet. *Guide to Brain-Computer Music Interfacing.* 01 2014.

[11] T. Petri. Exploring relationships between audio features and emotion in music. *Frontiers in Human Neuroscience*, 3(Escom):260–264, 2009.

[12] J. Pickles. *An introduction to the physiology of hearing.* Brill, 2013.

[13] J. A. Russell. Affective space is bipolar. *Journal of Personality and Social Psychology*, 37(3):345–356, 1979.

[14] L. A. Schmidt and L. J. Trainor. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition and Emotion*, 15(4):487–500, 2001.

[15] Y. Song, S. Dixon, and M. Pearce. Evaluation of musical features for emotion classification. *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012*, (Ismir):523–528, 2012.

[16] S. Valenzi, T. Islam, P. Jurica, and A. Cichocki. Individual Classification of Emotions Using EEG. *Journal of Biomedical Science and Engineering*, 07(08):604–620, 2014.