POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Exploring the correlation between music features and induced emotions using physiological signals in a listening experiment

Laurea Magistrale in Music and Acoustics Engineering - Ingegneria Acustica e della Musica

Author: Andriana Takić

Advisor: Prof. Augusto Sarti

Co-advisors: Sebastian Gonzalez, Alessandra Calcagno

Academic year: 2021-2022

1. Introduction

The goal of this study was to explore the human emotional responses to temporal features in music, focusing on harmonic tension, by means of acquiring various physiological data in a listening experiment, supported by self-report questionnaires. A listening experiment was designed and conducted with 10 recruited subjects wearing non-invasive equipment in order to have their physiological signals monitored and recorded, while they listen to certain music stimuli. The chosen signals to acquire were electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), and respiratory activity (RSP). The self-report questionnaires were added to the study with the goal to collect data about the subjects' lifestyle, musical expertise and preferences, and emotional competencies, as well we the emotions they feel while listening to the music stimuli. The analysis of various statistical features was performed, exploring the correlation between the physiological and self-report responses, as well as between the manually annotated tension in music and the recorded physiological data.

2. Emotions and music

Music is often related to emotions, as it has an ability to affect, shape, and even manipulate emotional states of individuals. One of the most important goals of the research on emotions in music is to understand how features in music composition and performance relate to inducing various emotional responses. Researchers are usually interested in cases where emotions in music are perceived similarly by different listeners, which is referred to as *listener agreement*. Results of several studies over the past century suggest that changes in music attributes correlate with changes in emotional responses [1, 3].

The most common and simple approach to measure emotional responses to music is through a first-person perspective or self-report. The main disadvantage of such self-report methods is that a subject might feel discomfort, insecurity, or even lack of self-awareness in recognizing and sharing their experiences related to the stimuli. To compensate for the disadvantages of self-reporting, various physiological approaches towards emotion recognition are often used as an alternative or an addition to self-reporting.

3. Physiological signals and sensors

The signals used in the physiological approach to emotion recognition are described below.

Electroencephalogram (EEG) is a method to record electrical activity of the human brain using electrodes placed on the scalp; it reflects how many neurons in the brain network communicate with each other via electrical impulses and in what way. Compared to other bioelectrical signals, EEG signals are characterized by their high temporal resolution, which allows capturing fast and precise cognitive, perceptual, emotional, and motor processes. The disadvantage of EEG signals is the low spatial resolution on the scalp, since the measured potential on the scalp is the average response of several spatial areas activated by a particular stimulus. Applications of EEG are mainly focused either on event-related analysis or on its spectral content (popularly referred to as "brain waves").

Electrocardiogram (ECG or EKG) produces a graph of voltage across time of the electrical activations that lead to the contraction of the heart muscles. The heart rate is regulated by the autonomic nervous system (ANS); upon excitation, the neurotransmitter norepinephrine gets released and causes the increase of the heart rate, which occurs throughout the sympathetic part of the ANS. The common features of the ECG signal are the heart rate (HR), the interbeat interval (IBI) and the heart rate variability (HRV). The data collected over a time interval cannot accurately relay information on the response to a stimulus but can rather help understanding the total reaction to a signal in the considered time interval.

Electrodermal activity (EDA) refers to changes in electrical conductance measured on the skin surface, usually as a result of emotional activation, increased cognitive workload, or physical exertion. Namely, the level of sweating gets increased as a response to the signals coming from the brain. Even though the change of the sweat on the skin surface may not be noticeable, the electrical conductance can increase significantly as the pores of the skin begin to fill. The EDA sensor measures the electrical conductance of the skin, by passing a small amount of direct current between two electrodes that are in contact with the skin. The skin conductance can be further divided into two components: the phasic skin conductance (PSC) corresponding to the rapid changes and the tonic skin conductance (TSC) corresponding to the raw conductance level and slow shifts. Although associated to fast external events, a phasic response can occur 1 to 5 seconds after the stimulus.

Respiratory activity (RSP) is related to the breathing of an individual and it is set and controlled by the respiratory center of the brain. The respiratory rate is a measure representing the number of breaths taken in one minute (BPM); a healthy adult in a resting state can have a respiratory rate between 12-15 BPM. The features mainly used to describe the RSP are the breath-to-breath interval (BBI) and the respiratory rate variability (RRV).

4. Spectral and statistical measures

Power spectral density (PSD), or simply power spectrum, gives information about a signal's power distribution across frequency. More precisely, it specifies the power levels of the frequency components present in a signal. The spectral density is usually estimated using Fourier transform methods, such as the Welch method.

Correlation is defined as any statistical relationship, whether causal or not, between two random variables. The correlation coefficient that indicates the strength of the relationship between two variables can be found as:

$$r_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 (y_i - \overline{y})^2}}, \tag{1}$$

where r_{xy} stands for the correlation coefficient of the linear relationship between the variables x and y, x_i and y_i represent their samples, and $\overline{x}$ and $\overline{y}$ represent their mean values, respectively. The relationship between more than two variables can be made possible through a correlation matrix. Correlation only assesses relationships between variables; thus, it does not imply that one variable causes the changes in another variable.

5. Methodology

A total number of 10 subjects aged from 22 to 36, out of which 7 males and 3 females, were recruited for the study. Upon invitation, a simple sorting process was conducted, mainly based on the applicant's basic health status, gender, and time availability. Before taking part in the study, the subjects were asked to read and agree to the Informed Consent approved by the research Ethical Committee of Politecnico di Milano, describing the process of acquiring, storing, and analyzing the data.

5.1. Self-reporting questionnaires

All subjects were required to fill out the questionnaires aimed to acquire data related to their lifestyle, musical expertise, music preference, and emotional competencies, as well as the emotions elicited while the subject listens to the music pieces during the data acquisition session. The self-report aspect was added to the study in order to give more context to the results and explore conditional correlations, with the assumption that the subject's preferred music genre, emotional competencies, and lifestyle habits could be correlated with their physiological responses to certain music stimuli.

5.2. Music stimuli

The music stimuli set created for this study contains recordings of four pieces of polyphonic choral music, performed in a live concert by "Discanto Vocal Ensemble". To have a part of music stimuli that is represented in a simpler and more direct fashion [2], the set was enriched with 12 chord progressions played on a MIDI keyboard. For each of the pieces, the different aspects of music tension through time were manually annotated by experts: the annotation process was done considering the corresponding music scores, but also the live performances.

5.3. Technical equipment

The physiological signals chosen for the continuous acquisition are electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), and respiratory activity (RSP). The equipment needed to acquire the described signals was connected to a PC for the real-time visualization of the acquired signals and their storage for further processing.

The **EEG** cap with 61 electrodes was placed on the subject's head according to the international 10-20 EEG placement system [5]. The signals coming from 61 channels were recorded with a sampling rate of 1024 Hz.

The **ECG** setup included placing three adhesive and disposable electrodes on the subject's thorax to record their cardiac activity. The ECG was acquired with a sampling rate of 256 Hz.

The **EDA** was measured using two clamps, each coupled with a sensor and wrapped around the medial phalanx of the index and middle finger of the non-dominant hand, in a way that ensures stable conductivity and is comfortable for the subject. The EDA signal was recorded with a sampling frequency of 2048 Hz.

The **RSP** was recorded using a belt with a sensor, wrapped around the subject's thorax, just under the sternum. The sensor tracks the movement of the thorax while the subject breathes in and out. The samples of the RSP were taken with a frequency of 256 Hz.

5.4. Experimental protocol

The data acquisition sessions were conducted in a light and quiet room of Politecnico di Milano, with a technical and educational support from the Brain Lab. Before the assigned acquisition session, the subjects were asked to fill out the provided online questionnaire and were given instructions related to the acquisition session. Upon arrival, the subjects were given deeper explanations about the experiment and asked to give a formal written consent of participation in the study. The average duration of each acquisition session was 1 hour 45 minutes, including the equipment setup, data acquisition and monitoring, and lastly cleaning.

The data acquisition included continuous recording of the four physiological signals and compiling the adjusted GEMS-9 questionnaires after each listened piece. During the session, the subjects were seated in a chair, at a comfortable distance from the table. The disposable electrodes for ECG were attached to the subject's body, the two EDA clamps were wrapped around their two fingers, and the respiratory belt was wrapped around their thorax. All the input signals were checked, accordingly adjusting the equipment if necessary.

Then, the EEG cap was set up on the subject's head, in such a way that it is centered and that the reference electrode is positioned at the midpoint of the nasion-inion distance. When the EEG cap was properly positioned and comfortably tightened with the strap around the subject's chin, the holes of the electrodes were filled with the conductive gel. The gel was applied into the holes in a systematic order, doing a circular motion with the syringe to move the hair and thus get closer to the skin.

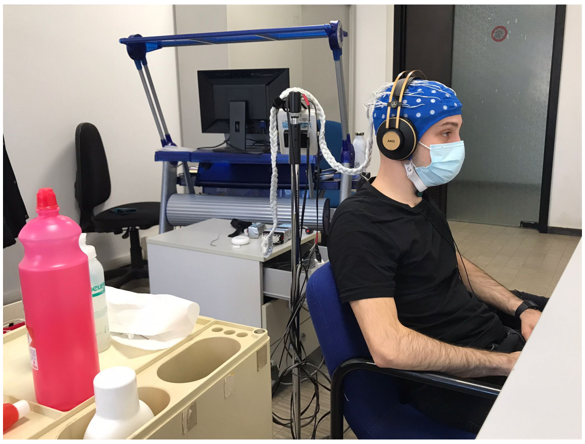

Figure 1: A subject with set-up equipment during an acquisition session

When the equipment was finally set up as represented in Figure 1, the subject was informed about the task in detail, asked to keep their eyes open, stay relaxed, and keep any physical movement at minimum, especially while listening to the music stimuli.

Before starting the task, it was needed to record the baseline signals, corresponding to the subject's resting state. One-minute-long EEG baseline signal was recorded for the case when the subject kept their eyes closed (EC), and five minutes of both EEG and EDA baseline signals were recorded while the subject keeps their eyes open (EO).

During the task, the music stimuli were played in randomized order, notes of which were accordingly documented. The time synchronization was established by pressing the external button at the same time as a keyboard button that played the music stimuli from the PC. During the data acquisition, the signals were continuously monitored on the screens.

After the data acquisition was completed, it was necessary to carefully take off the EEG cap of the subject's head, to partially remove the gel, and to take off the rest of the equipment. The subject was then asked for feedback about the whole procedure and allowed to leave the room.

In the end, the acquired signals were checked and saved, the disposable equipment thrown away, while the reusable equipment carefully cleaned and stored. Figure 2 represents the whole above-described experimental protocol.

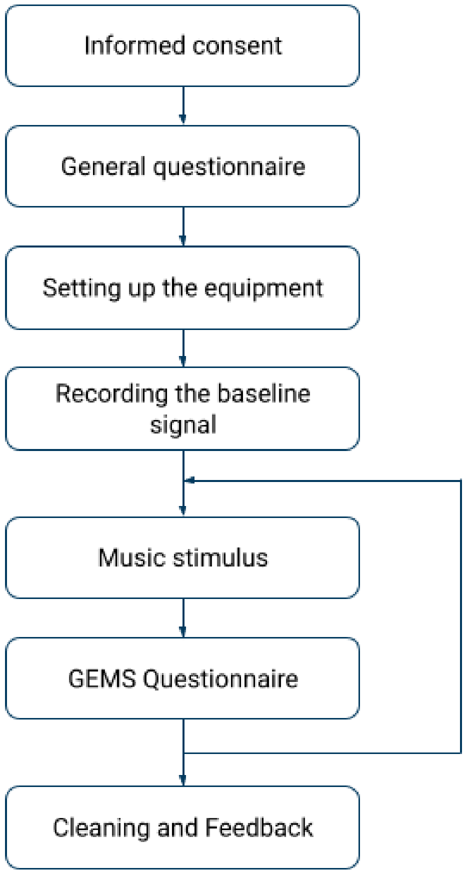

Figure 2: The diagram representing the designed experimental protocol

5.5. Data processing

The raw EEG data processing was performed using EEGLAB in MATLAB R2021b (MathWorks, Inc.), while the the acquired EDA, ECG, and RSP signals were processed in Python.

The created pipeline shown in Figure 3 includes importing and organizing data, preprocessing, feature extraction and analysis.

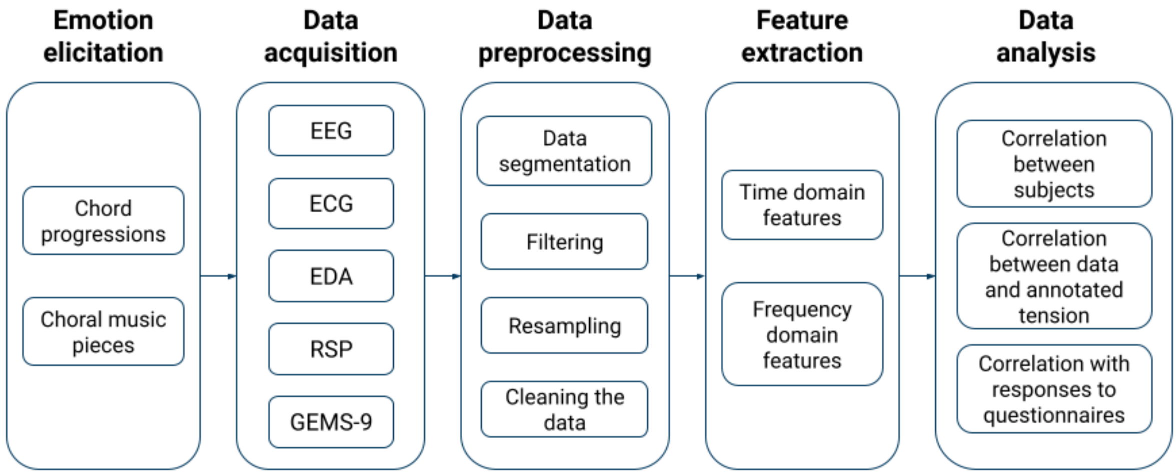

Figure 3: The diagram representing the data processing pipeline

6. Analysis

The temporal analysis was mainly focused on the EEG signals, due to the high temporal resolution. The correlation between various features and the questionnaires' responses was explored.

6.1. EEG

The main features extracted from the EEG signals were the estimated PSDs of theta (4-8Hz) and alpha (8-13Hz) frequency bands. The average alpha activity for the EEG signals corresponding to the music tracks is stronger compared to those corresponding to the baseline, an example of which is shown in Figure 4.

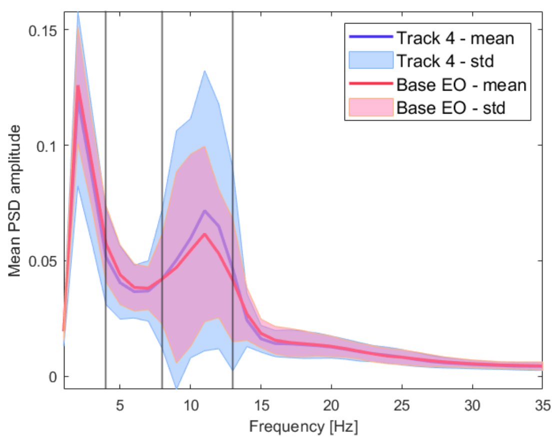

Figure 4: Estimated PSDs of the EEG signals corresponding to the baseline and a music track, averaged across subjects. The black lines mark the limits of theta and alpha frequency bands.

The inter-subjects analysis refers to comparing the EEG signals among subjects and exploring the correlation between their features, which also requires the process of standardization. The arrays containing values of estimated PSD for segments of 1 second were compared among all 10 subjects; the resulting correlation matrix measuring the correlation between the PSDs of the alpha band of 10 subjects (Figure 5) showed very low non-diagonal values, thus almost no correlation.

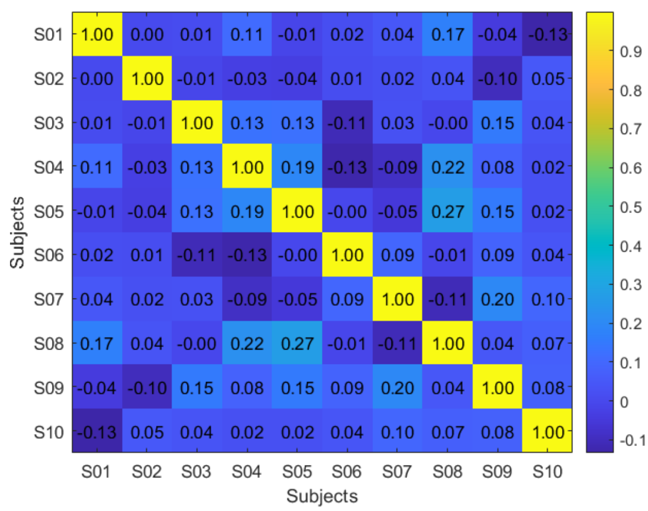

Figure 5: Correlation matrix between all subjects for "Intellige Clamorem Meum", across the whole temporal domain.

Therefore, it was decided to compute correlation matrices for shorter intervals of interest, instead for the whole temporal domain.

The correlation matrix for all subjects corresponding to the 3-seconds-long interval where the number of positive subject-to-subject correlations is the largest is shown in Figure 6.

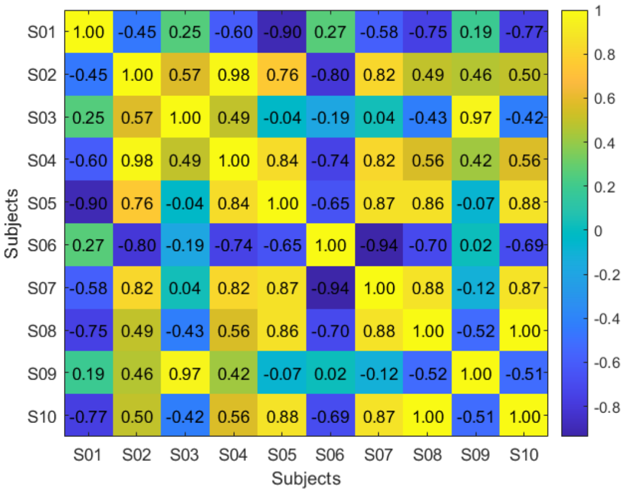

Figure 6: Correlation matrix between all subjects for "Abide with me" in a window ranging from 31 to 34 seconds of the track.

The interval corresponding to the strongest correlation among subjects also corresponded to the highest level of annotated tension (from 32.27 to 36.7 seconds) for the track "Abide with me".

6.2. EDA

Since the EDA signals are "slow", they might not be able to catch responses during continuous or overlapping stimuli. Several features were extracted, including the SCR onsets, SCR peaks, and SCR amplitude; however, due to not having observed significant correlation in preliminary visual and numerical inspection of the features, and due to the time limitations, further analysis of EDA signals has not been performed.

6.3. ECG

The analysis of ECG signals was mainly based on extracting long-term features, taking into account the whole temporal domain of a signal. The relevant features for the heart rate and the heart rate variability were extracted for each track and each subject.

6.4. RSP

The RSP and RRV features were extracted for each track and each subject, and further used in correlating with the responses to questionnaires.

6.5. Questionnaires

6.5.1 Gold-MSI

Responses of all 10 subjects to the Gold-MSI (Goldsmiths Musical Sophistication Index) [4] questionnaire have been collected and the results correlated with the features extracted from EEG, ECG and RSP signals. Positive correlation (**0.57**) was found between the PA (Perceptual Abilities) measure and the EEG alpha activity of the subjects. The BPM feature of ECG correlated with the Gold-MSI measures of MA (Musical Training), SA (Singing Ability), EM (Emotions), and GM (General Musical Sophistication), with coefficient values of **0.56**, **0.55**, **0.57**, and **0.63**, respectively. As for the RRV features, the largest values of correlation coefficients were found between PSD of the RRV LF (low frequency) band and several Gold-MSI measures: in particular, AE (Active Engagement), MT, SA, and GM, with values **0.62**, **0.55**, **0.69**, and **0.68**, respectively.

6.5.2 SREIT

Responses of all 10 subjects to the SREIT (Schutte Self-Report Emotional Intelligence Test) [7] questionnaire have been collected and the results correlated with the features extracted from EEG, ECG and RSP signals. Positive correlation was observed between the EEG theta activity and the SREIT values, with the correlation coefficient equal to **0.57**. From the ECG features, only SDNN was moderately positively correlated with the SREIT values, with the coefficient of **0.51**. No significant correlation was found between any of the RRV features and the SREIT values.

6.5.3 STOMP-R

The elaboration of responses to STOMP-R (Revised Short Test of Music Preferences) [6] questionnaire, according to the MUSIC model (standing for Mellow, Urban, Sophisticated, Intense, and Campestral music) has been performed. However, due to not having a standardized evaluation model, and due to the limited number of subjects in order to perform factor analysis, further processing was not performed on the responses to the STOMP-R questionnaire.

7. Conclusion

It can be concluded that the experimental protocol was well designed, as well as that the acquired data and the proposed pipelines can be used for further research. The preliminary analysis showed certain correlation between the explored features; however, it would be necessary to acquire additional data with significantly larger number of subjects to obtain more meaningful statistical measures and reach more objective conclusions about the correlations between music features and induced emotions.

The main contributions of this work are the designed experimental protocol with detailed instructions, the acquired datasets containing data for four physiological signals for 10 subjects and their responses to self-report questionnaires, the proposed pipeline for data processing and analysis, and the preliminary results. This study also showed the organizational difficulty of realizing interdisciplinary research and has set the minimum base for any future research project that involves the work of different laboratories.

When exploring the feature of tension in music, it is advised to use music pieces or excerpts that can have more objective representation of tension; additionally, having more diverse pieces would be useful in terms of exploring the variety of induced emotions. The exploration of more features and their connections, as well as the combination with fMRI or fNIRS data, could result in more advanced affective computing systems used in healthcare, therapy, entertainment, and other fields where it is desired to monitor or affect the emotional state of individuals.

8. Acknowledgements

I would like to thank my research supervisors, prof. Augusto Sarti, Sebastian Gonzalez, and Alessandra Calcagno, for their assistance and their educational and emotional support. I am also grateful to the Brain Lab of Politecnico di Milano for providing the necessary equipment for the study. Lastly, I would like to thank my colleagues, family, friends, all the people who contributed to my personal and professional growth, and all the opportunities to live a life of music, emotions, engineering, and exploration.

References

[1] Patrik N Juslin and Petri Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, 33(3):217–238, 2004.

[2] Stefan Koelsch, Simone Kilches, Nikolaus Steinbeis, and Stefanie Schelinski. Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS One*, 3(7):e2631, 2008.

[3] Leonard B Meyer. *Emotion and meaning in music*. University of chicago Press, 2008.

[4] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642, 2014.

[5] Marc R Nuwer, Giancarlo Comi, Ronald Emerson, Anders Fuglsang-Frederiksen, Jean-Michel Guérit, Hermann Hinrichs, Akio Ikeda, Fransisco Jose C Luccas, and Peter Rappelsburger. IFCN (International Federation of Clinical Neurophysiology) standards for digital recording of clinical EEG. *Electroencephalography and clinical Neurophysiology*, 106(3):259–261, 1998.

[6] Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.

[7] Nicola S Schutte, John M Malouff, Lena E Hall, Donald J Haggerty, Joan T Cooper, Charles J Golden, and Liane Dornheim. Development and validation of a measure of emotional intelligence. *Personality and individual differences*, 25(2):167–177, 1998.